OXFORD

Particle Detectors

fundamentals and applications

Hermann Kolanoski
and Norbert Wermes

Particle Detectors

Particle Detectors

Fundamentals and Applications

Hermann Kolanoski
Norbert Wermes

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Reprinted with corrections in 2022

Impression: 2

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2020930757

ISBN 978–0–19–885836–2

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface

New ideas and concepts in the development of particle detectors have often been the prerequisites for important experiments that led to discoveries or new perceptions in particle and astroparticle physics. Therefore the physical principles and techniques of detector development belong to the basic skills of an experimental physicist in this field. Often novel detector developments also pave the way for advancements in imaging methods using particles or radiation, for example, in medicine, biology or material science.

The physics of detectors interrelates to many other areas of physics and engineering. Developing and operating detectors requires knowledge of the interactions of particles with matter, of the physics of gases, liquids and solids, of the phenomena of charge transport and the formation of signals as well as of the techniques of electronic signal processing and microelectronics.

The idea for this book originated within the framework of a transregional educational project on the subject of detector physics, funded by the German 'Federal Ministry of Education and Research', in which the authors participated. The central elements of the project were lecture series and accompanying manuscripts on detector physics which became the seeds for this book. A German edition was published in 2016 [624]. Major parts have, however, been substantially revised and enlarged.

Originally conceived as a lecture book for advanced undergraduate and graduate students, the book evolved over the years with regard to content and depth of the treated material such that the content now goes beyond that of typical lectures on this topic. The target group for the book are both students who want to get an introduction or wish to deepen their knowledge on the subject as well as lecturers and researchers who want to extend their expertise. In addition, numerous tables and comparative synopses should serve as a reference for scientific work. We have made an effort, and hope to have succeeded, to treat the available abundant material on the various subjects as clearly as possible and as deeply as necessary.

At this point, we want to thank the many colleagues, co-workers and students who in one way or another contributed to the completion of this book. For support in producing the numerous figures we thank David Barney, Axel Hagedorn, Christine Iezzi, Jens Janssen, Manuel Koch, Edgar Kraft, Susann Niedworok, Philip Pütsch, Ludwig Rauch, Marco Vogt and Bert Wiegers. For the simulation of reaction events, cross checks of analytical formulae by simulations, and for assistance with using the GEANT4 program package we thank Timo Karg, Sven Menke, David-Leon Pohl and Yannick Dieter. For numerous discussions, information and/or proofreading of single chapters we want to thank Markus Ackermann, Peter Fischer, Eckhart Fretwurst, Fabian Hügging, Fabian Kislat, Hans Krüger, Teresa Marrodán Undagoitia, Peter Lewis, Michael Moll, Rainer Richter, Ludwig Rauch, Jochen Schwiening and Peter Wegner as well as Ted Masselink for enlightening discussions on the topic of charge transport in solids. We are grateful to Sabine Baer for the support in obtaining copyrights and to Martin Köhler for advice in questions concerning the publication of a

book. For the English edition we obtained information and help in language questions from Summer Blot, Ian Brock, John Kelly, Jakob van Santen and Andrew Taylor.

For the support during the genesis and the realisation of this book we thank the German Federal Ministry of Education and Research and the Helmholtz Research Centre Deutsches Elektronen-Synchrotron (DESY).

In particular, we would like to deeply thank our wives, Marion and Sonja, for their continuous support for our work on the book over almost two decades.

Hermann Kolanoski and Norbert Wermes

Berlin and Bonn, March 2020

Contents

1
Introduction

The visual perception of objects is due to the fact that light, or more generally speaking electromagnetic radiation, interacts with matter. The light is first scattered at the object, is then absorbed in the eye and subsequently transformed into neural signals which are further processed by the brain which finally generates the object image perceived by us. The detection of elementary particles, nuclei and high-energetic electromagnetic radiation, in this book commonly designated as 'particles', proceeds similarly through interactions of the particles with matter. In general, though, we do not directly perceive particles with our sense organs. Instead we need an external 'detector' in which the particles interact and which derives perceptible signals from the interaction.

The *electromagnetic interaction* of particles with matter is by far most frequently employed for detection. Charged particles are detected through the ionisation of the matter along their trajectory or also through emission of electromagnetic radiation like bremsstrahlung or Cherenkov and transition radiation. Photons and electrons develop electromagnetic showers in matter which can be used for energy measurement. The *strong interaction* is exploited, for example, for the detection of neutrons or for the energy determination of high energetic hadrons through the development of hadronic showers. Finally, the *weak interaction* is exploited for the detection of neutrinos.

Depending on the application, particle detectors have to fulfil different tasks with quite different requirements. A simple example is the measurement of the radiation flux with a Geiger counter for the detection of radioactivity. In particle physics experiments one usually wants to measure in addition the particle kinematics (direction, momentum, energy) and preferably also determine the identity of a detected particle.

The requirements on the performance of particle detectors, which are closely related to the costs, vary over a wide range. For the specification of a detector the following criteria come into consideration depending on the intended application purpose:

- high detection probability;
- small perturbation of the process to be measured;
- high signal-to-background ratio;
- good resolution (position, time, energy, momentum, angle, ...);
- fast, deadtime-free electronic signal processing;
- simple online monitoring and control;
- justifiable costs.

The development of detectors and detection methods for particles was mainly driven by applications in basic research, as in particle and nuclear physics. The progress in these fields crucially depends on the state of the detector technology. This has also been recognised by the Nobel Committee which repeatedly awarded the Nobel Prize for decisive breakthroughs in detection methods, such as the cloud chamber:

C.T.R. Wilson 1927, the advancements of the cloud chamber method: P. Blackett 1948, nuclear emulsions: C.F. Powell 1950, the coincidence method: W. Bothe 1958, the bubble chamber: D.A. Glaser 1960 and the multiwire proportional chamber: G. Charpak, 1992. Besides the use of detectors in particle, nuclear and astroparticle physics, there is meanwhile a variety of detector applications in other fields, for example, in medicine, geology, archaeology and material science. The detector sizes vary between a cubic-centimetre, for example, of a dosimeter in the format of a ball-point pen, and cubic-kilometre detectors for the detection of air showers initiated by cosmic rays.

Besides referencing the original literature we have made an effort to also point to literature for further reading, lecture books and compact reviews of the respective fields. Comprehensive introductions into the subject of detectors are the books by Kleinknecht [612] and Grupen and Shwartz [489]. The books by Knoll [616] and Leo [652] are particularly suited to learn about the classical methods of construction and operation of detectors. The applications in modern particle physics experiments are covered, for example, in the books by Leroy and Rancoita [653] and Green [473]. Expert articles related to the different chapters of this book can be found in various collections, for example, the ones issued by Ferbel [397] or by Sauli [850]. A comprehensive collection of many detection methods are the two volumes 'Handbook of particle detection and imaging' edited by Grupen and Buvat [488].

This book originates from lectures which the authors gave repeatedly for the specialisations in 'Experimental Particle Physics' and 'Astroparticle Physics' at the Humboldt University in Berlin and at the Friedrich-Wilhelms University in Bonn. The volume of this book, however, has evolved far beyond what can be presented in such a lecture comprising typically two hours per week for one semester. However, the book should be well suited as a basis for such a lecture, for going more deeply into the subject matter and for getting prepared for instrumental work in particle and astroparticle physics as well as in many fields which are addressed in section 2.4.

Besides the introductory and overview chapters (chapters 1 and 2), the book is divided into five subject areas:

- fundamentals (chapters 3 to 5),
- detection of charged tracks (chapters 6 to 9),
- phenomena and methods mainly for particle identification (chapters 10 to 14),
- energy measurement (accelerator and non-accelerator experiments) (chapters 15 and 16),
- electronics and data acquisition (chapters 17 and 18).

Comprehensive lists of literature, keywords and abbreviations can be found at the end of the book.

2
Overview, history and concepts

2.1 On the history of detectors

Progress in nuclear and particle physics is based on the development of detectors with which particles and radiation can be detected and their properties can be measured (see table 2.1). The discovery of radioactivity by H. Becquerel in 1896 marks the beginning: Becquerel concluded from the observation of blackening of a photographic plate, which he kept in the dark close to a sample of uranium salt, that radiation is coming from the uranium, a phenomenon which was later called radioactivity. Such an integral measurement of radiation (in contrast to the detection of single quanta or particles) was also the basis for the discovery of cosmic radiation by V. Hess in 1912 (Nobel Prize 1936 [520]). During a balloon flight he observed the discharge of an electrometer with increasing altitude which he interpreted as being due to ionising radiation originating from outer space (chapter 16).

In contrast to this indirect evidence for radiation it is important for basic research that particles are individually measured with as many details about their properties and kinematics as possible. Around 1900, photographic plates and scintillating coatings of screens became the first detectors for the newly discovered radiation—besides the α, β and γ radiation from nuclei, also for cathode rays (electron beams) and X-rays. In scattering experiments with α particles, Rutherford, Geiger and Marsden detected the scattered particles on a scintillating zinc sulfide screen (ZnS) thereby determining their scattering angle [841,453]. Figure 2.1 shows the apparatus with which the scintillation flashes on the screen were recorded by eye using a microscope [453].

Also, the photoemulsion technique has been refined further so that the kinematics of single particles of cosmic radiation could be reconstructed by analysing photographic pictures (see section 6.3). Using this method, C.F. Powell and co-workers discovered the pion in 1947 (see fig. 16.2). The Nobel Prize in Physics 1950 was awarded to Powell 'for his development of the photographic method of studying nuclear processes and his discoveries regarding mesons made with this method' [793].

For the electrical recording of single particles H. Geiger developed tubes based on the principle of gas amplification of the ionisation charge in strong electric fields (H. Geiger 1908 [842]). The strong field is obtained by a high voltage applied between

Table 2.1 Some breakthroughs in the history of detector development. Since for such developments precise time specifications are often not possible the specifications of years in the first column are given for orientation only. The assignment of discoveries to persons is also sometimes arbitrary. For example, the principle of gas amplification, which is here associated with the name of Geiger, was essential for the works on the classification of the radioactive radiation for which Rutherford received the Nobel Prize in chemistry in 1908. Some detector principles, like gas amplification, coincidence method and wire chambers, have been so fundamental that it is not possible to link them to specific discoveries. In more recent times it became also increasingly difficult to associate detector developments with single individuals, as for example in the case of microstrip detectors with which important discoveries involving heavy fermions were made.

Year	Name	Detector principle	Discovery	Nobel Prize
1896	H. Becquerel	photographic plate	radioactivity	1903
1908	H. Geiger	gas amplification		
1911	E. Rutherford	scintillation screen	atomic nucleus	
1912	C.T.R. Wilson	cloud chamber	many new particles	1927
1912	V. Hess	electrometer	cosmic rays	1936
1924	W. Bothe	coincidence method		1954
1933	P. Blackett	triggered cloud chamber	e^+e^- pairs	1948
1934	P.A. Cherenkov	Cherenkov radiation	ν oscillation	1958
1947	C.F. Powell	photoemulsion	pion	1950
1953	D.A. Glaser	bubble chamber	Ω^-, neutral currents	1960
1968	G. Charpak	multiwire prop. chamber		1992
1980		Si microstrip detector	$B\overline{B}$ oscillation	

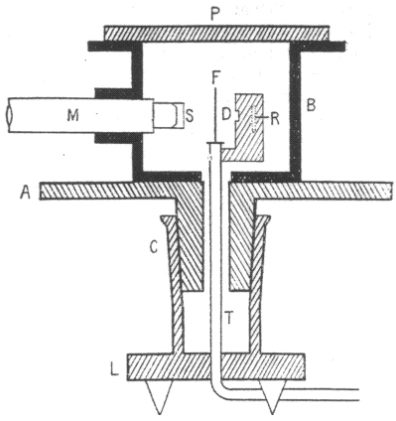
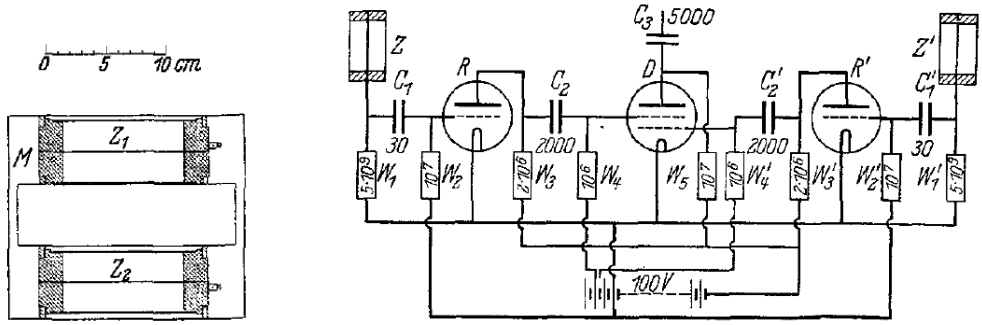


Fig. 2.1 Apparatus for the observation of the scattering of α particles off a gold foil (Rutherford scattering) as described in [453]. An α source R radiates through a thin diaphragm D onto the foil F. The scattered particles are observed on a screen using a microscope whose objective carries a small scintillation screen S. The microscope is rigidly connected to the box B which is closed by the plate P and evacuated through the tube T. The scattering angle is set by turning the box with the microscope relatively to foil and source (the base plate A turns in the mount C fixed to the base plate L). Reprint from [453], with kind permission of Taylor & Francis Ltd.

the wall of a cylinder (cathode) and a thin wire (anode) which is strung along the cylinder axis, see section 7.2.2. Gas amplification by secondary ionisation leading to the development of avalanches in gases [943,944] had already been studied by J. Townsend since about 1900 (see references in [842]). These developments led to the radiation monitor known as the Geiger–Müller tube [454] (often just called 'Geiger counter', in particular when referring to the ready-to-use instrument) and later to the 'proportional counter' which employs also gas amplification but in a more modest operation mode



(a) (b)

Fig. 2.2 First applications of the coincidence method: (a) Experimental set-up for the investigation of the properties of cosmic rays (in German 'Das Wesen der Höhenstrahlung' by Bothe and Kolhörster 1929 [239]). In order to test different absorber materials, the signals of counter tubes above and below the absorber are recorded when they occur in coincidence. (b) Vacuum tube circuit for the detection of coincidences of the two counters Z and Z' [238]. The bias voltage of the tube D in the middle is adjusted such that a signal is only generated at the output capacitor C_3 if the potentials of both grids are simultaneously high. The grids are controlled by the signals of both counters via the symmetric circuits on the left and right. These circuits amplify and differentiate the counter signals to yield sufficiently high, short pulses. The separation accuracy of this circuit is about 1 ms. In the original application of the circuit the output signal at C_3 was transferred to a unit counter of a telephone. Reprint of the pictures with the kind permission of Springer Science and Business Media.

(see section 7.4). At the beginning the amplified charges were recorded by eye through the deflection of an electrometer needle. However, the generation of electric signals with a detector cleared the way for the automatic recording and finally for electronic data processing in general.

A first fundamental step towards the electronic recording of events was the development of the 'coincidence method' in the 1920s by W. Bothe and others (Nobel Prize 1954 [237]). The method employed the temporal correlation of the occurrence of events to draw conclusions on the underlying physical process. For example, the penetration ability of radiation could be measured for individual quanta or particles by observing signals simultaneously above and below an absorber block (fig. 2.2(a)). Based on this method coincident signals can be used to generate 'triggers' selecting interesting events which can then be automatically recorded. The first coincidence circuit was realised by Bothe in 1929 using vacuum tube electronics [237] (fig. 2.2(b)). However, before the development of transistor electronics starting in the 1950s, the potential of such circuits was quite limited as compared to today's standards.

Until the 1980s, without the availability of highly integrated electronics, optical recording by photographic exposure remained the only choice to store pictures of complex reactions in large detector volumes. The first detector type with which particle reactions could be made visible was the cloud chamber (chapter 6). In cloud chambers ionisation along particle trajectories becomes visible as a 'condensation trail' in a supersaturated gas. As an example, fig. 2.3 shows the renowned cloud chamber picture of the discovery of the first antiparticle, the positron, by Anderson in 1932 [94]. The determination of the charge sign of the particle by observing the track curvature in

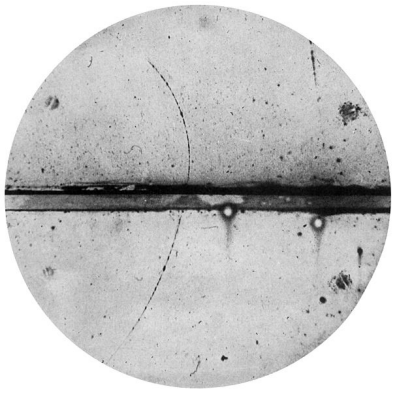


Fig. 2.3 Discovery of the positron by C.D. Anderson in 1932 [94]. The picture shows a particle track in a cloud chamber passing a 6 mm thick lead plate. The curvatures of the two track segments due to the applied magnetic field of 1.5 T indicate that the particle came from below and lost energy in the plate. Flight direction, curvatures and track length are compatible with belonging to a positively charged particle with the mass of the electron; a proton would have a ten-times shorter range. From Wikimedia Commons (original in [94]).

a magnet field was decisive for the identification as antiparticle. The cloud chamber was developed around 1900 by C.T.R. Wilson and was recognised by the Nobel Prize 1927 [998]. The full capability of the apparatus only became available after P. Blackett had automated the sequence of chamber expansion and picture taking with the aid of a selective trigger. For the advancement of the 'cloud chamber method' Blackett won the Nobel Prize in 1948 [207].

Around 1950 the bubble chamber took over the leading role as a large-volume detector with which particle reactions could be recorded quite completely (fig. 2.4). The bubble chamber works according to a similar principle as the cloud chamber but rather with bubble formation in a superheated liquid which could be photographed. For this development D.A. Glaser received the Nobel Prize in 1960 [462]. The bubble chamber picture in fig. 2.4 illustrates the discovery of the Ω^- baryon which, consisting of three strange quarks, was pivotal for the confirmation of the quark model. Cloud and bubble chambers are mostly operated in a magnetic field such that momentum

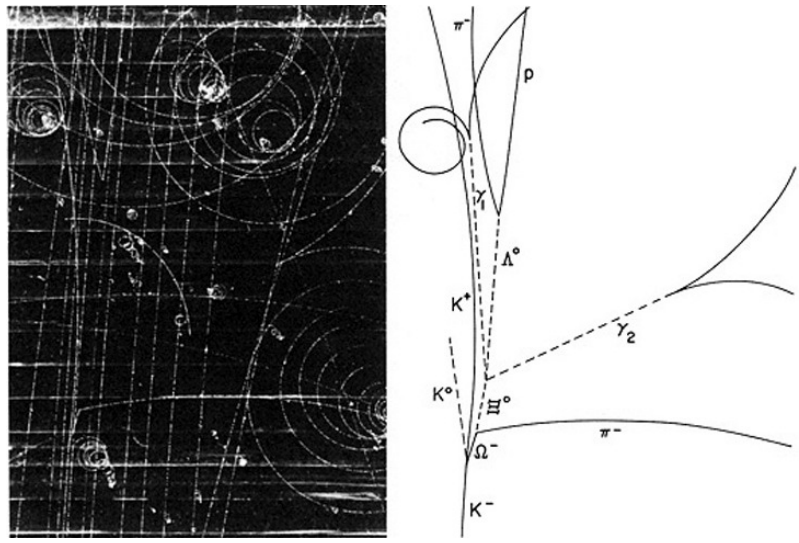


Fig. 2.4 Visualisation of particle reactions. Bubble chamber picture of the discovery of the Ω^- baryon [156] (courtesy Brookhaven National Laboratory).

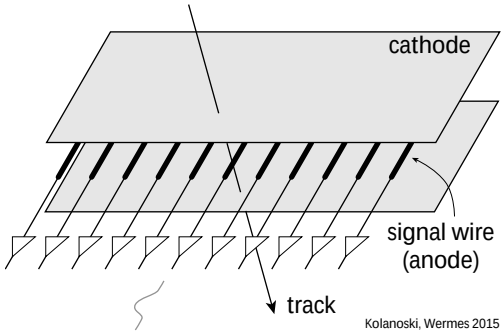


Fig. 2.5 Multiwire proportional chamber. Each wire corresponds to a single proportional counter tube, but with planar cathodes below and above the wires. The signal from each wire is readout separately.

and charge sign can be determined from the track's curvature.

The transition to detectors with completely electronic readout proceeded since the 1960s in parallel to the progressing development of highly integrated electronics. The first step in this direction was the development of the multiwire proportional chamber (MWPC) by G. Charpak (Nobel Prize 1992 [290]). The MWPC is in principle a planar side-by-side arrangement of counter tubes (fig. 2.5). In the 1980s, this detector type and variants thereof, in particular drift chambers, finally replaced bubble chambers (chapter 7). Figure 2.6 shows an example for the display of an event in a drift chamber. In experiments with high radiation levels, as at the LHC (see table 2.2), mostly semiconductor detectors (chapter 8) instead of gas-filled wire chambers are employed. Semiconductor detectors, based on silicon or germanium, had already been employed in nuclear physics for high-resolution spectroscopy since the 1960s. In particle physics spatially resolving silicon detectors were introduced in the 1980s for precise measurements of decay vertices, especially of heavy fermions. The required accurate spatial resolution became feasible by micro-structuring the electrodes, adapting methods from the rapidly developing microelectronics. With the progress in the miniaturisation of electronics, semiconductor detectors became more affordable, allowing also their use as large volume tracking detectors, in particular in high rate environments (e.g. at LHC).

So far we have discussed developments which mainly concern the reconstruction of ionisation tracks of charged particles. Another development line is detectors for the

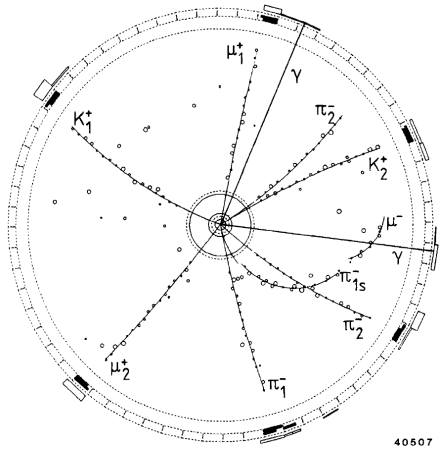


Fig. 2.6 Visualisation of particle reactions. Collision event recorded with the cylindrical drift chamber of the ARGUS experiment (cut perpendicular to the colliding electron-positron beams) visualising the decay of an $\Upsilon(4s)$ resonance with two identical B mesons in the final state (source: DESY). It is one of the completely reconstructed events of the type $\Upsilon(4s) \to B^0\overline{B^0} \to B^0B^0$ (or : $\overline{B^0B^0}$) in which unambiguously a particle has turned into its antiparticle or vice-versa. Such events are the basis for the discovery of matter-antimatter mixing in systems with bottom quarks [74].

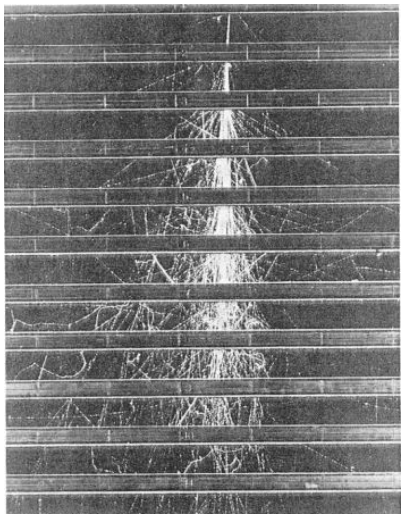


Fig. 2.7 Cloud chamber picture of an electromagnetic shower which was initiated by a high-energetic photon or electron. The detector exhibits a sandwich architecture. The absorber layers are brass plates; the shower particles can be observed in the gaps between the plates. Source: 'MIT Cosmic Ray Group', published in [836].

measurement of particle energies, which are called calorimeters because their principle is based on the total absorption of a particle's energy. In dense materials the particles deposit their energy by generating particle cascades (chapter 15). Initially, calorimeters were developed for electromagnetically interacting particles, like electrons and photons. Electromagnetic showers had already been observed and measured in cloud chambers, see fig. 2.7. Since the 1940s, particularly good energy resolutions were obtained with scintillating crystals (see chapter 13). However, the production of calorimeter crystals is quite demanding, in particular for large quantities, and is relatively expensive. A compromise between resolution and cost efficiency is found by employing *sandwich calorimeters* which are built from alternating layers of absorber material, such as lead, and layers of readout detectors, such as scintillators. The cloud chamber picture in fig. 2.7 shows the principle of a sandwich calorimeter (we refer here to a cloud chamber just because there the shower development is particularly well visible).

The accuracy of momentum measurements at high energies is limited by the achievable strength of magnetic fields. Therefore, the continuous growth of accelerator energies and thus of the average particle energies requires that also the energies of hadrons have to be measured calorimetrically. Besides the detection of charged tracks, hadron calorimetry became a key aspect of today's detector development.

The discovery of the Cherenkov effect by P. A. Cherenkov in 1934 (Nobel Prize 1958 [302]) was not related to a specific detector development. But it was soon realised that the velocity dependence of the effect combined with a momentum measurement can be exploited for a mass determination and thus usually for the identification of a particle (chapter 14). The ability to identify particles and to determine at least some quantum numbers was crucial for the development of the particle systematics of the Standard Model. Beyond the aspect of particle identification, Cherenkov radiation offers a cost-effective possibility to instrument large detector volumes, as they are often required in astroparticle physics (chapter 16).

The detection of Cherenkov and scintillation light requires detectors which are sensitive to few photons. Since the 1930s single photon sensitivity was achieved by photomultiplier tubes (PMTs) which exploit the photoeffect in combination with the effect of secondary electron emission (section 10.2.1) [1006]. A PMT basically works

as follows: photons hitting a thin metal foil generate electrons which are accelerated to produce secondary electrons; after several acceleration steps the electrons are sufficiently amplified to deliver a measurable output signal, even for single photons. A PMT in combination with a scintillation counter was employed as a particle detector for the first time in 1944 [652]. Only in combination with a PMT did scintillation detectors find their place alongside counter tubes, cloud chambers and photoemulsions.

Primarily, detectors have been developed for the measurement of particles in nuclear and particle physics experiments. In such experiments, particles are mostly scattered off each other. From the observation of the scattering products one tries to gain insight into the properties of the participating particles and their interactions[1].

In the beginning the only available particle sources were radioactive nuclear decays and cosmic radiation. A scattering experiment took place using a stationary 'target'. For example, in the famous Rutherford scattering experiment α particles were scattered off a gold foil (fig. 2.1); in the case of cosmic radiation the molecules of the atmosphere served as the target. When particle accelerators became available more target-oriented experiments could be performed at higher intensities. The development of accelerators began in the 1920s[2] and since the beginning of the 1930s they were also employed for nuclear physics experiments. In particle physics, accelerators had their breakthrough only after the Second World War. The turning point is marked by the discovery of the antiproton at the proton synchrotron Bevatron in Berkeley with an energy of 6.2 GeV [288]. The beam energy was in the first place chosen such that antiprotons could be produced. Until then, all new members of the steadily growing 'particle zoo'—positrons, muons, pions, kaons, hyperons (Λ, Σ, Ξ)—had been discovered in the cosmic radiation with the help of cloud chambers, photoemulsions and Geiger–Müller counters which were brought to mountains or by balloons to high altitudes.

The most important development lines of detectors have now been sketched where we have, however, omitted the many interesting developments for special applications. The progress was always driven by the demands at the frontier of science, wherever it was at the respective time. For accelerator-based experiments the general trend always went towards higher energies and higher intensities allowing us to study higher mass scales and rarer processes. For astrophysical experiments the energy spectrum and intensity is given and cannot be changed so that the scientific reach can only be extended by making the detectors bigger and more sensitive.

Currently, the requirements of the experiments at the Large Hadron Collider (LHC) at CERN (Geneva) are stimulating the detector developments. In the LHC protons in two oppositely directed beams with the currently highest energies of up to 7 TeV are scattered off each other. The prominent challenges for the detector technology are the very high reaction rates and particle multiplicities as well as the resulting high radiation loads. In recent years, important progress in electronic data acquisition, processing and storage was achieved (chapter 18). Data are selected on computer farms with hundreds to thousands of processors, which are fed by data from up to 100 million readout channels. The yield of stored data reaches petabytes per year corresponding to several 100 megabytes per second. An end of this development is not yet in sight.

---

[1] In this book we try to presume as little knowledge on nuclear, particle and astroparticle physics as possible. Some background in these fields is certainly helpful to better understand the motivations for the detector developments. Lecture books on these subjects are for example the books by D. Perkins [772] and A. Bettini [190].

[2] Although the development of accelerators is closely related to the subject 'detectors' we do not want to dwell on it but refer to the respective literature, see for example [994, 999].

2.2 Detectors at accelerators

The foundations of particle physics have been laid by observing elementary reactions in the cosmic radiation. Since the 1950s beams of accelerated particles became available in the laboratory. The advantages of cosmic radiation, like general availability and highest energies, are often outweighed by the possibility offered by accelerators to carry out controllable, high-rate experiments. Also, the energy frontier has been steadily moved upwards.

2.2.1 Accelerators

Stable charged particles can be accelerated by electric fields while magnetic fields can keep them within compact beams on desired orbits [994, 999]. A particle with the elementary charge e gains in an electric field E over the path length L the kinetic energy

$$T = e \int_0^L E \, ds = eV \,, \tag{2.1}$$

where V is the potential difference between 0 and L. Therefore, in nuclear and particle physics the energy is quoted in units of electronvolt (eV):

$$1\,\mathrm{eV} \approx 1.602 \times 10^{-19}\,\mathrm{C\,V} = 1.602 \times 10^{-19}\,\mathrm{J}\,.$$

Since DC voltages (DC = direct current) in accelerators are limited to about 1–10 MV, the accelerators in high energy physics operate with high frequency electromagnetic fields which are stored in periodic structures. The particles are accelerated in short bunches when they pass the structure in phase with the accelerating field. The bunch structure of the beams affects the design of the time structure of the data acquisition systems of the experiments (chapter 18).

The high-frequency structures could be linearly (linear accelerator) or circularly (cyclotron, synchrotron) arranged (fig. 2.8). A prominent example for an electron linear accelerator is the machine at the 'Stanford Linear Accelerator Center' (SLAC) built in the 1960s for the measurement of the structure of nucleons (deep-inelastic scattering). Except for this one, all other high energy accelerators have been designed as synchrotrons, either for electrons or for protons. A synchrotron offers the possibility of 'strong focusing' by applying inhomogeneous magnetic fields to constrain the beam size in both transverse directions (see e.g. [999]). With the introduction of strong focusing around 1950, the preconditions were given for the construction of the proton synchrotrons in the 30 GeV domain at CERN and at Brookhaven. Currently the largest synchrotron complex is the proton–proton collider at CERN, named Large Hadron Collider (LHC). This position was held for a long time by the proton–antiproton collider Tevatron at the Fermi National Accelerator Laboratory (FNAL, Fermilab) in the USA before the LHC started operation.

The classical synchrotron generates a particle beam which is sent onto a fixed target. The target can be positioned internally inside the machine in the fringe region of the beam or externally in an ejected beam. There are accelerators for electrons, protons and their antiparticles as well as for heavy ions. In addition secondary beams of practically all sufficiently long-lived particles (photons, pions, kaons, hyperons, muons, neutrinos) can be generated by showering the primary beam in a dense conversion target followed by a filter selecting the desired particles. Hadron beams are generated by primary protons, photon beams from the bremsstrahlung of electrons and muon

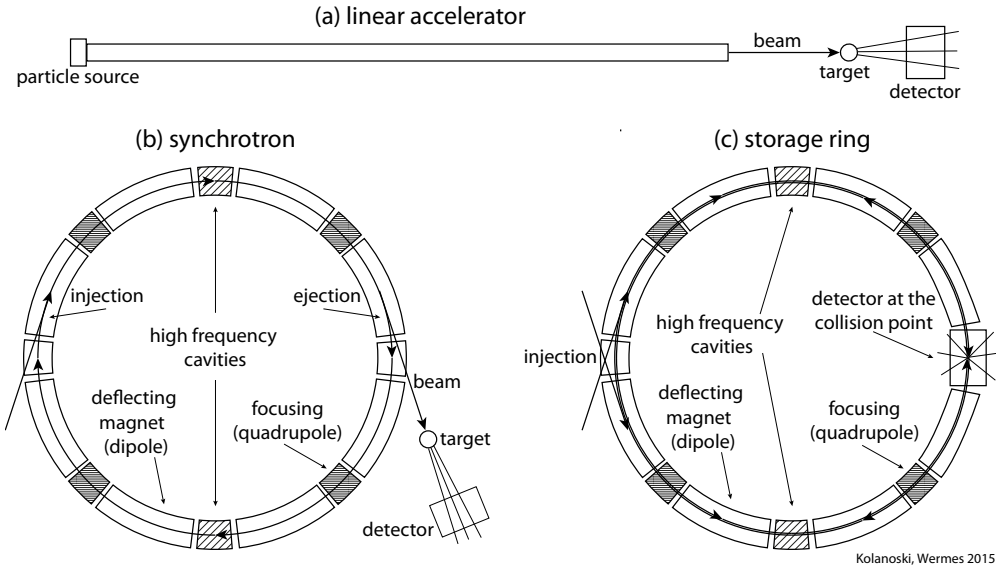


Fig. 2.8 Schematic sketches of three important accelerator types. The first classification separates into linear accelerators (a) and circular accelerators (b, c). Under the circular accelerators one distinguishes the scheme of a synchrotron (b) with a beam in the machine and an ejected external beam and the scheme of a storage ring (c) with two oppositely circulating particle beams which are collided at a crossing point.

and neutrino beams from decays of secondary pions and kaons. Secondary beams are also used for testing and calibrating detectors. At CERN the highest energy external beams are delivered by the 'Super Proton Synchrotron' (SPS, up to 450 GeV). The Fermilab 'Main Injector', a proton synchrotron, provides beams up to 120 GeV .

Since in a stationary target a large portion of the beam particle's energy is lost as recoil energy, the idea of 'colliders' emerged where particle bunches are accelerated in opposite directions and are collided at specific interaction points (fig. 2.8(c)). Colliders are usually realised as rings, called storage rings, in which particles are kept as long as possible in circular orbits. An exception thereof is the 'Stanford Linear Collider' (SLC, table 2.2). The experiments at LEP (CERN, see table 2.2) and the SLC have studied the Z-boson resonance in electron–positron collisions. Because in circular accelerators electrons lose their energy by radiation (synchrotron radiation) for which the emitted power strongly increases with energy, future high-energy electron–positron colliders are planned as linear colliders as well [555, 314]. Colliders have been built for the following particle combinations, see table 2.2: electron–positron, electron–electron, proton–proton, proton–antiproton, proton–electron, proton–positron.

Characteristic parameters of colliders are, amongst others, the energy of the beam particles, the time structure of the particle bunches, the spread of the interaction zone and the luminosity. The luminosity L is a measure for the interaction rate at a crossing point of the beams and is defined as the reaction rate $\dot{N}$ per cross section σ of the reaction (see section 3.1):

$$\dot{N} = L\,\sigma \qquad \text{or} \qquad L = \frac{\dot{N}}{\sigma}\,. \tag{2.2}$$

The luminosity is proportional to the particle currents and inversely proportional to the overlap area of the beams at the crossing point. In the simplest case, for equal beam profiles following two-dimensional Gaussian distributions with widths σ_x and σ_y (in the plane of the machine and perpendicular thereof) one obtains:

$$L = \frac{b\, N_1 N_2\, f}{4\pi\, \sigma_x \sigma_y}\,, \tag{2.3}$$

where b is the number of bunches per beam which circulate with the frequency f and contain N_1 and N_2 particles, respectively [994]. The beam profiles depend on parameters of the beam optics which is defined by the bending and focusing magnets of the accelerator. Accordingly the luminosity can be determined by measuring the beam parameters, for example, by moving the beams in the collision zone relative to each other and recording the corresponding rates ('Van-der-Meer scan' [154]). Since this scan cannot be executed during the running of an experiment, for a continuous determination of the luminosity particle rates are measured by a so-called luminosity monitor and converted into a luminosity according to formula (2.2). The cross section entering in the formula is either known—for example for Bhabha scattering at small angles in electron–positron colliders—or the rate was calibrated according to (2.3) by means of the Van-der-Meer Scan and measurements of the currents using induction coils. The highest achieved collider luminosities lie at $L \approx 10^{34-35}\ \mathrm{s}^{-1}\,\mathrm{cm}^{-2}$, meaning that for a cross section of $\sigma = 1\,\mathrm{fb} = 10^{-39}\,\mathrm{cm}^2$ (fb = femtobarn) up to 10 events will be produced per year.

2.2.2 Detector concepts

Particle detectors at accelerators are usually 'magnet spectrometers', meaning that they include one or several magnets and tracking detectors with which the momenta of charged particles can be determined by measuring the curvature of their trajectories (chapter 9). The first experiments at accelerators were of the 'fixed-target' type, which means that the beam is shot on a stationary target. The detectors searching for the reaction products were restricted to detect single particles in a small solid angle of a few millisteradian and in a limited momentum window. Figure 2.9 shows as an example a spectrometer for the measurement of electrons. This apparatus was employed in the classical experiment of deep-inelastic scattering at the linear accelerator SLAC where the nucleon structure functions have been measured [935]. Besides the dipole magnets for momentum analysis such a spectrometer also comprises focusing quadrupole magnets for mapping, for example, different momenta independently of the scattering angle onto the focal plane. A hit in a counter of a counter array in the focal plane yields the scattered energy while the spectrometer direction relative to the incoming beam yields the medium scattering angle. Thus with only a relatively small number of detector elements and without explicitly reconstructing the complete particle trajectory, it is possible to determine the particle's kinematics.

At the same time, since the 1950s complete events could also be made visible with bubble chambers, see for example fig. 2.4. Certainly, the analysis of the huge number of photographs, which mostly contained nothing interesting, was very laborious. As sketched in the historical section 2.1 the miniaturisation of semiconductor electronics allowed for equipping ever larger detector volumes with ever finer segmentation of the electronic readout. Therefore, the goal in modern particle physics experiments is often to detect, identify and kinematically reconstruct all particles of a scattering reaction

Table 2.2 List of colliders with centre-of-mass energies (E_{cm}) above 8 GeV (this limit is arbitrary, but ensures that the machines with most of the experiments mentioned in this book as examples are included). For the heavy ion collider RHIC and for heavy ions in LHC the energies are given per nucleon (N = number of nucleons in an accelerated nucleus). More details about colliders can be found in the review articles 'High Energy Collider Parameters' in the 'Review of Particle Properties' of 2014 and 1996 issued by the 'Particle Data Group' (PDG) [746, 157].

Name	Status	Location	Particles	E_{beam} (GeV)	E_{cm} (GeV)
ISR	1971–1984	CERN (Geneva)	$p\,p$	35 + 35	70 GeV
SPEAR	1972–1990	Stanford (USA)	$e^+\,e^-$	4 + 4	8 GeV
DORIS	1973–1993	DESY (Hamburg)	$e^+\,e^-$	5.6 + 5.6	11.2
CESR	1979–2002	Cornell (USA)	$e^+\,e^-$	6 + 6	12
VEPP-4M	1994–	Novosibirsk (Russia)	$e^+\,e^-$	6 + 6	12
PETRA	1978–1986	DESY (Hamburg)	$e^+\,e^-$	20 + 20	40
PEP	1980–1990	Stanford (USA)	$e^+\,e^-$	15 + 15	30
TRISTAN	1987–1995	KEK (Japan)	$e^+\,e^-$	32 + 32	74
SpS̄pS	1981–1990	CERN (Geneva)	$p\,\bar{p}$	250 + 250	500
PEP II	1999–2008	Stanford (USA)	$e^+\,e^-$	3.1 + 9.0	10.58
KEK-B	1999–2010	KEK (Japan)	$e^+\,e^-$	3.5 + 8.0	10.58
LEP	1989–2000	CERN (Geneva)	$e^+\,e^-$	100 + 100	200
SLC	1989–1998	Stanford (USA)	$e^+\,e^-$	50 + 50	100
Tevatron	1987–2011	Fermilab (USA)	$p\,\bar{p}$	1000 + 1000	2 000
HERA	1992–2007	DESY (Hamburg)	$e\,p$	30 + 920	330
LHC	2008–	CERN (Geneva)	$p\,p$	7000 + 7000	14 000
LHC	2008–	CERN (Geneva)	$A\,A$	2500/N + 2500/N	5000/NN
RHIC	2001–	Brookhaven (USA)	$A\,A$	100/N + 100/N	200/NN
ILC	planned	not decided	$e^+\,e^-$	250 + 250	500

and to achieve that in as much of the full 4π solid angle as possible. This has natural limitations which we will point out at several places in this book.

In the following we sketch concepts of typical 'general-purpose detectors' while variants and concepts for other applications will be discussed in the corresponding sections of the book. Detector configurations differ in whether they are placed behind a stationary target (fixed target) or in the interaction zone of a collider.

The *fixed-target experiments* have forward detectors, meaning that they cover only a part of the solid angle in the forward direction of the beam (fig. 2.10(a)). The reaction products preferentially go into this direction because of the Lorentz boost which the beam particle transfers to the centre-of-mass system of the reaction. The angular acceptance of the detector can be adjusted such that nearly the full 4π solid angle in the centre-of-mass system is covered.

For *collider experiments* with symmetric beam energies the laboratory and the centre-of-mass system are the same so that the detector is best arranged spherically around the interaction point (fig. 2.10(b)). For reasons which are related to the mechanical construction, the configuration usually becomes cylindrical with the beams as the symmetry axis. Since the entrance and exit for the beams has to be left open, a full solid angle coverage cannot be reached but rather more like 95–99% thereof. A her-

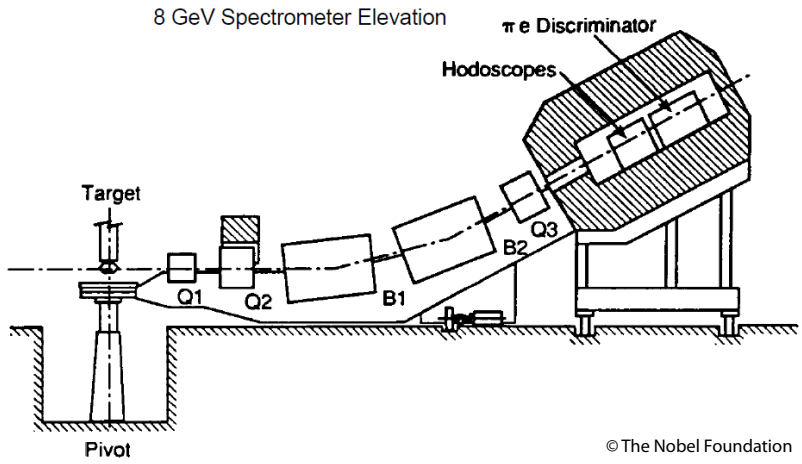


Fig. 2.9 Example of an imaging magnet spectrometer. The 8 GeV spectrometer was one of three spectrometers with different momentum acceptances with which the structure functions of the nucleons were measured for the first time [935]. From these structure functions evidence for quarks inside the proton was concluded. The structure functions are measured by inelastically scattering of electrons off nucleons (protons or neutrons). The spectrometer only measures the scattered electrons (without detecting the recoil system) which is sufficient to determine the relevant kinematical variables. The figure shows a side view of the spectrometer which deflects the electrons upwards. The plane of the view contains the axis through the target around which the spectrometer can be rotated to define the scattering angle. The scattered electrons are deflected by two dipole magnets (B1, B2) and are focused by three quadrupole magnets (Q1–Q3) such that the momenta can be determined by an array of scintillation counters ('hodoscope') arranged in the focal plane of the magnet system. The shielded housing which surrounds the hodoscope also contains trigger counters and elements for the identification of the electrons, in particular for separating them from pions. Source: Nobel Foundation.

metic coverage is particularly important when searching for new phenomena involving invisible (non-interacting) particles.

General-purpose detectors which are striving for a complete detection of all reaction products are typically constructed in layers (planar layers in the fixed-target case and shell layers in the collider case) each taking over a specific task, see fig. 2.10. The detector layers closest to the target or interaction point should be as 'thin' as possible in order to minimise scattering and absorption of particles. In fig. 2.10 a typical layer sequence is sketched for fixed-target and collider configurations:

(1) Vertex detectors for the detection of secondary vertices with separation from the primary vertex of the reaction.
(2) Tracking detectors for track reconstruction yielding momentum determination and (sometimes) particle identification. They can be inside (usually at colliders) or in front and behind a magnetic field.
(3) Optional detectors for particle identification, for example, Cherenkov detectors.
(4) A calorimeter for the measurement of electromagnetic showers (electrons and photons).

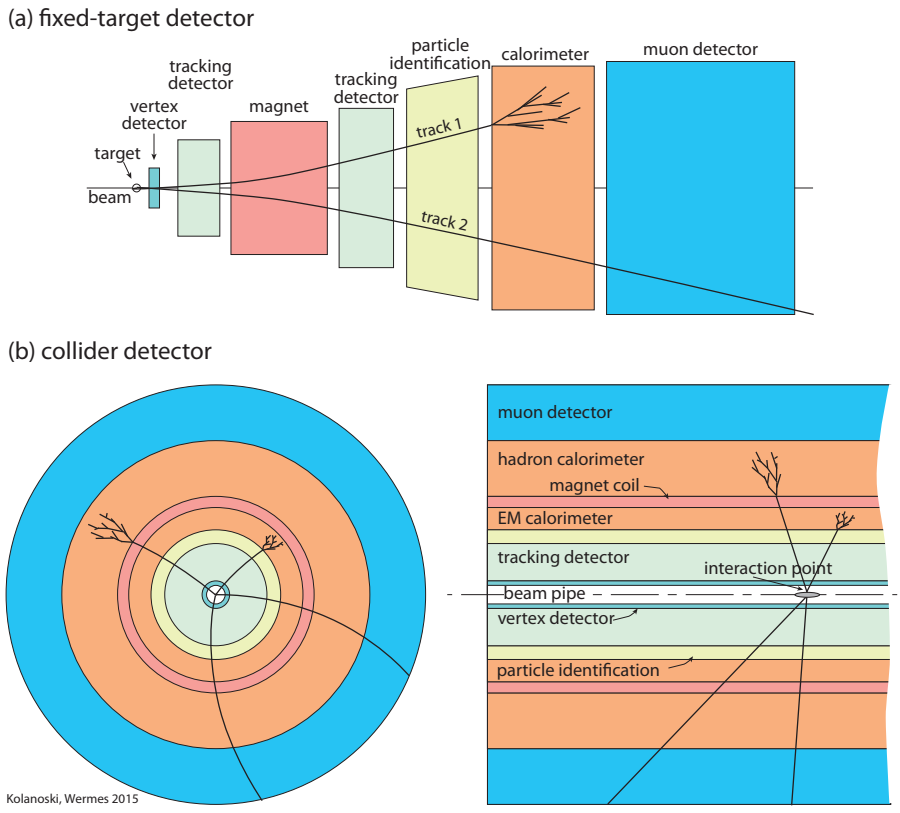


Fig. 2.10 Typical configuration of detector components in (a) fixed-target (side view parallel to the beam) and (b) collider experiments (cuts perpendicular and parallel to the beams). The collider detector is composed of cylindrical shells around the interaction point. Explanations in the text.

(5) A magnet coil in collider experiments; sometimes also in front of the electromagnetic calorimeter.
(6) A hadron calorimeter, usually outside the magnet coil.
(7) Muons are the most penetrating charged particles and therefore they are identified by their detection behind the calorimeters. In collider experiments the muon detectors are often layers of tracking devices installed between iron blocks which at the same time serve as magnet yoke, as in fig. 2.11.

As an example, fig. 2.11 shows a slice of the CMS detector at the LHC. The actual dimensions of the components can be inferred from the scale on the picture. For the CMS detector the high magnetic field of 4 T is characteristic (rather typical is ≤ 2 T). It is also unusual that the coil of the solenoid field (axis parallel to the beams, i.e. perpendicular to the shown view) is placed outside the two calorimeters, usually at least the hadron calorimeter lies outside the coil. The magnetic flux of the solenoid magnet is returned through an iron yoke outside of the coil. The iron is subdivided into several layers between which tracking detectors for the measurement of high-energetic muons are installed.

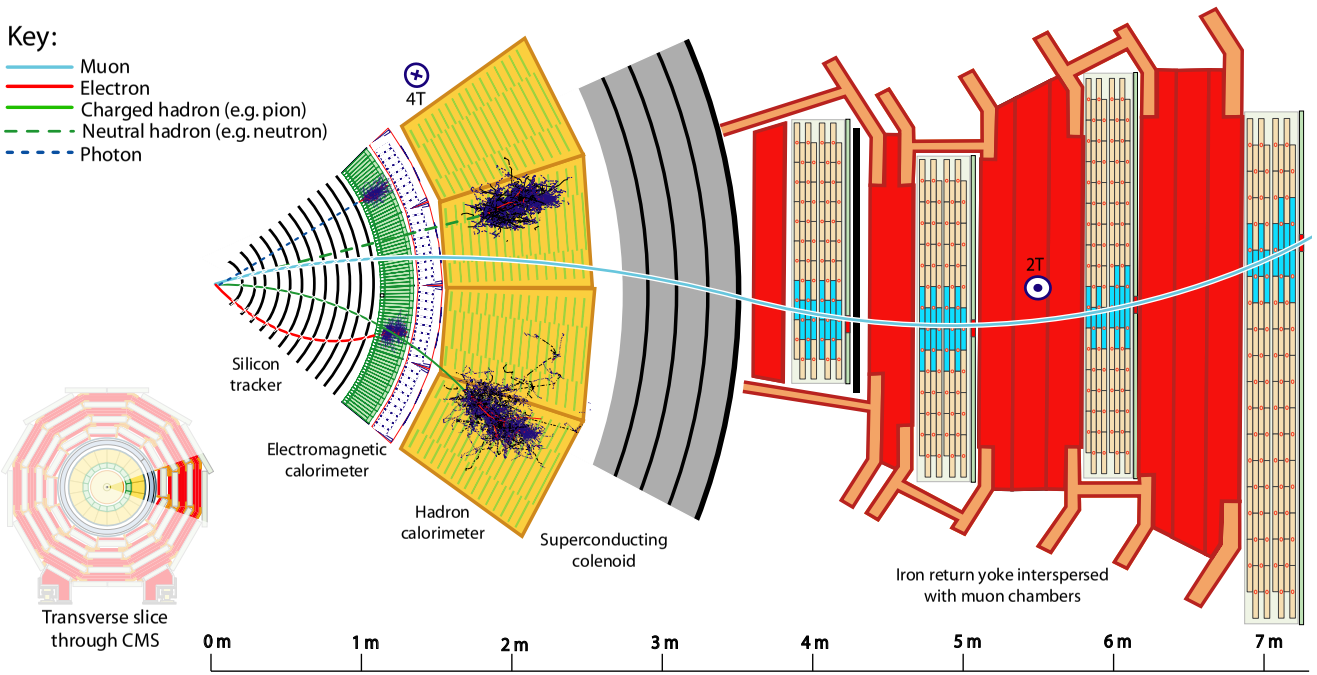


Fig. 2.11 Slice of the CMS detector in the projection perpendicular to the beams (source: CERN/CMS Collaboration). The detector has a layer structure as explained for fig. 2.10. In the innermost layer all charged tracks are detected. Due to a magnetic field which is oriented perpendicular to the projection plane the tracks are bent proportionally to $1/p$ thus providing a momentum measurement. In the following layer, the electromagnetic calorimeter, the electrons lose nearly all their energy, which can be exploited for energy measurements. Similar energy deposits are left by photons in the calorimeter, but they can be distinguished from electrons because they do not create hits in the tracking system. The stable charged and neutral hadrons lose their energy preferentially in the following hadron calorimeter where their energy can be determined. The following layer is the superconducting coil which generates a solenoid field of 4 T inside and an oppositely oriented magnetic field of 2 T outside in the iron. Muons are the only charged particles that can penetrate all these material layers up to behind the coil. Direction and momentum of the muons are measured by tracking chambers between the iron layers of the magnet's return yoke.

2.3 Detectors in astroparticle physics

2.3.1 Detector applications

The research field of astroparticle physics is the detection of radiation and particles originating from outer space in a wide range of energies, from solar neutrinos of some 100 keV to cosmic rays of more than 10^{21} eV (chapter 16). Accordingly, the respective detector technologies are very different, and there is no detector configuration which would be typical for a majority of experiments. However, similarities can be found amongst detectors in the following categories:

- Balloon and satellite experiments (section 16.2): The detectors must be particularly compact and light. They usually have a layer structure similar to detectors at accelerators (fig. 2.10), with specific emphasis on components for particle identification.
- Air shower detectors (section 16.4): These are in principle calorimeters which measure cosmic ray initiated showers only in one plane (the Earth's surface) and with a coarse sampling. The high primary energy requires the instrumentation of very large areas and thus the optimisation of costs per detector unit.
- Cherenkov telescopes for the measurement of TeV gamma radiation (section 16.5): These are mostly imaging telescopes with high time resolution of the cameras for the suppression of ambient background light. Stereoscopic imaging using an array of several telescopes yields an accurate determination of the gamma ray direction.
- Neutrino detectors for low energies (sun, supernovae, ...) (section 16.6): Because of the low detection probability the detector mass has to be maximised and the background radiation has to be minimised. In order to suppress background, laboratories are set up in mountain tunnels and deep mines.
- Detectors for high energy neutrinos of cosmic origin (section 16.6.5): Because of the low reaction probability and low flux of the neutrinos, these detectors have particularly large volumes, up to cubic-kilometres. All of the so far constructed detectors observe Cherenkov light produced by the neutrino reaction products in water or ice.
- Instruments for the detection of exotic particles (dark matter, monopoles, Majorana neutrinos, ...) (section 16.7): The detection methods are very different for different searches. Since in general the background causes the largest problems, the experiments are carried out in underground laboratories as in the case of neutrino experiments.

2.3.2 Observatories for astroparticle physics

The most important observatories for the measurement of air showers that are initiated by charged cosmic particles (section 16.4) and high-energetic photons (section 16.5) are listed in the following:

- Pierre Auger Observatory: Detection of cosmic radiation at the highest energies over an area of about 3000 km in the Pampa Amarilla near Malargüe, Argentina.
- Telescope Array Observatory: Detection of cosmic radiation at the highest energies on an area of about 762 km in the high altitude desert in Millard County, Utah (USA).
- IceCube Neutrino Observatory: 1-km^3-sized Cherenkov detector about 2 km deep in the ice at the geographic South Pole.

- Tunka EAS Cherenkov Light Array: An about 1-km^2-sized air shower detector in the Tunka valley near Lake Baikal, detects Cherenkov light generated in the atmosphere.
- Kaskade-Grande: 0.5-km^2-sized detector at Karlsruhe Institute of Technology (KIT) for the measurement of air showers.
- Imaging Cherenkov telescopes for TeV gamma astronomy: H.E.S.S. (Namibia), MAGIC on La Palma (Spain), VERITAS (Arizona, USA).
- High-Altitude Water Cherenkov (HAWC) Gamma-Ray Observatory: Non-imaging Cherenkov detector at Sierra Negra Volcano near Puebla, Mexico, at an altitude of 4100 m.
- YBJ International Cosmic Ray Observatory: The facility in the Yangbajing valley (YBJ) in the Tibetan highlands at an altitude of 4300 m comprises several experiments, including the air shower experiment ARGO-YBJ for the detection of charged cosmic rays and photons.

2.3.3 Underground laboratories

Prominent underground laboratories are:

- Baksan Neutrino Observatory (BNO): located in the Baksan River gorge in the Caucasus mountains in Russia, 3500 m deep; includes the SAGE Experiment (Soviet-American Gallium Experiment for solar neutrino detection).
- China Jinping Underground Laboratory (CJPL): located in the Jinping Mountains of Sichuan, China, at a depth of 2400 m or 6720 m we (water overburden); houses the dark matter experiment PandaX.
- Kamioka Observatory, Japan: about 1 km deep, observation of oscillations of neutrinos from the sun and the atmosphere, includes the neutrino experiments Super-Kamiokande, KamLand and T2K.
- Laboratori Nazionali del Gran Sasso (LNGS), Italy: one of the largest underground laboratories for particle physics; 1400 m deep in a tunnel through the Gran Sasso massif, includes neutrino experiments, experiments searching for dark matter and double-beta decay.
- Laboratoire Souterrain de Modane (LSM), France: in the Frejus tunnel at a depth of 1700 m (4800 m we); includes the experiments NEMO (double-beta decay) and EDELWEISS (dark matter search).
- Sanford Underground Research Facility (SURF), Homestake Mine, USA: in the Homestake Mine first indications for neutrino oscillations were found with the 'Davis Experiment'; contains now experiments searching for dark matter (LUX) and double-beta decay.
- Soudan Underground Laboratory: located in the Soudan Mine USA, 714 m deep; it accommodates the experiments MINOS (running until 2016) which is positioned in a neutrino beam from the 735-km-distant Fermilab (long-baseline neutrino experiments) and the CDMS Experiment searching for dark matter.
- SNOLAB: in the Vale Inco's Creighton Mine in Sudbury, Ontario, Canada; 2073 m deep; the laboratory is home to the Sudbury Neutrino Observatory (SNO), an experiment which confirmed the oscillation hypothesis for solar neutrinos by detecting the flux of all flavours.

2.4 Other detector applications

Detectors, which had been developed for research in nuclear and particle physics, have also been successfully employed in other fields, such as medicine, biology, material science, radiation and environment protection, archaeology, geology and astronomy. In these areas detectors are mostly used for relatively low-energetic particles, emitted from radioactive nuclei, X-ray sources or low-energy particle accelerators.

Dosimetry. Whenever radiation is involved, for example in research, engineering or medicine, the monitoring of the radiation with dosimeters is mandatory (see e.g. [911]). Dosimetry instruments are, for example, Geiger counters, ionisation chambers, photographic films and semiconductor detectors. Each one of them has specific capabilities: a Geiger counter or semiconductor can detect single radiation quanta, the current of an ionisation chamber yields the radiation level and on a film the time-integrated radiation dose is recorded. The most important measurement quantities and units of dosimetry are compiled in appendix A.1.

Tomography, radio-labelling. Tomography is a method of producing a three-dimensional image of the spatial structures of an object, such as the human body, the earth or machine parts, by measuring the direction dependence of radiation transmission. For example, particle detectors are employed for X-ray computed tomography (CT) by analysing the direction dependence of X-ray absorption. In positron emission tomography (PET, fig. 2.12) and in scintigraphy using single-photon-emission computed tomography (SPECT) the examined biological structures are labelled by atoms or compounds which contain radioactive nuclides. The nuclides emit particles whose directions are analysed yielding a picture of the spatial distribution of the radioactivity.

Tracer method and radiometric dating. 'Tracers' are radioactive additives to materials allowing us to trace material transport. For example, flows of liquids or the abrasion debris of mechanical machine parts can be studied in this way.

In geology and archaeology the dating of rocks and of archaeological finds by measurement of the activity of specific nuclides became an important research tool. An example is the radiocarbon dating where the β activity of ^{14}C in an organic specimen is measured using a proportional counter.

X-rays and synchrotron radiation. Scintillators and semiconductor detectors are employed in structure analysis with X-rays and synchrotron radiation. Similar methods are utilised in astronomy for the detection of cosmic X-rays by satellite experiments.

Diagnostics with neutrons. The scattering of low-energy neutrons off hydrogen nuclei is particularly efficient (because of their equal mass). Therefore, in various areas neutron scattering is used as an indicator for the abundance of water or, more generally, of hydrogen. For example, in geology the neutron method is used for rock examinations in boreholes. Neutron detectors (section 14.5) are counting tubes or scintillators that are combined with materials which have a high cross section for neutrons.

2.5 Units and conventions

2.5.1 Units

Following a common use in particle physics, formulae in this book are often expressed using a system of units where $\hbar$ and c are dimensionless with $\hbar = c = 1$ ('natural units'). Sometimes, we also write explicitly $\hbar$ and c if in the literature the corresponding

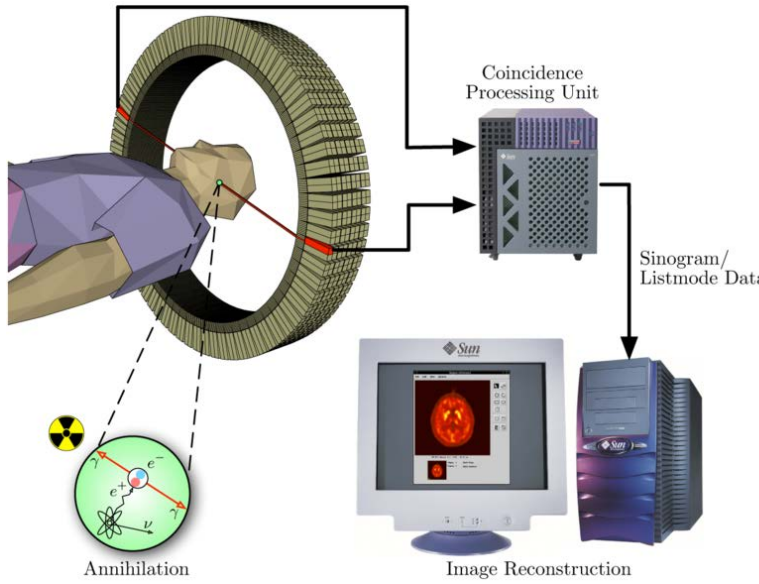


Fig. 2.12 Scheme of a positron emission tomograph (PET). PET uses a positron emitter as marker. The positrons are stopped in the tissue near the source and annihilate with electrons into two back-to-back photons. The photons are detected by a ring-shaped array of scintillator crystals. The sample of all measured photon directions allows for a tomographic reconstruction of the emitter distribution. Source: Wikimedia Commons, author J. Langner [643].

expressions are usually written in this way (e.g. atomic physics formulae in chapter 3 on interactions) or if numerical calculations have to be carried out using a certain formula. Numerical values are then given in SI units (SI = Système International d'Unités, known in English as the International System of Units) [260]. For the conversion the following relations are helpful:

$$\hbar c = 197.3 \text{ MeV fm} \approx 200 \text{ MeV fm}, \quad (\hbar c)^2 = 0.3894 \text{ GeV}^2 \text{ mb} \approx 0.4 \text{ GeV}^2 \text{ mb} \quad (2.4)$$

The relativistic relation

$$E = \sqrt{p^2 + m^2} \quad (2.5)$$

is numerically fulfilled if, for example, the energy E is given in GeV, the momentum p in GeV/c and the mass m in GeV/c^2.

2.5.2 Standards for material properties

Quoting material properties we often refer to the collection 'Atomic and Nuclear Properties' of the Particle Data Group [762]. The specific values are given under different ambient conditions which have to be accounted for. This is most important for gas properties like density and refractive index. The Particle Data Group mostly refers to two standard conditions for temperature and pressure. See [481] from which we quote the following definitions:

STP Standard temperature and pressure (STP) for a gas is defined by the International Union of Pure and Applied Chemistry (IUPAC) as 0 °C (273.15 K), 100 kPa.

This differs from an earlier definition that used 1 atm (101.325 kPa) as the pressure reference.

NTP Normal temperature and pressure (NTP) for a gas is defined as 20 °C (293.15 K), 1 atm (101.325 kPa).

Other standards abound, and the definitions are not necessarily consistent (see [481] and references therein).

2.5.3 Particle kinematics

In particle physics the kinematics is mostly evaluated using relativistic formulae. The Lorentz variables γ, β are defined as (E, m, p in natural units)

$$\gamma = \frac{E}{m} = \frac{1}{\sqrt{1-\beta^2}}, \quad \beta = \frac{p}{E} = \frac{v}{c}\,. \tag{2.6}$$

The kinetic energy is

$$T = E - m \;\rightarrow\; \frac{p^2}{2m} \quad \text{for } \gamma \to 1\,, \tag{2.7}$$

where $\gamma \to 1$ means the non-relativistic limit.

With most general-purpose detectors energy, momentum and angle of the particles, sometimes also the mass, can be determined. Momenta are often distinguished as longitudinal or transverse momenta (p_L, p_T) relative to a preference direction. For example, this preference direction could be the direction of a magnetic field where p_T lies in the plane transverse to the field containing the curvature of the charged particles. Often the preference direction is the beam direction which for collider experiments is usually also the magnetic field direction.

In head-on collisions the vector sum of the transverse momenta of all reaction products relative to the beam direction vanishes. The transverse momentum conservation is exploited to calculate from the observed transverse momentum $\vec{p}_T^{\,obs}$ the missing transverse momentum, $\vec{p}_T^{\,miss} = -\vec{p}_T^{\,obs}$, and to estimate from that the kinematics of unobserved particles, like neutrinos or invisible exotic particles. For calorimetric measurements the 'transverse energy' E_T and the 'missing transverse energy' E_T^{miss} are used instead. Both are derived by treating energies like momentum vectors with the vector pointing to an energy deposition and using the relativistic relation $E \approx p$. In general, for longitudinal momenta a similar constraint as for transverse momenta does not hold since there are always gaps in the instrumentation around the beams so that particles are lost in forward and backward directions. In particular for hadron beams, both in fixed-target or collider configurations, the reaction products are strongly collimated in the forward direction.

Relative to a preference direction also polar and azimuth angles θ, ϕ are defined (see e.g. fig. 9.6). In collider experiments with cylinder symmetry the two angles are defined by the cylinder coordinates. Instead of the polar angle θ the 'rapidity'

$$y = \frac{1}{2}\ln\left(\frac{E+p_L}{E-p_L}\right) = \operatorname{arctanh}\left(\frac{p_L}{E}\right) \tag{2.8}$$

is also used. At high energies, with $E \approx p$ and with $\cos\theta = p_L/p$, the equation becomes

$$y \to -\ln\left(\tan\frac{\theta}{2}\right) =: \eta\,. \tag{2.9}$$

The so defined 'pseudo-rapidity' η is also used independently of the validity of the high-energy approximation if the mass of the respective particle is not known.

2.6 Content overview

In the chapters 3 to 5 the essential theoretical and technical foundations for the conception and construction of detectors are presented. It starts with the 'interactions of particles with matter' in chapter 3, the basic prerequisite of particle detection. Essential for detector operation are also the phenomena of charge transport in electric and magnetic fields, discussed in chapter 4, and the signals resulting from the charge movement, discussed in chapter 5. The next four chapters 6 to 9 treat the detection of tracks generated by charged particles. Beginning with the classical 'non-electronic detectors', which are mostly of historical interest, detector types are presented whose sensitive media are either gas (chapter 7) or semiconductors (chapter 8). Track reconstruction with and without magnetic field is explained in chapter 9. In the following four chapters detector effects and methods are described which are in a wide sense related to particle identification. The Cherenkov, transition and scintillation effects are discussed in chapters 11, 12 and 13, respectively. The detection of these effects requires photodetectors, which are presented in chapter 10. In chapter 14 an overview of methods for particle identification is given. Energy measurements, also referred to as *calorimetry*, are covered by chapter 15 for accelerator-based experiments and by chapter 16 for astrophysical experiments or, more generally, for non-accelerator experiments. The last two chapters, 17 and 18, deal with the electronics of detectors comprising the signal readout, trigger and data acquisition systems which make the data available in digital form for physical analysis.

3
Interactions of particles with matter

Particles, charged or neutral, can only be sensed by interacting with matter. Detectors usually exploit the following processes:

- ionisation and excitation of atoms in media by charged particles;
- bremsstrahlung: photon radiation emitted by charged particles in the fields of atomic nuclei;
- photon scattering and photon absorption;
- Cherenkov and transition radiation;
- nuclear reactions: hadrons (p, n, π, α, ...) with nuclear matter;
- weak interactions constituting the only possibility to detect neutrinos.

Generally, a particle will undergo more than just one interaction process on its path through matter if it is not absorbed in its first interaction. For example, charged particles usually lose energy in a large number of subsequent ionisation and excitation interactions with atoms of the medium they pass through.

The probability that a particle interacts with the atoms of a medium is defined by the cross section of a reaction. The definition of the cross section and related terms will be given in the following section.

3.1 Cross section and absorption of particles and radiation in matter

The cross section is a measure for the probability of a particle reaction, which in turn depends on the kind and strength of the interaction(s) between the scattering partners. It can be interpreted as an effective interaction area, as sketched in fig. 3.1. The cross section σ represents the effective area of a target particle 'seen' by an incoming particle beam. We assume for simplicity that the beam particles have no spatial extent. The beam enters the target with area S and length l with a rate $\dot{N}_{in}$. There are

$$N_T = \frac{\rho V}{A} N_A \tag{3.1}$$

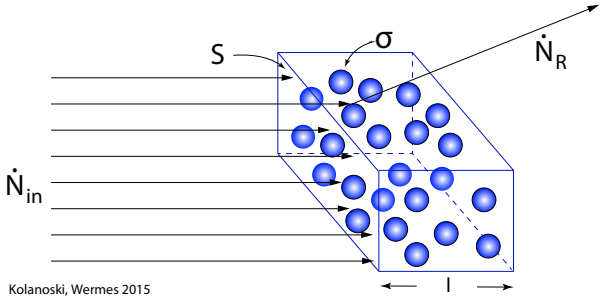


Fig. 3.1 Illustration for the cross section definition: $\dot{N}_{in}$ and $\dot{N}_R$ are incoming particle rate and reaction rate, respectively; S and l are total area and target length, respectively; σ denotes the cross section.

particles in the target volume $V = S\,l$ where ρ is the target mass density, A the atomic mass per mole of the target particles[1] and $N_A = 6.02214076 \times 10^{23}/\text{mol}$ is Avogadro's number. Since 2019 the numerical value of N_A is fixed and is used for the definition of the mole as the amount of substance which contains N_A entities [260]. The target can consist of any type of particle (atom, molecule, electron, ...) if A is taken as the corresponding mass per mole.

The particle number density is given by

$$n = \frac{N_T}{V} = \frac{\rho}{A}\,N_A\,. \tag{3.2}$$

The beam 'sees' a total area $N_T\,\sigma$ of target particles. The probability to hit a target particle is $P = N_T\,\sigma/S$. This probability can also be expressed by the scattering or reaction rate $\dot{N}_R$ relative to the rate of incoming beam particles $\dot{N}_{in}$, provided that the target is sufficiently thin such that any change of the beam while passing the target can be neglected:

$$P = \frac{\dot{N}_R}{\dot{N}_{in}} = \frac{N_T\,\sigma}{S} = n\,\sigma\,l\,. \tag{3.3}$$

The cross section can then be expressed as

$$\sigma = \frac{\dot{N}_R}{\dot{N}_{in}}\,\frac{1}{n\,l}\,. \tag{3.4}$$

Hence the reaction rate is proportional to the cross section, the proportionality constant being the luminosity L:

$$\dot{N}_R = \sigma\,L, \qquad L = \dot{N}_{in}\,n\,l\,. \tag{3.5}$$

The term 'luminosity', here defined for a so-called fixed-target experiment, is more commonly used for colliding beams (see the corresponding discussion on page 11).

If the assumption of a thin target is no longer valid one must take into account that the number of particles $N(x)$ that have thus far not interacted with the target decreases exponentially with the penetration depth x (fig. 3.2(a)):

[1] The use of the symbol A in this context is not unique in the literature. Either it denotes the *(atomic) mass number*, that is, the number of nucleons of a nuclide, or it denotes the *atomic mass*. In this book we use for the (dimensionless) mass number the symbol $\mathcal{A}$. For the symbol A we adopt the convention used by the PDG in their review of particle interactions with matter [201] where A is the *atomic mass per mole* [762]. The definition accounts for the isotopic composition of an element and the atomic composition of molecules and mixtures. Due to the definition of the mole the numerical values in units g/mol are close to the respective mass number, nearly equal ('exactly equal' until the SI-unit system redefinition in 2019 [260]) for the nuclide ^{12}C.

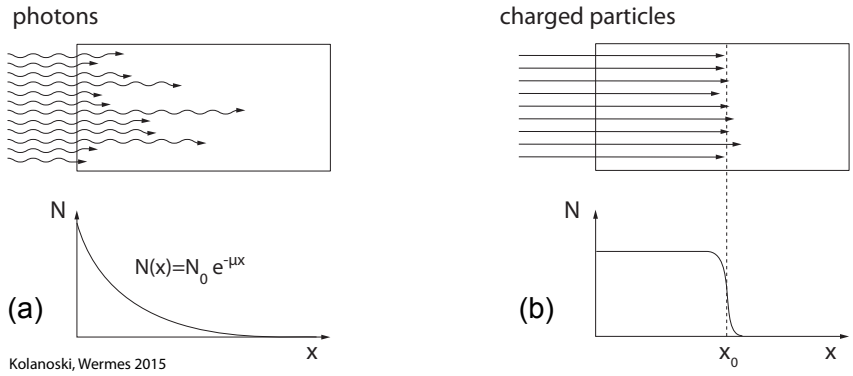


Fig. 3.2 Number of particles in a beam as a function of the penetration depth. (a) Loss of particles due to absorption (typical for photons) leads to an exponential decrease of the beam intensity, (b) energy loss in matter (typical for charged particles) leads to a finite range with a fairly local drop to zero in intensity.

$$\frac{dN}{N} = -n\sigma dx \qquad \Rightarrow \qquad N(x) = N_0 \mathrm{e}^{-\mu x}\,, \tag{3.6}$$

where N_0 is the original number of particles and $\mu = n\sigma$. If the beam particles are absorbed when they interact with target particles then μ is called the *absorption coefficient* or *linear attenuation coefficient*. Equation (3.6) is also known as the *Beer–Lambert law*. The *mean free path* of a particle is defined as

$$\lambda = \frac{1}{\mu} = \frac{1}{n\sigma}. \tag{3.7}$$

The cross section has dimensions of area; its unit is called *barn*:

$$1\,\mathrm{barn} = 1\,\mathrm{b} = 10^{-24}\,\mathrm{cm}^2. \tag{3.8}$$

Since nuclear densities are largely independent of the mass of a nucleus, the nuclear volume is roughly proportional to the number of nucleons $\mathcal{A}$ and the cross-sectional area is proportional to $\mathcal{A}^{2/3}$. Because of the short-range nature of nuclear forces their geometrical area also governs the order of magnitude of their interaction cross sections:

$$\sigma_{nucl} \approx \pi r_0^2\, \mathcal{A}^{\frac{2}{3}} \approx 45\,\mathrm{mb}\,\mathcal{A}^{\frac{2}{3}}\,. \tag{3.9}$$

The typical order of magnitude is in the range of femtometres with $r_0 \approx 1.2$ fm. Correspondingly the magnitude of atomic cross sections can be estimated using the Bohr radius a_0 ($a_0 \approx 0.05\,\mathrm{nm}$):

$$\sigma_{\mathrm{atom}} \approx \pi a_0^2 \approx 10^8\,\mathrm{b}\,.$$

Due to the long-range nature of the electromagnetic interaction, however, a simple dependence on $\mathcal{A}$ as for nuclei is no longer given. Owing to the shell structure of the atoms, the atomic radius increases within a group of atoms with Z electrons; however, within an atom period the radius shrinks from alkali metals to ideal gases.

For our applications we distinguish between elastic scattering on the one hand and inelastic scattering on the other, where for the former the incoming (beam) particle

is still present in the final state, while it has been absorbed in the latter. An example for the elastic/inelastic case is the scattering of a charged particle off atomic shell electrons which can be either entirely elastic or inelastic when ionisation or excitation of the atom occurs. An example for absorption is the photoelectric effect, by which a photon is absorbed by an atom emitting an electron. For the two cases very different dependencies of the number of beam particles as a function of the penetration depth result, as shown in fig. 3.2: in the case of absorption an exponential decrease occurs, whereas for continuous energy loss the beam intensity remains practically constant until the entire energy of the particle has been transferred to the medium (see also section 3.2.4).

3.2 Energy loss of charged particles by ionisation

3.2.1 The Bethe–Bloch formula for the average energy loss

When penetrating a medium charged particles lose energy by ionisation and excitation of the medium's atoms (fig. 3.3). This is the dominant energy loss process of charged particles up to very high particle velocities when radiation effects start to play a role (see section 3.3).

The energy loss per path length is described by the *Bethe–Bloch formula*. The first calculations using classical methods were already published by Bohr in 1913 [226]. In the 1930s Bethe [187] and Bloch [214,213] provided a quantum mechanical treatment of the problem, the results of which are to a large extent still valid today. Subsequently the calculations have been refined; in particular corrections for various kinematic regimes were computed (see e.g. [227, 392] and references in [746]).

Energy loss of a particle passing through matter arises from a sequence of individual stochastic processes. The *average energy loss* per path length depends on the properties of the medium a as well as on the mass M and the velocity β of the particle:

$$-\left\langle \frac{dE}{dx} \right\rangle = n \int_{T_{min}}^{T_{max}} T \, \frac{d\sigma_a}{dT}(M, \beta, T) \, dT \, . \tag{3.10}$$

Here n is the target number density and $d\sigma_a/dT$ the differential cross section for a loss of kinetic energy T in the collision. The integral from T_{min} to T_{max} comprises the region of all possible energy transfers. Energy transferred to the atom can be spent to release an electron from the atom by ionisation (providing binding energy and kinetic energy of the electron), but also to merely excite the atom. We will see that the computation of (3.10) is far from straightforward. At low energy transfers, corresponding to large impact parameters where atom excitations play the dominant role, the integral cannot be solved classically.

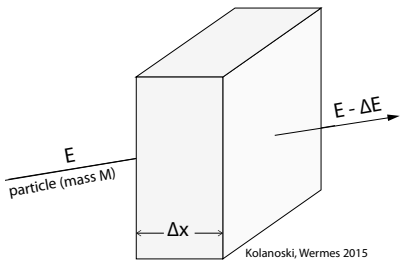


Fig. 3.3 Energy loss of a particle passing through matter.

3.2.1.1 Energy loss through Rutherford scattering off shell electrons

In what follows we restrict ourselves to 'heavy' charged particles (henceforth called 'the particle') passing through the medium. These represent all charged particles except electrons and positrons (see page 35) for which exchange and annihilation must be considered in addition. To compute (3.10) we subdivide the range of energy transfers into a high energy region ($T \geq T_1$) with large losses T of the particle's kinetic energy and a low energy region ($T < T_1$), where binding energies of the electrons cannot be neglected and a quantum mechanical treatment is needed. Note that in contrast to a classical treatment in quantum mechanics T cannot become zero. We hence obtain the total mean energy loss from two parts:

$$-\left\langle \frac{dE}{dx} \right\rangle = n_e \int_{T_1}^{T_{max}} T \, \frac{d\sigma}{dT} \, dT \; - \left\langle \frac{dE}{dx} \right\rangle_{T<T_1} . \tag{3.11}$$

The energy T_1 dividing these regions is typically between 0.01 and 0.1 MeV [835], but will eventually drop out in the sum of both contributions (see eq. (3.25)). Since scattering takes place on all shell electrons the relevant target density is the electron density

$$n_e = Z \frac{\rho N_A}{A} \tag{3.12}$$

with Z being the number of electrons per atom or molecule.

To begin with the high energy region, we assume that the velocity of the particle is large compared to the electron orbit velocity and that the energy transferred to an electron is large compared to its binding energy. The electron is thus regarded as 'quasi free'. Then the scattering process can be described by the Rutherford scattering formula (see e.g. [564], ch. 13). As is well known, the classical and the (first order) quantum theoretical derivation of the Rutherford formula yield the same result. Therefore quantum effects must only be taken into account in energy loss computations when the above assumptions are no longer valid, that is, for small energy transfers in the same order of magnitude as the binding energies or for particle velocities as low as the electron orbit velocities. We shall come back to these cases later.

We consider a projectile particle with mass M ($M \gg m_e$), charge ze and 4-momentum vectors P and P' before and after the interaction, respectively. The 4-momentum vectors of the electron are p_e and correspondingly p'_e. With the 4-momentum transfer

$$Q^2 = -(P - P')^2 = -(p_e - p'_e)^2 \tag{3.13}$$

one obtains the Lorentz-invariant expression for the Rutherford cross section [564]:

$$\frac{d\sigma}{dQ^2} = \frac{4\pi \, z^2 \alpha^2 \hbar^2 c^2}{\beta^2} \frac{1}{Q^4} . \tag{3.14}$$

Here βc is the velocity of the particles relative to each other, which in their respective rest systems is identical and can be defined in a Lorentz-invariant way. In the system in which the scattered shell electron is initially at rest, it has an energy $E_e = m_e c^2$ before and $E'_e = T + m_e c^2$ after the collision. The squared 4-momentum transfer hence is

$$Q^2 = -(p_e - p'_e)^2 = 2 m_e c^2 T . \tag{3.15}$$

Putting this into (3.14) yields the differential cross section (differential in the particle's energy loss T):

$$\frac{d\sigma}{dT} = \frac{2\pi\, z^2\alpha^2\hbar^2}{\beta^2\, m_e}\,\frac{1}{T^2}\,. \tag{3.16}$$

Extension of (3.16) to the Mott cross section which takes into account a possible spin flip of the target electron[2] is obtained by applying a factor $(1-\beta^2 T/T_{max})$ [564]:

$$\frac{d\sigma}{dT} = \frac{2\pi\, z^2\alpha^2\hbar^2}{\beta^2\, m_e\, T^2}\left(1-\beta^2\frac{T}{T_{max}}\right). \tag{3.17}$$

For the assumed high energy region $T_1 \le T \le T_{max}$, where the lower limit is much larger than the binding energies of the shell electrons, we can solve the integral (3.10):

$$\begin{aligned} -\left\langle\frac{dE}{dx}\right\rangle_{T>T_1} &= n_e\int_{T_1}^{T_{max}} T\;\frac{2\pi\, z^2\alpha^2\hbar^2}{\beta^2\, m_e\, T^2}\left(1-\beta^2\frac{T}{T_{max}}\right)dT \\ &= \frac{2\pi\, z^2\alpha^2\hbar^2}{\beta^2\, m_e}\, n_e\;\left(\ln\frac{T_{max}}{T_1}-\beta^2\right). \end{aligned} \tag{3.18}$$

In (3.18) a term $\beta^2(1-T_1/T_{max})$ has been approximated by β^2, assuming $T_1 \ll T_{max}$.

Maximum energy transfer T_{max}. The maximally possible energy transfer T_{max} can be computed from the kinematics of an elastic collision of the particle with the shell electron. The transfer is maximal for a central collision, that is, when the vectors $\vec{P}$, $\vec{P}'$, $\vec{p}_e'$ are all parallel. For the incoming particle, energy and momentum can be expressed by the Lorentz factors β, γ:

$$E = \gamma Mc^2, \qquad |\vec{P}| = \beta\,\gamma\, Mc\,. \tag{3.19}$$

The kinetic energy of the outgoing electron is

$$T = E_e' - m_e c^2\,. \tag{3.20}$$

Neglecting the binding energy we then obtain for the maximum kinetic energy (derived in section 3.2.2):

$$\begin{aligned} T_{max} &= \frac{2\, m_e\, c^2\, \beta^2\, \gamma^2}{1+2\,\gamma\, m_e/M+(m_e/M)^2} \\ &\approx \begin{cases} 2\, m_e\, c^2\,(\beta\gamma)^2 & \text{for}\;\; \gamma m_e \ll M \\ \gamma\, M\, c^2 \;=\; E & \text{for}\;\; \gamma\to\infty \\ m_e c^2(\gamma-1) \;=\; E - m_e c^2 & \text{for}\;\; M = m_e\,. \end{cases} \end{aligned} \tag{3.21}$$

For this high energy regime, the maximally possible energy transfer in (3.18) and the maximum kinetic energy of the scattered electron in (3.21) can be assumed to be approximately the same and equal to the maximum energy loss of the particle.

It should be noted that for both the highly relativistic limit ($\gamma\to\infty$) and also for $M = m_e$ the full energy of the incoming particle can be transferred to a shell electron.

[2] To a good approximation one can neglect the contribution from a spin flip of the projectile particle because the responsible magnetic coupling is suppressed by a factor Q^2/M^2.

If the incoming particle is an electron then scattering and scattered particle cannot be distinguished (see also page 35). In this case the maximal possible energy transfer is only $T_{max}/2$, since the 'scattered' electron is always defined as the one with the smaller energy.

Minimum energy transfer. Classically the energy transferred to a free electron can become arbitrarily small. Quantum mechanically, however, it must be considered that below the ionisation threshold only discrete energy transfers are allowed in a single collision. Furthermore, for particle velocities similar or smaller than the electron orbit velocities, interference effects play a role. The quantum mechanical treatment of atomic excitations and corrections caused by screening effects of atomic shells is the most difficult problem in the computation of energy loss in matter. A considerable amount of literature on this subject is available summarised by the Particle Data Group [746,482] following [187,835,392,200]. A detailed commentary is given in [479]. Already in 1913 Bohr explained a passing electrons's energy loss by the generation of 'vibrations' of the atom electrons of the absorbing material [226]. The first rigorous treatment was given by Bethe [187] describing the energy loss by atomic excitations to different energy levels below the ionisation threshold down to (close to) zero momentum transfer and yields [482]:

$$-\left\langle \frac{dE}{dx} \right\rangle_{T<T_1} = \frac{2\pi\, z^2 \alpha^2 \hbar^2}{\beta^2\, m_e}\, n_e \left[\ln \frac{2\, m_e\, c^2\, \beta^2\, T_1}{I^2} - \ln \frac{1}{\gamma^2} - \beta^2 \right] . \tag{3.22}$$

In essence (3.22) corresponds to an integral over an energy region between an effective minimal energy transfer $T_{min} = I^2/2m_e c^2 \beta^2$ (typically 0.1 eV to a few eV) and T_1 at the upper end. The quantity I is the mean excitation energy and is defined via its logarithm $\ln I$ as the sum of the logarithms of the excitation energies E_k weighted by the corresponding oscillator strengths f_k for a transition to the energy level k [187,392]:

$$\ln I = \sum_k f_k \ln E_k \,. \tag{3.23}$$

Logarithms are used because T_1 also enters logarithmically in (3.22). In principle I is defined by (3.23), but usually it is taken from a fit to experimental data. The determination of I is a very difficult part in the computation of the mean energy loss [482, 885], such that I values found in the literature depend on the publication date. In fig. 3.4 I/Z is displayed as a function of the atomic number Z using recent values [545]. Their dependence on Z can be roughly described by a simple power law fit yielding

$$I \approx 17.7\, Z^{\,0.85}\ \text{eV}. \tag{3.24}$$

The term $\ln 1/\gamma^2$ in (3.22) accounts for the relativistic growth of the electric field extension perpendicular to the direction of the moving charge. Since the penetration depth of the transverse field increases linearly with γ the maximal impact parameter of the passing particle contributing to energy transfers increases. Classically, the integration over all contributing impact parameters leads to the logarithmic γ dependence (see e.g. the corresponding discussion in [564] and in the appendix of [84]).

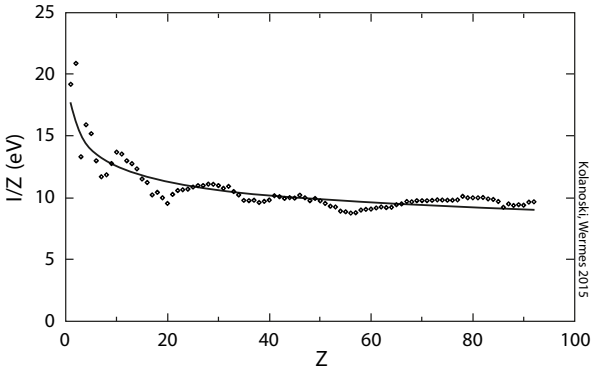


Fig. 3.4 Mean excitation energy I devided by the atomic number Z as a function of Z. The values for I are taken from table 1 in [545]. The curve represents a fit of a simple power law in Z following (3.24).

3.2.1.2 The Bethe–Bloch formula

With the ansatz in (3.11) and the respective parts for the high energy regime in (3.18) and the low energy regime in (3.22) as well as substituting the electron density n_e as given in (3.12), the *Bethe–Bloch formula* is obtained (see [188,835]) as quoted in [746]:

$$-\left\langle \frac{dE}{dx} \right\rangle = K \frac{Z}{A} \rho \frac{z^2}{\beta^2} \left[\frac{1}{2} \ln \frac{2\, m_e\, c^2\, \beta^2\, \gamma^2\, T_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} - \frac{C(\beta\gamma,\, I)}{Z} \right] . \quad (3.25)$$

The formula contains two additional correction terms $\delta/2$ and C/Z, relevant at high and low energies, respectively (see below). The various quantities in the equation are defined as follows:

- $K = 4\pi N_A r_e^2 m_e\, c^2 = 0.307$ MeV cm^2/mol, using the classical electron radius:

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2} = \alpha \frac{\hbar}{m_e c} \approx 2.8\,\text{fm}\,. \quad (3.26)$$

- z, β are charge and velocity of the projectile particle.
- Z, A are atomic number and atomic mass of the medium, respectively.
- I is the mean excitation energy.
- T_{max} is the maximum possible energy transfer to a shell electron, occurring in a central collision (see above).
- δ is the so-called density correction, important at high energies.
- C/Z is a shell correction, relevant for small β values.

In tables one usually finds the mean energy loss normalised to the density of the medium (omitting here and in what follows the brackets for the mean value):

$$\frac{dE}{\rho\, dx} \quad \text{in the units} \quad \frac{\text{MeV}}{\text{g cm}^{-2}}\,. \quad (3.27)$$

Often one still just writes dE/dx with x being a density weighted length in units of mass per area, for example g/cm^2, also called *column density*:

$$x \quad \longrightarrow \quad \rho\, x\,. \quad (3.28)$$

Equation (3.25) is valid for 'heavy' particles which includes all particles except electrons and positrons. For electrons/positrons (3.25) must be modified (see page 35).

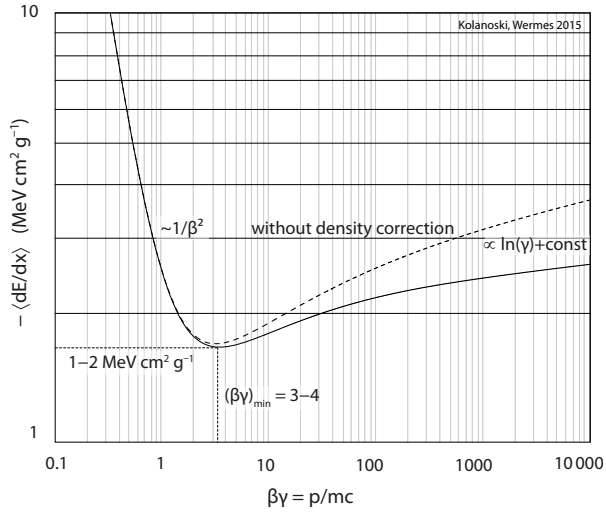


Fig. 3.5 Mean energy loss of charged particles by ionisation as a function of $\beta\gamma = p/mc$, here given for charged pions in silicon. The range indicated for the minimum of the energy loss is valid for most media. At high energies the density effect is evident as the deviation from the log γ trend due to the polarisation of the medium by the charged particle and hence screening further extension of the transverse electric field.

The Bethe–Bloch formula describes how particles are stopped in matter. The mean energy loss in matter is hence also called *stopping power*.[3] The stopping power determines for example the range of a particle in matter or the energy lost after penetration of a layer of material. One should distinguish the stopping power from the measurable deposited energy in a detector seen for instance as measured ionisation charges, thus neglecting atom excitations when their de-excitation happens through low energetic photons. The latter do, however, contribute to the total stopping power.[4] In a thin detector layer it can also happen that a high energetic fraction of the kicked-off shell electrons, called δ *electrons*, leaves the detector thus reducing the mean observed energy.

3.2.1.3 Energy dependence of $\langle dE/dx \rangle$

Figure 3.5 shows the mean energy loss $\langle dE/dx \rangle$ as a function of the energy of the incoming particle for a typical example ($\pi^{\pm}$ in silicon) following (3.25) with a parametrisation of the density and shell correction as in [914]. At low energies the $1/\beta^2$ term, at high energies the $\ln\gamma$ term is dominant. In between both regions there is a broad minimum around $\beta\gamma \approx 3-3.5$, respectively $\beta \approx 0.95$, depending on Z. Particles in this kinematic range are thus called *minimum-ionising particles* (*mips*). Since the increase in dE/dx for energies corresponding to $\beta\gamma > 3.5$ is only moderate compared to the steep rise $\propto 1/\beta^2$ towards energies lower than the minimum, it is common practice to use the term *mip* for all charged particles with energies larger than those at the minimum. In table 3.1 one finds besides other parameters also the minimum dE/dx values for some materials.

Energy loss at low energies. The $1/\beta^2$ dependence can be explained by the fact that the momentum transfer increases with the effective interaction time $\Delta t \simeq b/\gamma v$ (b = impact parameter, $v = c/\beta$) which is longer for slower particles. Since the momentum

[3] *Mass stopping power* is the stopping power divided by the density as in (3.27).

[4] Atomic excitations can, depending on the dielectric properties of the medium, lead to coherent radiation phenomena like Cherenkov and transition radiation (Chapters 11 and 12). A unified description of energy loss is attempted by the PAI model (PAI = *photoabsorption and ionisation*) [84].

Table 3.1 Parameters to compute the energy loss for some selected materials [762]. 'Nuclear emulsion' in [762] corresponds to photoemulsion type Ilford G.5 (see section 6.3). The bottom part contains parameters for the density correction (3.29) taken from [762] (Muon Energy Loss Tables).

Material	Z	A $\left(\frac{\text{g}}{\text{mol}}\right)$	ρ $\left(\frac{\text{g}}{\text{cm}^3}\right)$	I (eV)	$\left\langle \frac{dE}{dx} \right\rangle_{min}$ $\left(\frac{\text{MeV cm}^2}{\text{g}}\right)$	$\hbar\omega_P$ (eV)
C (graphite)	6	12.01	2.21	78	1.74	30.28
Al	13	26.98	2.70	166	1.62	32.86
Si	14	28.09	2.33	173	1.66	31.05
Fe	26	55.85	7.87	286	1.45	55.18
Pb	82	207.2	11.35	823	1.12	61.07
Ar (gas)	18	39.95	0.00166	188	1.52	0.79
CsI	108	259.8	4.51	553	1.24	39.46
polystyrene	56	104.2	1.06	68.7	1.94	21.75
water	10	18.0	1.00	79.7	1.99	21.48
(standard) rock	11	22.0	2.65	136.4	1.69	33.2
nuclear emulsion			3.82	331.0	1.42	38.0

Material	a	k	ζ_0	ζ_1	$-C_D$	δ_0
C (graphite)	0.208	2.95	−0.009	2.482	2.893	0.14
Al	0.080	3.63	0.171	3.013	4.240	0.12
Si	0.149	3.25	0.202	2.872	4.436	0.14
Fe	0.147	2.96	−0.001	3.153	4.291	0.12
Pb	0.094	3.16	0.378	3.807	6.202	0.14
Ar (gas)	0.197	2.96	1.764	4.486	11.95	0
CsI	0.254	2.67	0.040	3.335	6.28	0
polystyrene	0.165	3.22	0.165	2.503	3.30	0
water	0.091	3.48	0.24	2.800	3.502	0
(standard) rock	0.083	3.41	0.049	3.055	3.774	0.00
nuclear emulsion	0.124	3.01	0.101	3.487	5.332	0.00

transfer Δp is proportional to Δt it follows that the energy transfer $\Delta E = (\Delta p)^2/2m$ is proportional to $1/v^2 \propto 1/\beta^2$ since $\gamma \approx 1$ at low energies. If, however, the particle velocity comes in the range of shell electron velocities, the approximation that the minimal energy transfer is effectively determined by the mean excitation energy as assumed in (3.22) is no longer valid. As has been mentioned already, these low energy excitations must be analysed in detail and quantum mechanical interference effects, like the so-called Ramsauer effect (see section 4.6.2), must be considered. We shall discuss such effects in the context of drift and diffusion of electrons in gases in section 4.6.4.3.

The energy dependent shell corrections C/Z in (3.25) become relevant below $\beta \approx 0.3$. A review about the usage of shell corrections can be found in [201] with a further reference to a detailed discussion in [182]. The shape of the energy loss curve at low β values can be taken from fig. 3.6 for protons in argon. Below a kinetic energy of about 1 MeV, corresponding to $\beta \approx 0.05$, the curve deviates from the $1/\beta^2$ dependence of the Bethe–Bloch equation, reaching a maximum and falling off steeply thereafter (note the logarithmic scales!). At very low kinetic energies below some 100 eV, the proton

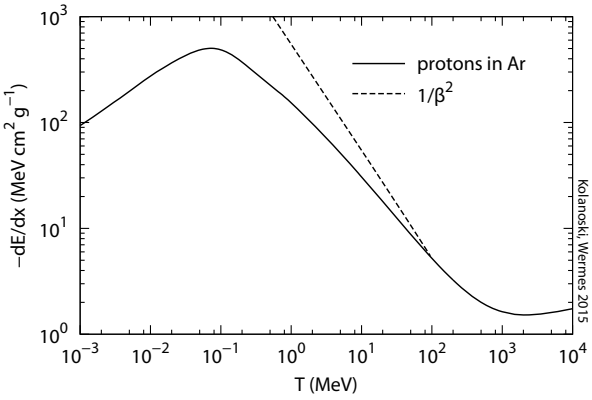


Fig. 3.6 Mean energy loss of protons in argon as a function of the proton kinetic energy (created using the interactive program PSTAR [182]). At low energies the influence of the shell corrections becomes evident as deviation from the $1/\beta^2$ behaviour.

loses energy mainly by elastic collisions with nuclei until it finally thermalises.

Energy loss at high energies. The reason for the rise of the energy loss at high energies is twofold. The first one is the asymptotic increase of the maximum energy transfer T_{max} with γ (see eq. (3.21)), which is a purely kinematic effect. The other is the mentioned relativistic effect, namely the increasing transverse extension of the electric field with γ leading to an increase of contributing large impact parameters and hence entering the Bethe–Bloch equation via the minimum possible energy transfer.

The increase of the E-field extension for highly relativistic particles, is however limited by the screening effect of nearby atoms as a consequence of the polarisation of the medium (density effect) leading to a flattening of the dE/dx curve at high energies. The density effect is taken into account in the Bethe–Bloch equation by the $\delta(\beta\gamma)$ term. A parametrisation is given in [914] for different energy regions:

$$\delta(\beta\gamma) = \begin{cases} 2\,\zeta\,\ln 10 + C_D & \text{for } \zeta \geq \zeta_1\,, \\ 2\,\zeta\,\ln 10 + C_D + a\,(\zeta_1 - \zeta)^k & \text{for } \zeta_0 \leq \zeta < \zeta_1\,, \\ \delta_0\,10^{2\,(\zeta-\zeta_0)} & \text{for } \zeta < \zeta_0\,, \end{cases} \tag{3.29}$$

$$\text{with} \qquad C_D = 2\,\ln(\hbar\omega_P/I) - 1 \quad \text{and} \quad \zeta = \log_{10}\beta\gamma\,.$$

For some materials the parameters in this equation are listed in table 3.1. A comprehensive collection for different materials can be found interactively on the web page of the Particle Data Group[5] [762].

For large impact parameters, corresponding to small energy transfers, the charge flying by acts coherently on the surrounding electrons and excites them to collective oscillations about the less mobile ions. The frequency of this oscillation, the so-called *plasma frequency*

$$\omega_P = \sqrt{\frac{n_e e^2}{\epsilon_0 m_e}}\,, \tag{3.30}$$

depends—apart from universal constants—only on the electron density. The lower limit for the transferable energy is therefore given by the plasma excitation energy $\hbar\omega_P$, asymptotically replacing the mean excitation energy I of a single atom.

[5]The parameters are somewhat hidden in the 'Table of Muon dE/dx and Range' in [762], entered there individually for each medium.

At the high energy end for large γ values, the density correction in (3.29) takes the form

$$\delta \to 2 \ln \frac{\hbar\omega_P}{I} + 2 \ln \beta\gamma - 1 \,, \tag{3.31}$$

employing the definition of the parameter C_D. In this extreme case the Bethe–Bloch formula (3.25) can be written in the form

$$-\frac{dE}{dx} = K \frac{Z}{A} \rho z^2 \left[\frac{1}{2} \ln \frac{2 m_e c^2 T_{max}}{(\hbar\omega_P)^2} - \frac{1}{2} \right] \quad (\gamma \to \infty, \beta \to 1) \,. \tag{3.32}$$

The mean excitation energy under the logarithm is replaced by the plasma energy $\hbar\omega_P$ and the $\beta^2\gamma^2$ term vanishes. A γ dependence of dE/dx then only enters via an increase of $T_{max} \propto \gamma$, such that the energy loss curve flattens off (fig. 3.5). The increase of T_{max} means that higher and higher energetic δ electrons can be produced. However, their energy contribution to an ionisation measurement often cannot fully be accounted for because δ electrons can either leave a (very thin) detector or their energy loss contribution to a measurement is not included. Motivations are for example to improve a position measurement (limiting the cluster size to suppress δ electrons emitted perpendicular to a particle track) or to achieve a better mean energy loss measurement $\langle dE/dx \rangle$. The measured energy loss is then 'restricted' and a corresponding definition of *restricted energy loss* is defined including only energy losses up to a maximum energy T_{cut}. Equation (3.25) is changed accordingly:

$$-\left.\frac{dE}{dx}\right|_{T<T_{cut}} = K \frac{Z}{A} \rho \frac{z^2}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 \beta^2 \gamma^2 T_{cut}}{I^2} - \frac{\beta^2}{2} \left(1 + \frac{T_{cut}}{T_{max}} \right) - \frac{\delta}{2} - \frac{C}{Z} \right] . \tag{3.33}$$

The restricted energy loss (3.33) saturates at high energies ('Fermi plateau') because T_{max} does not increase any longer with γ in (3.32), but is rather replaced by a fixed T_{cut} (see also [746] and fig. 3.13 on page 44). The Fermi plateau reaches a value higher than the minimum by typically about 40% in gases and less than 10% in solids.

3.2.1.4 Scaling laws and particle identification

The average energy loss per unit length (3.25), also called 'specific energy loss' (or 'specific ionisation' if, as it is often the case, only the ionisation part of the energy loss is measured) can approximately be expressed as a function of the particle velocity and charge:

$$\frac{dE}{dx} \approx z^2 Z f_\beta(\beta) = z^2 Z f_p \left(\frac{p}{M} \right) = z^2 Z f_T \left(\frac{T}{M} \right) , \tag{3.34}$$

with functions $f_{\beta,p,T}$ of β, p, or T, respectively. The direct proportionality to the atomic number Z results from the incoherent nature of the scattering off the shell electrons, meaning that the contributions of individual electrons must be added. The ratio ρ/A which also enters (3.25) can be assumed to be roughly constant. With the approximation $Z/A \lesssim 1/2\,\mathrm{mol/g}$, which is a reasonable assumption at least for lighter elements, the density normalised energy loss $dE/(\rho dx)$ is roughly independent of the material as can also be seen from the column $\langle dE/dx \rangle_{min}$ in table 3.1.

The dependencies on momentum and kinetic energy of the particle result because for a particle mass M the ratios $p/(Mc) = \gamma\beta$ and $T/(Mc^2) = \gamma - 1$ are functions of velocity only. For a given medium the functions $f_{\beta,p,T}$ are thus independent of the particle's identity and mass.

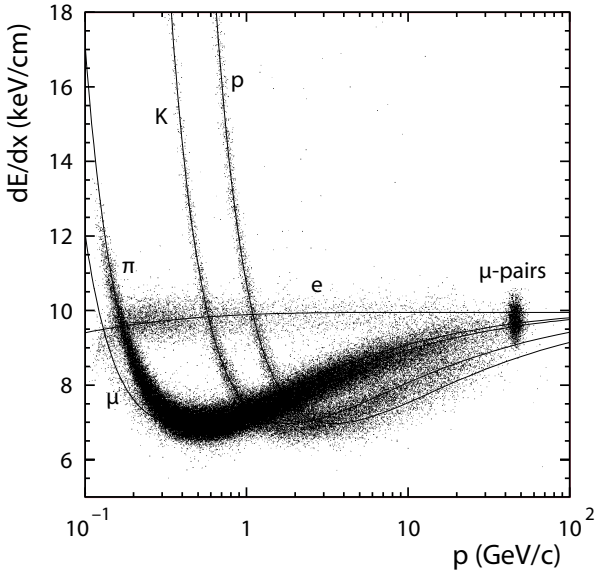


Fig. 3.7 Average ionisation energy loss of charged particle tracks measured as a function of their momentum on 159 anode wires (and averaged) in a drift chamber filled mainly with argon. For the averaging the 'truncated mean' method is employed by discarding the 30% highest values of a track (see section 14.2.2 and fig. 14.9). Every dot represents a measurement for one particle track, observed in decays of the Z^0 boson ($m_Z = 91\,\mathrm{GeV/c^2}$) (LEP, OPAL detector [509], with kind permission by Elsevier).

This property of dE/dx is exploited for particle identification: particles with the same momentum but different masses have different β and γ. Therefore different dE/dx curves for each mass result when plotted as a function of momentum (fig. 3.7, see also fig. 14.11). Hence knowing the momentum a particle's mass can be identified by measuring the mean energy loss $\langle dE/dx \rangle$ (see also chapter 14).

3.2.1.5 Energy loss by electrons and positrons

For electrons and positrons, due to their much smaller mass compared to heavier particles, energy loss by photon radiation (bremsstrahlung) becomes relevant already at relatively small energies. Since ionisation and bremsstrahlung of electrons or positrons can to a good approximation be treated independently we here restrict ourselves to the energy loss by ionisation. Bremsstrahlung is covered in section 3.3.

The ionisation energy loss by electrons and positrons differs from that by heavy particles due to kinematics, spin and their identity or respectively their antiparticle character with respect to the electrons of the medium. For positrons a complete loss via e^+e^- annihilation into photons must also be accounted for. Generally, two regimes of energy transfer can be distinguished: in the regime where the atomic energy levels cannot be neglected energy loss is treated by averaging, as done above for heavy particles; in the regime with large energy transfers discrete processes, that is Møller scattering ($e^-e^- \to e^-e^-$) or Bhabha scattering ($e^+e^- \to e^+e^-$), are computed.

Figure 3.8 shows the comparison of energy loss by protons and electrons in silicon. The difference near the minimum of the ionisation is about 10 %, increasing by small amounts towards smaller and larger velocities.

One should note that the Bethe–Bloch formula (3.25) includes energy loss by Cherenkov radiation (see also section 11.3 on page 447). These losses occur in transparent media for energy transfers of the passing particle below the ionisation threshold if refractive indices and frequencies fulfil the Cherenkov condition (see chapter 11). This has already been discussed by Fermi in 1940 [398] and is explicitly described for ex-

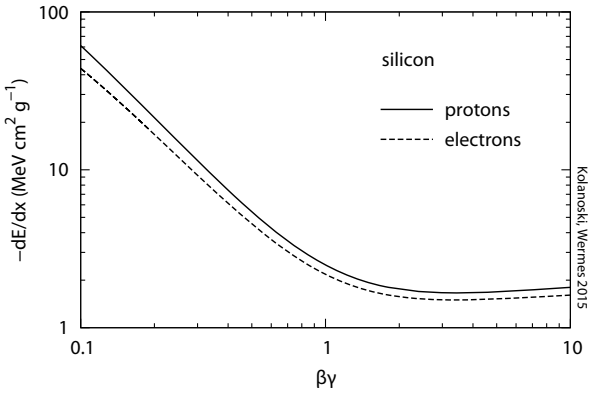


Fig. 3.8 Comparison of the energy loss by electrons (excluding bremsstrahlung) and by protons in silicon. The curves have been computed with the help of the programs ESTAR and PSTAR [182] (called '*collision energy loss*' for electrons and '*total energy loss*' for protons in these programs).

ample in [84]. The contribution to the total energy loss is, however, only of the order of a few per cent.

Transition radiation (chapter 12) results in a similar way if the traversed medium contains inhomogeneities.[6] It is generally not included in the energy loss described by the Bethe–Bloch formula (3.25), since homogeneous media are assumed in its derivation. More importantly, (3.25) does not include energy loss by bremsstrahlung which becomes dominant at high $\beta\gamma$ values (see section 3.3.2).

3.2.1.6 Parameters of the Bethe–Bloch formula for mixtures and compounds

For different components i of a medium the dE/dx values are summed, weighted by their mass (and neglecting atomic corrections):

$$\frac{dE}{\rho dx} = \sum_i w_i \left(\frac{dE}{\rho_i dx}\right)_i . \tag{3.35}$$

For compounds the weights w_i result from the abundance a_i of atom i with atomic mass A_i:

$$w_i = \frac{a_i A_i}{\sum_j a_j A_j} . \tag{3.36}$$

Also for the parameters in the Bethe–Bloch formula effective values for mixtures and compounds can be computed:

$$\begin{aligned} Z_{\text{eff}} &= \sum_i a_i Z_i \,, \\ A_{\text{eff}} &= \sum_i a_i A_i \,, \\ \ln I_{\text{eff}} &\approx \frac{1}{Z_{\text{eff}}} \sum_i a_i Z_i \ln I_i \,, \\ \delta_{\text{eff}} &\approx \frac{1}{Z_{\text{eff}}} \sum_i a_i Z_i \delta_i \,. \end{aligned} \tag{3.37}$$

[6] See e.g. in [392] the discussion of 'boundary effects' in section 7.6.

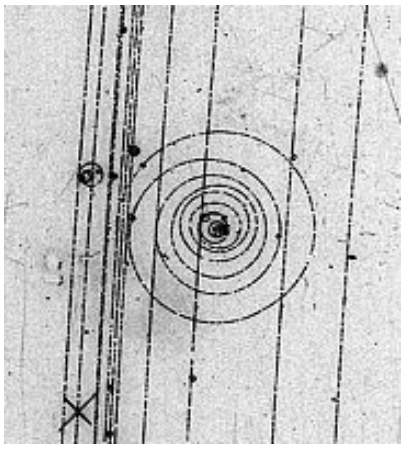


Fig. 3.9 Detail of a bubble chamber picture showing a δ electron emitted off a track to the right causing a spiralling secondary track in the magnetic field of the chamber (source CERN).

These averages have to be taken with some care, in particular for the parameters I and δ which depend on electron densities (even more so for the low energy correction C which is not listed here for this reason). The authors of [885] find that the above formula for I_{eff} works well if a compound only contains the elements H, C, N, O, F or Cl. For other contributing elements they suggest to increase the corresponding I value in the I_{eff} formula by 13%. The review 'Passage of particles through matter' e.g. in [746] discusses these issues and refers to the corresponding literature on measurements or, if they do not exist, to improved calculations (see also appendix A in [482]).

3.2.2 Delta electrons

So-called δ *electrons* or *high energy knock-on electrons* are emitted when the collisions of the projectile particle with shell electrons are close to central, causing high energy transfers. The kicked-off electrons then have comparatively high energies. According to (3.16) the electron kinetic energy follows a distribution proportional to $1/T^2$. The tails of this distribution can reach fairly large values up to T_{max} given in (3.21) especially for relativistic projectiles. Delta electrons with kinetic energies in the keV range and above can be observed in detectors having high spatial resolution like bubble chambers or emulsions (fig. 3.9), but also in modern electronic detectors with high granularity. If a detector is unable to resolve δ electrons emitted off a track, this potentially leads to a deterioration of the spatial resolution. Rarely occurring emission of very high energetic δ electrons also leads to larger fluctuations in dE/dx measurements for particle identification and hence to a worse resolution for the mean value $\langle dE/dx \rangle$. Because of their relevance to detectors and detection methods we discuss the energy and angular distributions of δ electrons in more detail in the following sections.

Relation between energy and emission angle. It is appropriate to consider the high-energetic processes leading to δ-rays by assuming elastic collisions with 'quasi free' electrons. Using the 4-vectors P, P', p_e, p'_e of the incoming and outgoing collision partners the relation between energy and emission angle of the electron in this elastic scattering process can be easily derived. Energy-momentum conservation

$$P + p_e = P' + p'_e \tag{3.38}$$

yields

$$(P - p'_e)^2 = (P' - p_e)^2 \qquad \Rightarrow \qquad P\,p'_e = P'\,p_e$$
$$\Rightarrow \qquad E\,E'_e - E'\,m_e c^2 = |\vec{P}c||\vec{p}'_e c|\,\cos\theta\,. \tag{3.39}$$

Using the Lorentz factors γ, β of the incoming particle and the relations

$$E = \gamma M c^2, \quad |\vec{P}| = \gamma\beta M c, \quad E' = E - T, \quad T = E'_e - m_e c^2, \quad |\vec{p}'_e c| = \sqrt{T^2 + 2T m_e c^2} \tag{3.40}$$

one obtains the sought-for relation between the kinetic energy T of the δ electron and the emission angle θ (fig. 3.10):

$$\cos\theta = \frac{T(\gamma + m_e/M)}{\gamma\beta\sqrt{T^2 + 2T m_e c^2}}\,, \tag{3.41}$$

$$T(\theta) = \frac{2\,m_e c^2\,\beta^2\,\gamma^2\,\cos^2\theta}{\gamma^2\,(1 - \beta^2\,\cos^2\theta) + 2\,\gamma\,m_e/M + m_e^2/M^2}\,. \tag{3.42}$$

The maximum energy transfer T_{max} to the electron is reached at $\theta = 0°$ and has already been quoted in (3.21). The minimum kinetic energy of the electron is $T_{min} = 0$, corresponding to $\theta = 90°$. One should note, however, that in this kinematic regime of minimal energy transfer the assumed scattering off a 'free' electron is no longer valid. The minimum amount of energy transferred to the atom is rather given by the minimum atom excitation energy.

For highly relativistic energies, $\gamma \gg 1$ ($\beta \to 1$), (3.42) becomes:

$$T(\theta) = \frac{2\,m_e c^2\,\cos^2\theta}{1 - \cos^2\theta} = \frac{2\,m_e c^2}{\tan^2\theta}\,. \tag{3.43}$$

This expression diverges for $\theta \to 0°$, that is, $T \to T_{max}$. Therefore, for $\theta \lesssim 1/\gamma$ one rather first computes the limiting case for the first term in (3.42),

$$\gamma^2(1 - \beta^2\,\cos^2\theta) \;\stackrel{\theta\to 0}{\longrightarrow}\; 1\,, \tag{3.44}$$

thus yielding $T_{max} = T(\theta = 0) \to \gamma M$ for $\gamma \to \infty$ as in (3.21).

It is remarkable that in the highly relativistic case ($\gamma \gg 1$ and $\theta \gg 1/\gamma$) the energy–angle relation (3.43) is no longer dependent on the properties of the incoming particle, neither its energy nor its mass (solid line in fig. 3.10).

Energy and angular distributions. To estimate the influence of δ electrons on detector space resolutions we examine the energy–angle relation a bit closer. Though large δ-electron emission angles correspond to small electron energies and hence usually have short ranges, the ionisation density of low energy electrons lying in the $1/\beta^2$ region of the Bethe–Bloch formula (fig. 3.5) is high, the higher the slower the electron moves. Therefore ionisation clusters along the δ electron's path deteriorate the resolution for hit reconstruction of the mother particle.

But δ electrons emitted under small angles can also lead to a shift of the centre of the measured charge cluster from a passing track due to multiple scattering (increasing with $1/(p\beta)$, section 3.4) or due to magnetic deflection (as in fig. 3.9). For a more quantitative estimation we examine the energy and angular distributions of the electrons a bit closer in the following.

The rate of δ electrons per energy interval dT and path length dx is

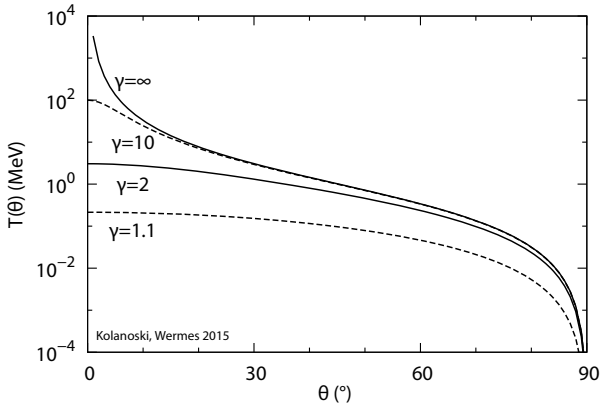
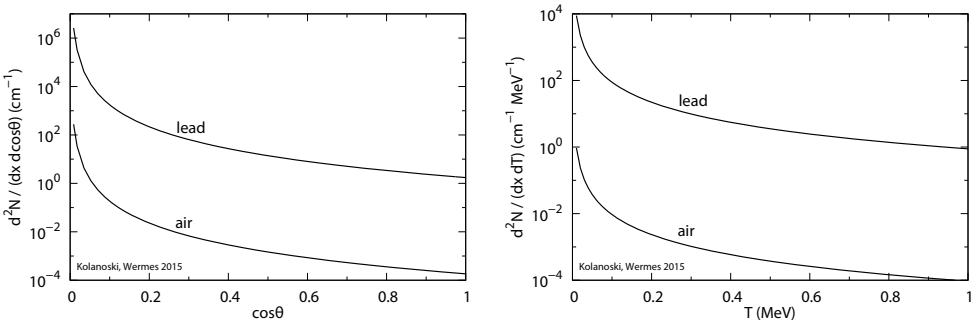


Fig. 3.10 Kinetic energy of δ electrons as a function of the emission angle for protons as projectile particles for different Lorentz factors γ. For the derivation elastic scattering off free electrons at rest has been assumed. This assumption is no longer justified for small energies of the scattered electron, that is, near $\theta = 90°$.



(a) δ electron rate as a function of emission angle. (b) δ electron rate as a function of kinetic energy.

Fig. 3.11 Angle and energy distributions for δ electrons in lead and in air. The distributions have been generated using (3.49) and (3.46), valid for highly relativistic projectiles and small kinetic energies of the electrons.

$$\frac{d^2N}{dx\,dT} = n_e \frac{d\sigma}{dT}\,. \tag{3.45}$$

With the electron density n_e in (3.12) and the differential cross section in (3.17) the differential rate of δ electrons per path length and energy becomes

$$\frac{d^2N}{dx\,dT} = \frac{1}{2}\,z^2\frac{Z}{A}\,K\,\rho\,\frac{1}{\beta^2}\,\frac{F(T)}{T^2}. \tag{3.46}$$

Here $K = 4\pi N_A r_e^2 m_e c^2 = 0.3071\,\mathrm{MeV\,g^{-1}\,cm^2}$ is the already known constant in the Bethe–Bloch formula (3.25). The function $F(T)$ from (3.17) takes care of the spin dependence and is of order 1, such that we assume $F(T) = 1$ in the following for reasons of simplicity.

Integration of (3.46) over T and x yields the number of δ electrons with energies between T_{min} and T_{max} in a medium of thickness Δx:

$$N = \frac{1}{2}z^2\frac{Z}{A}K\rho\,\Delta x\,\frac{1}{\beta^2}\left(\frac{1}{T_{min}} - \frac{1}{T_{max}}\right) \approx 0.077\,\frac{\mathrm{MeV\,cm^2}}{\mathrm{g}}\,z^2\rho\,\Delta x\frac{1}{T_{min}}\,. \tag{3.47}$$

In the last step we have used $\beta \approx 1$, $Z/A \approx 1/2\,\mathrm{mol/g}$ and $T_{max} \gg T_{min}$.

In the relativistic approximation we also obtain the emission angle dependence from (3.43):

$$\frac{dT}{d\cos\theta} = 4m_e c^2 \frac{\cos\theta}{\sin^4\theta} \,. \tag{3.48}$$

Putting this into (3.46) yields:

$$\begin{aligned} \frac{d^2N}{dx\, d\cos\theta} &= \frac{1}{2}\, z^2 \frac{Z}{A}\, K\, \rho\, \frac{1}{\cos^3\theta}\, \frac{1}{m_e c^2} \\ &\approx 0.15\, \frac{\mathrm{cm}^2}{\mathrm{g}}\, z^2\, \rho\, \frac{1}{\cos^3\theta}\,, \end{aligned} \tag{3.49}$$

assuming again $Z/A \approx 1/2\,\mathrm{mol/g}$. This expression diverges for $\theta \to 90°$, corresponding to $T \to 0$, an extreme case for which the approximation of a free electron (as chosen) is no longer valid. The energy and emission angle distributions obtained from (3.46) and (3.49) are displayed in fig. 3.11. The number of low energy δ electrons per solid angle scattered near 90° is larger by many orders of magnitude than those scattered in the forward direction. In spatially resolving detectors this leads to a broadening of ionisation clusters along a track and hence to a deterioration of the spatial resolution.

3.2.3 Statistical fluctuations of energy loss

3.2.3.1 Overview

The Bethe–Bloch formula determines the average energy loss per path length. In fact, however, the energy loss process has a statistical nature exhibiting fluctuations. The energy loss ΔE along a distance Δx is composed of many small contributions δE_n corresponding to individual ionisation or excitation processes:

$$\Delta E = \sum_{n=1}^{N} \delta E_n \,. \tag{3.50}$$

Statistical fluctuations occur for the number N of ionisation/excitation processes but also for the emitted energy δE in these processes. Both together produce fluctuations in the energy lost by a particle and also in the energy deposited in a material thickness Δx. These are commonly called *Landau fluctuations*.

Number fluctuations. In thin detectors, as often used to measure tracks of charged particles, Poisson statistics is relevant for the number of ionisations N. For example, in argon a minimum-ionising particle generates about 94 electron–ion pairs per cm (see table 7.1). The mean energy loss is $\langle \Delta E \rangle = 2.44\,\mathrm{keV/cm}$. This results in an average energy input to produce an electron–ion pair of $\langle \delta E \rangle = 26\,\mathrm{eV}$ and is distinctly larger than the minimum energy of 15.7 eV for ionisation. Besides ionisation, atomic excitations (threshold 11.6 eV), remaining mostly unobserved, contribute to the average energy demand for the production of one e^-–ion pair. The resolution on ΔE by a measurement of the number of e^-–ion pairs can be estimated using Poisson statistics. For a minimum-ionising particle passing through 1 cm of argon this is:

$$\frac{\sigma(\Delta E)}{\Delta E} \approx \frac{1}{\sqrt{N}} \approx 10\% \,. \tag{3.51}$$

In comparison, for semiconductors the averagy energy needed to create an electron–hole pair is much smaller, for example $\langle \delta E \rangle = 2.85\,\mathrm{eV}$ for Ge and $\langle \delta E \rangle = 3.65\,\mathrm{eV}$ for

Si (see chapter 8). Hence, for the same energy loss many more electron–hole pairs in a semiconductor are produced than electron–ion pairs in a gaseous detector and hence a better energy resolution can be reached (about three times better for Ge than for Ar at the same ΔE).

Energy transfer fluctuations. An important contribution to the total energy loss fluctuations originates from fluctuations in the amount δE of energy transferred in an individual process. The distribution of δE between a minimal and a maximal value δE_{min} and δE_{max} has a $1/(\delta E)^2$ shape (eq. (3.46) and fig. 3.11(b)). As discussed already, the minimum value δE_{min} is given by the minimum excitation/ionisation energy, the maximum value by (3.21). The most probable energy transfer (the maximum of the distribution) is in fact near δE_{min}, but from time to time large energies up to $\delta E_{max} \approx E$ are transferred in central collisions with atom electrons (see (3.21)) leading to a long tail of the distribution to high energies.

Influence on measurements. Energy loss fluctuations can influence the performance of detectors in an unfavourable way:

- The momentum resolution of a charged particle is deteriorated if the particle loses energy before or during the momentum measurement. Usually one can correct for the average energy loss using the Bethe–Bloch equation. The remaining fluctuations can sometimes be suppressed by additional measurements of energy depositions in the detector. Some possibilities how to suppress fluctuations are described on page 45 ff.
- For particle identification by measuring $\langle dE/dx \rangle$ the resolution to separate different particle species depends substantially on the width of the dE/dx distribution, see section 14.2.2 and the measurement example in fig. 3.7.
- Tracking detectors suffer a reduction in space resolution by the statistical fluctuations of the ionisation clusters along a track. Amongst other effects, the space resolution is also reduced because the knocked-on δ electrons move away from the track and there produce secondary ionisations. For this reason the produced electron–ion pairs most often appear in clusters: primary created electrons generate secondary ionisations nearby. For example, in argon only 1/3 of the 94 electron–ion pairs created per cm are primary ionisations. Since this has an essential impact on the statistical distribution of the ionisations along a track, it plays an important role for spatially resolving detectors (see chapters 7 and 8).

While for the momentum resolution (first item of the list) the fluctuations of the total energy loss play a role, independent of whether the fluctuation results from excitation or ionisation fluctuations, for both latter items only fluctuations in the measurable ionisations have an effect.

3.2.3.2 Landau–Vavilov distribution

The energy loss ΔE over a fixed distance Δx follows a probability density $f(\Delta E; \Delta x)$, which is normalised in an interval between minimal and maximal energy transfer:

$$\int_{\Delta E_{min}}^{\Delta E_{max}} f(\Delta E; \Delta x)\, d\Delta E = 1\,. \tag{3.52}$$

If in (3.50) the individual contributions δE_n to ΔE are statistically independent, the *central limit theorem* of statistics states that ΔE is normally distributed for $N \to \infty$

with a variance that is, N times the variance of the individual processes. This limit, however, is never reached for relativistic particles, as will become clear in the following.

Generally, the energy transfer fluctuations in individual collisions along the particle's path lead to an asymmetric distribution $f(\Delta E; \Delta x)$ which has a Gaussian part, corresponding to the many ionisation processes with small energy loss, and a tail to large energy loss values (see fig. 14.9 on page 554 or fig. 3.12). Large energy loss corresponds to the comparatively rarely occurring hard collisions in which much energy is transferred to individual electrons, that is, the δ electrons of section 3.2.2. We distinguish the *most probable value* (mpv) of the energy loss, that is, the maximum of the distribution, from the average value $\langle dE/dx \rangle$ situated to the right of the mpv by an amount depending on the asymmetry of the distribution.

The exact form of the distribution depends on the ratio of the mean energy loss, as given by the Bethe–Bloch formula (3.25), and the maximum energy loss (3.21). A measure for this ratio (but not the ratio itself) has been introduced by Vavilov with the parameter [956]:

$$\kappa = \frac{\xi}{T_{max}} \,, \tag{3.53}$$

where ξ is the factor in front of the logarithm of the Bethe–Bloch formula (3.25) multiplied by the path length Δx:

$$\xi = \frac{1}{2} K \frac{Z}{A} \rho \frac{z^2}{\beta^2} \Delta x \,. \tag{3.54}$$

The two limiting cases are:

κ large $\longrightarrow$ $f(\Delta E; \Delta x)$ a symmetric Gaussian distribution (for $\kappa \gtrsim 1$),
κ small $\longrightarrow$ $f(\Delta E; \Delta x)$ a very asymmetric distribution.

The first analytic form for the energy loss distribution has been derived by Landau in 1944 [640] for thin layers of material (corresponding to small κ values) making the following assumptions:

(i) T_{max} can become infinitely large corresponding to $\kappa \to 0$;
(ii) the electrons are treated as quasi free, that is, shell effects at small energy transfers are neglected;
(iii) the decrease of the particle's energy while penetrating the medium layer can be neglected.

The so derived *Landau distribution* is defined as the definite integral:

$$f_L(\lambda) = \frac{1}{\pi} \int_0^\infty \mathrm{e}^{-t \ln t - \lambda t} \sin(\pi t)\, dt \,. \tag{3.55}$$

This distribution (fig. 3.12) is very asymmetric with a tail towards $\lambda = \infty$; it has a maximum at $\lambda = -0.22278$ and a full width at half maximum (FWHM) of $\Delta\lambda = 4.018$. For computations it is available as a programmed function [625], for example in the program package ROOT [832]. The standard form of the Landau distribution (3.55) is transformed by scaling and shifting to a distribution that properly describes the observed energy loss of particles in matter. The correspondence between the energy loss ΔE and the variable λ is given by

$$\lambda = \lambda(\Delta E_w, \xi) = \frac{\Delta E - \Delta E_w}{\xi} - 0.22278. \tag{3.56}$$

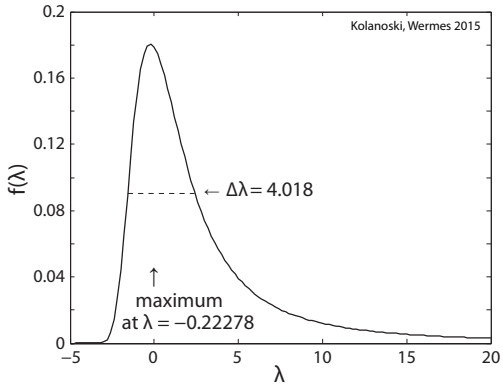


Fig. 3.12 Landau distribution as defined by (3.55). Also shown are the position of the maximum and the full width $\Delta\lambda$ at half maximum of the distribution.

Here ΔE_w is the most probable energy loss corresponding to the maximum of the Landau distribution at $\lambda = -0.22278$; ξ as defined in (3.54) characterises the width of the distribution. The FWHM is given by

$$W_{\mathrm{FWHM}} = \Delta\lambda\,\xi = 4.018\,\xi\,. \tag{3.57}$$

The most probable energy loss ΔE_w is a more stable energy loss measure than the mean value due to the strong fluctuations in the tail of the distribution (see fig. 3.13) and is given by the expression [746] (nomenclature as in (3.25)):

$$\Delta E_w = \xi\left(\ln\frac{2m_e c^2\beta^2\gamma^2}{I} + \ln\frac{\xi}{I} + j - \beta^2 - \delta\right) \xrightarrow{\gamma\gtrsim 100} \xi\left(\ln\frac{2m_e c^2\xi}{(\hbar\omega_P)^2} + j\right). \tag{3.58}$$

Here—in the high energy limit $\gamma \gtrsim 100$ and due to the density correction—the average excitation energy I is replaced by the plasma energy $\hbar\omega_P$ as was done also in (3.31). For the correction term j a value of $j = 0.2$ is given in [746]. Equation (3.58) shows that the most probable energy loss, normalised to the layer thickness, $\Delta E_w/\Delta x$, depends logarithmically on the layer thickness ($\xi \propto \Delta x$) due to the $\ln\xi$ dependence. This dependence is reproduced in fig. 3.13 by the three lower curves.

The Landau distribution can be approximated by the so-called *Moyal distribution* [965]:

$$f_M(\lambda) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-0.5\left(\lambda+\mathrm{e}^{-\lambda}\right)}\,. \tag{3.59}$$

The Moyal form is often used to fit 'Landau-like' asymmetric distributions by a simple formula. Note, however, that λ in (3.59) does not have the same physical meaning as in the Landau distribution (3.55) (a comparison of both is made e.g. in [625]). For the physical interpretation of measurements reference to the Landau distribution (3.55) is thus to be preferred.

The Landau distribution is a good approximation for small values of the parameter κ, roughly for $\kappa \lesssim 0.01$. A generalisation allowing a realistic maximum energy transfer T_{max}, hence being also valid for larger values of κ, has been given by Vavilov [956]. However, shell effects at small energy transfers are also neglected in Vavilov's work (Landau's assumption (ii) above). As a function of the Landau parameter λ the *Vavilov distribution* is defined via a Laplace transform (Laplace transforms in another context are discussed in appendix H) [831]:

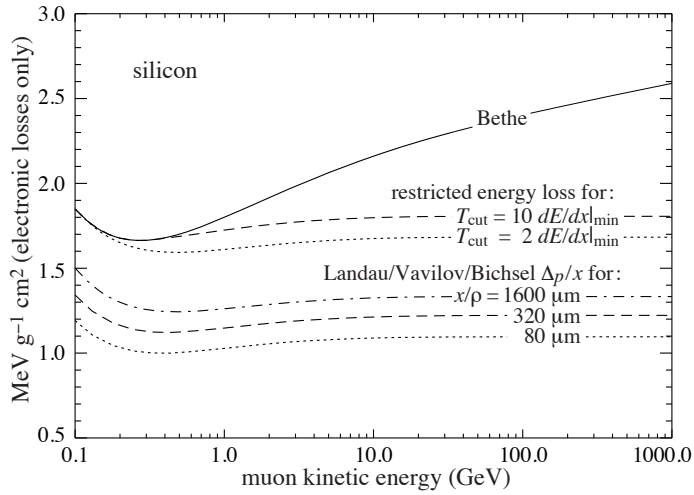


Fig. 3.13 Energy loss of muons in silicon (excluding radiation losses) [746]. The solid line is the mean energy loss of the Bethe–Bloch formula (3.25). The two curves below show the mean energy loss with the constraints on T_{cut}, the cut-off value for the energy transfer per collision, as described in the text (see page 45). The three bottom curves show the most probable energy loss per path length for silicon layers of various thicknesses. The shortcut notation Δ_p is the same as ΔE_w in (3.58).

$$p(\lambda;\kappa,\beta^2) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \phi(s)\mathrm{e}^{\lambda s}ds\,, \tag{3.60}$$

where

$$\phi(s) = \mathrm{e}^{C}\mathrm{e}^{\psi(s)}\,, \qquad C = \kappa(1+\beta^2\gamma_E)\,,$$
$$\psi(s) = s\ln\kappa + (s+\beta^2\kappa)\left(\int_0^1 \frac{1-\mathrm{e}^{\frac{-st}{\kappa}}}{t}\,dt - \gamma_E\right) - \kappa\,\mathrm{e}^{\frac{-s}{\kappa}}\,,$$
$$\gamma_E = 0.5772\ldots \qquad \text{(Euler's constant)}\,,$$

with s being a complex and c a real number. For small κ values the Vavilov distribution goes over to the Landau distribution and for large κ values to a Gaussian distribution. Validity over the entire kinematics and thickness ranges is achieved because the Vavilov distribution has two more parameters in addition to the two Landau parameters ΔE_w and ξ, namely κ and β^2. The Vavilov distribution is also available in the program package ROOT with the mathematical formulation documented on the corresponding web page [831].

Examples. For a 2 GeV muon a 1 m thick layer of water constitutes a 'moderate' thickness with $\kappa \approx 0.025$ (a typical scenario in airshower experiments like Auger and IceTop, section 16.4.3). For the same muon a 300 µm thick silicon layer has $\kappa \approx 1.7\times10^{-5}$. The examples in fig. 3.14(a,b) show that for small κ values the Landau and also the Vavilov distribution yield a reasonably good description; in fact for fig. 3.14(a) there is even no visible difference between the Landau and the Vavilov fit. Examining closer, however, one finds that for the very thin Si layer in fig. 3.14(a) both distributions underestimate the width somewhat.

Much larger values of κ are reached with low energetic heavy particles. For example, for a proton with a kinetic energy of 145 MeV when traversing 1 cm of water a value of $\kappa \approx 1$ results leading to an approximate Gaussian distribution of the energy loss, see fig. 3.14(d). In the limit of large κ values, the variance of the Gauss distribution is

$$\sigma^2 \approx \xi T_{max} \left(1 - \frac{\beta^2}{2}\right) . \tag{3.61}$$

The transition region with medium κ values is described by the Vavilov distribution (see the example for $\kappa = 0.27$ in fig. 3.14(c)). For $\kappa > 1$ Vavilov usually is the best description, mostly due to the fact that it has two more parameters than Landau and Gauss. Consequently, the resulting value for κ cannot always be interpreted as defined in (3.53) (see also fig. 3.14(d)).

For most practical usage a combination of Landau or Vavilov and Gauss distributions yields good approximations. There are, however, parameter regions in which these approximations fail, as shown e.g. in [199] for thin silicon layers. For this reason a wealth of theoretical and experimental investigations on energy loss distributions has been published after the first publications of Landau, Vavilov, and others. Their main focus was to improve the quantum mechanical treatment of the atomic bound states which become relevant at low energy transfers. Unfortunately, no simple formulae exist describing these more accurate results but rather computer programs must be employed to obtain the distributions.

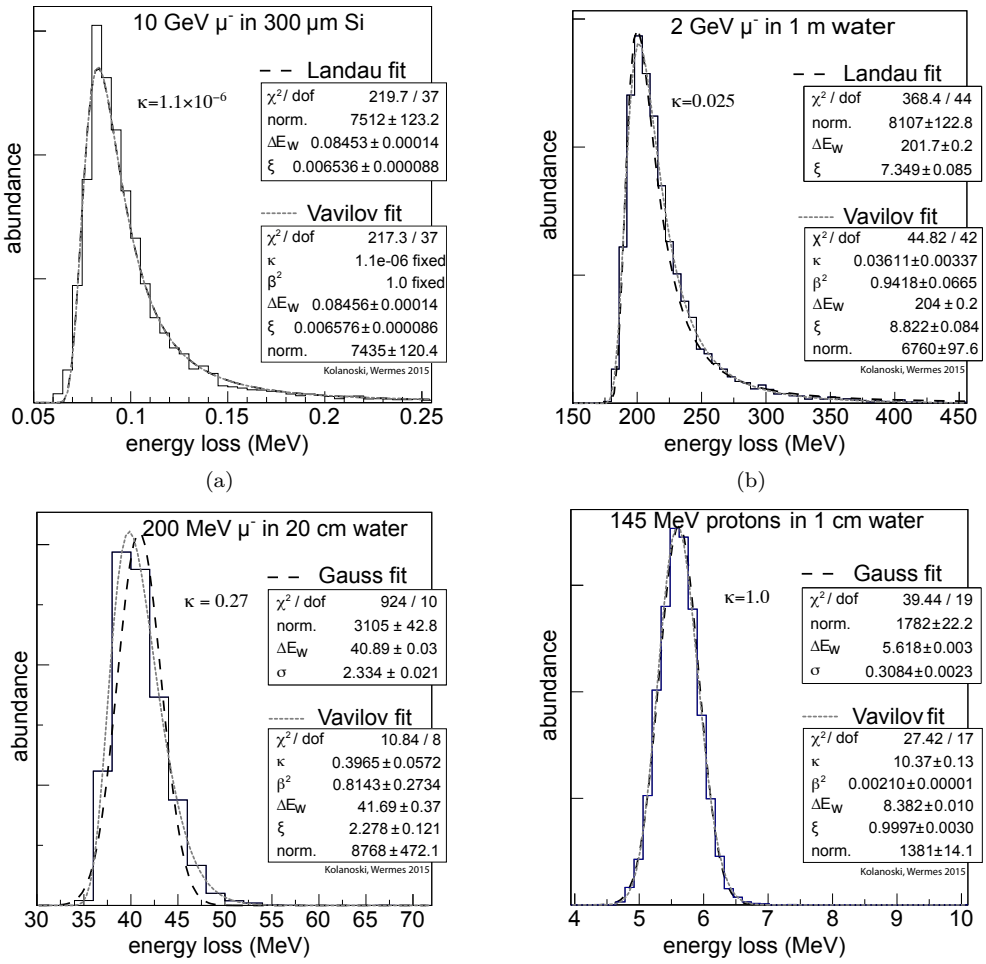
In general the energy loss of particles in detectors is computer-simulated. The universal simulation program Geant4 [452] covers all κ ranges with a remarkably simple model. The model assumes that there are only two excitation levels plus the continuum, that is, ionisation. The corresponding cross sections are adjusted such that the theoretical calculations and—if available—the data on energy loss distributions are correctly reproduced. Above an energy T_{cut} the δ electrons are traced individually. At large κ values a Gaussian approximation with a width as in (3.61) is used.

3.2.3.3 Suppression of fluctuations

As mentioned before, the strong fluctuations of the energy loss limit the resolution on energy-momentum measurements, the resolution of the particle identification via dE/dx measurements and the spatial resolution in detectors that measure space points by ionisation along a particle's path. The fluctuations can be reduced if the most probable value (maximum of the Landau distribution) can be used as an energy loss estimator rather than the average because the former is more stable with respect to individual fluctuations. From fig. 3.13 is evident that the average energy loss is fairly constant with increasing energy of the projectile particle if one restricts the maximally allowed energy loss in a collision to a value $T < T_{cut}$ (*restricted energy loss*, see (3.33)). The most probable energy loss has a similar dependence on the particle energy as the restricted energy loss and is thus a more stable quantity than the average energy loss which rises towards high energies (solid line 'Bethe' in fig. 3.13).

Hence the resolution can be improved by suppressing the fluctuations:

(a) if δ electrons can be detected and be excluded from the measurements, possible for instance in a bubble chamber or with many thin and independent detector layers;



	Particle, medium	T	d	κ	ΔE_w	ξ	σ_{gauss}
(a)	muon, silicon	10 GeV	300 μm	1.1×10^{-6}	83.7 keV	5.35 keV	-
(b)	muon, water	2 GeV	1 m	0.025	200 MeV	8.55 MeV	-
(c)	muon, water	200 MeV	20 cm	0.27	37.0 MeV	1.72 MeV	2.8 MeV
(d)	proton, water	145 MeV	1 cm	1.0	5.69 MeV	0.34 MeV	0.32 MeV

Fig. 3.14 Four examples of energy loss distributions for muons and protons after passing through a medium with thickness d (from simulations using the program package Geant4 [452]). The table shows the parameters of the distributions, ordered according to increasing κ values. The parameters κ, ΔE_w, ξ and $\sigma_{\text{Gauß}}$ are computed using (3.53), (3.58), (3.54) and (3.61), respectively. They can be compared with the fit parameters given in (a) and (b) for the Landau and Vavilov distributions and in (c) and (d) for the Vavilov and Gauss distributions. For the very small value of κ in (a) the Landau and Vavilov distributions are practically identical. The Gaussian and Vavilov fits in (d) can hardly be distinguished by eye, but the κ parameter needed for the Vavilov fit is much larger than the input value of κ (see comment in the text).

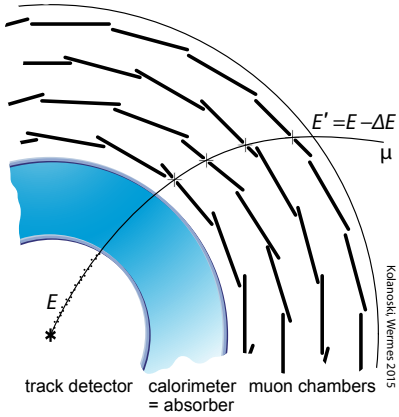


Fig. 3.15 Measurement of muon momenta in a collider detector. The sketch assumes the same magnetic field in the inner and outer tracking detectors. The energy loss, in particular in the massive calorimeter, leads to a more and more increasing curvature of the track.

(b) if the largest (possibly also the smallest) values of a series of measurements are discarded from the calculation of the mean (*truncated mean*) (see fig. 14.9 in section 14.2.2);

(c) if the very large energy deposits along a particle track can be measured and be used for a correction of the energy loss.

Regarding (c), to optimally reconstruct particle tracks for a precise momentum measurement one usually corrects for the energy loss along the particle's flight path. To do this one usually must take the average dE/dx lost over some distance of this path. The momentum resolution can, however, be improved if one is able to measure the actual energy lost (e.g. to δ rays).

An example for such a procedure is the reconstruction of muons in the ATLAS experiment at the LHC [4]. The muon's momentum can be measured stand-alone in the outer air toroid region of the detector in a large volume with magnetic field and with little material (fig. 3.15). Before reaching the muon spectrometer the muons traverse the inner tracking detector and the calorimeters. The largest amount of energy is lost in the calorimeters including about 17 cm of lead and 125 cm of iron in the central region. For this thickness we have $\kappa = 0.001$ for 100 GeV muons meaning that the

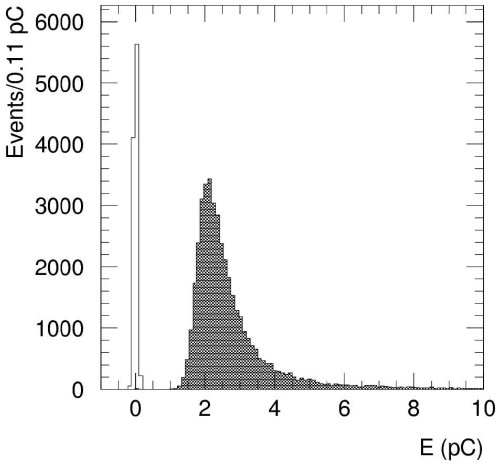


Fig. 3.16 Measurements of the energy loss by muons in the hadron calorimeter of the ATLAS detector [122] using 180 GeV muons in a test beam. The energy is given in units of the measured charge signal with 1 pC roughly corresponding to 1 GeV. The entries around 0 show the electronic noise with an equivalent width of about 40 MeV. The width of the energy loss distribution receives contributions from the dE/dx fluctuations, but also from the calorimeter resolution. The distribution can be described well by a Landau distribution folded with a Gaussian as shown in [341].

energy loss distribution is strongly asymmetric due to the large maximum energy loss T_{max}. The applicability of the central limit theorem, which would lead to a symmetric Gaussian distribution, is obviously not yet valid despite the very large number of individual ionisation processes. This is demonstrated by fig. 3.16 which is a measured distribution of the energy loss in the hadron calorimeter of ATLAS (≈125 cm iron) for 180 GeV muons ($\kappa = 0.0006$). Measuring the energy loss of the muons in the active part of the absorbing medium can improve the momentum resolution of muons significantly, as shown e.g. in [122].

3.2.4 Range

Due to the energy loss by ionisation the range R of a charged particle in matter has a specific value with only a small dispersion. In contrast, when absorption is the dominant interaction (e.g. for photons) the number of beam particles decreases exponentially (see fig. 3.2). If, on the other hand, absorption processes can be neglected for the passage of a charged particle beam through matter, as is the case for (heavy) charged particles, the number of beam particles remains constant up to a relatively abrupt drop to zero (fig. 3.17(b)). In fig. 3.17(b) a slow decrease by about 10% is observed, which can be explained by the absorption of beam protons by nuclear reactions. The brim at the end of the range is statistically smeared by range fluctuations (*straggling*). The width of the distribution in fig. 3.17(b) is smaller than 1 mm, that is, about 1% of the total range of about 77 mm. Because of the $1/\beta^2$ shape of the Bethe–Bloch curve at low energies (fig. 3.5), at the end of the particle's path much energy is lost over a relatively small distance (fig. 3.17(a)). This enhancement in deposited energy, called *Bragg peak*, is exploited for example in the radiation therapy of deep lying tumours (see page 51f).

In this section we study the range of particles in a medium assuming that they lose energy only by ionisation and excitation and that other processes, in particular nuclear reactions, can be neglected. In a sufficiently thick piece of material the particles come to rest when they have lost all of their initial kinetic energy T_0.

The range R results from the integration of the energy loss along the particle's path, always respecting that dE/dx is a function of the momentary kinetic energy T:

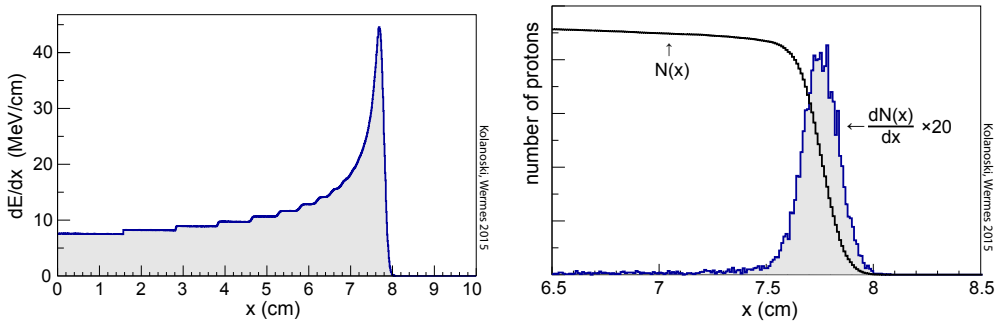
$$dT = \left\langle \frac{dE}{dx}(T) \right\rangle dx \quad \Rightarrow \quad dx = \left\langle \frac{dE}{dx} \right\rangle^{-1} dT \quad \Rightarrow \quad R = \int_{T_0}^{0} \left\langle \frac{dE}{dx} \right\rangle^{-1} dT\,. \tag{3.62}$$

Figure 3.18 shows the range of particles heavier than electrons for different media as a function of $\beta\gamma$. In this double logarithmic representation two slopes can be observed for all media (except for hydrogen) which cross in the $\beta\gamma$ region of the ionisation minimum. In the integration of (3.62) these slopes correspond to the $1/\beta^2$ behaviour of the energy loss at low energies and the $\ln\gamma$ rise after the minimum, respectively.

The ranges in fig. 3.18 are normalised to the masses of the projectile particles. The range proportionality M can be seen as follows: For a given medium dE/dx only depends on the velocity of the particle. Substituting

$$T = (\gamma - 1)\,Mc^2 \qquad \Rightarrow \qquad dT = Mc^2\,d\gamma \tag{3.63}$$

the range in (3.62) can be expressed as a function of γ_0, the γ value at the enter point of the medium:



(a) (b)

Fig. 3.17 Energy loss and range in water for 10,000 monochromatic protons with 100 MeV kinetic energy simulated with Geant4 [452]. (a) Energy loss per path length showing the Bragg peak at the end of the range at about 7.7 cm. Only the energy loss of the proton itself is shown and not the actual energy deposition; δ electrons and photons for example can carry energy away from the actual point where the energy was lost. The steps along the curves are due to discretisation in the simulation. The program adjusts the step lengths to the changes in energy loss such that near the Bragg peak the steps are very small whereas elsewhere they are larger. (b) Histograms of the number of remaining protons $N(x)$ and of the number of lost protons per unit length $dN(x)/dx$ as a function of penetration depth.

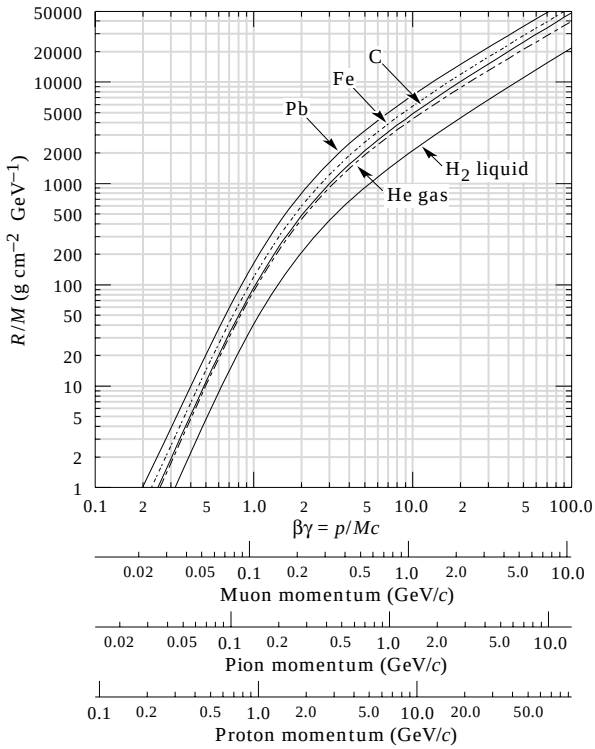


Fig. 3.18 Mass normalised range of particles in different media as a function of $\beta\gamma$ (from [746]), valid for masses M larger than the electron mass. For example, a charged kaon with mass $M = 0.494\,\mathrm{GeV}/c^2$ and momentum 700 MeV/c, i.e. $\beta\gamma = 1.42$, has in lead a range of $R/M = 396\,\mathrm{g\,cm^{-2}\,GeV^{-1}}\,c^2$, which for $\rho = 11.35\,\mathrm{g\,cm^{-3}}$ yields a range $R = 17.2\,\mathrm{cm}$.

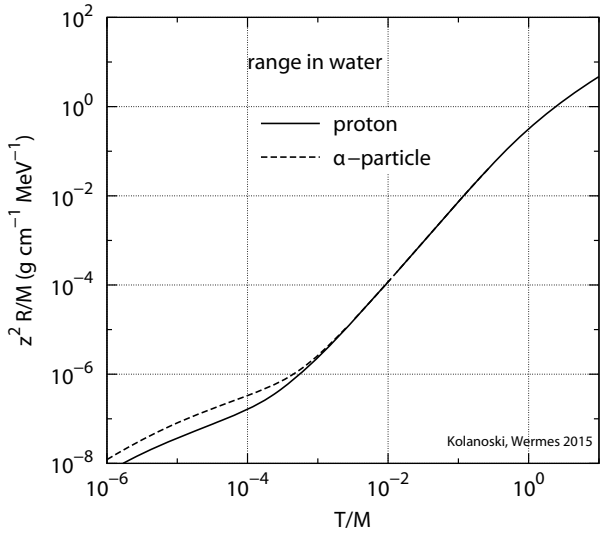


Fig. 3.19 Scaled ranges (after eq. (3.65)) of protons and α particles in water. The curves have been computed using the programs PSTAR and ASTAR [182]. For protons the abscissa corresponds to kinetic energies of about 1 keV to 10 GeV, for α's the energy scale is shifted by a factor of about 4, corresponding to the α/p mass ratio. For proton energies of about 1 MeV to almost 1 GeV both curves perfectly coincide following the power law of (3.66).

$$R = Mc^2 \int_{\gamma_0}^{1} \left\langle \frac{dE}{dx} \right\rangle^{-1} d\gamma = \frac{M}{z^2} f(\gamma_0)\,. \tag{3.64}$$

For a given medium, the function $f(\gamma_0)$ is independent of the mass and the charge of the projectile particle as is also evident from fig. 3.18 showing the range in dependence on $\beta\gamma = p/Mc$. Using instead of γ_0 the kinetic energy $T_0 = (\gamma_0 - 1)Mc^2$, one obtains a scaling relation for ranges of different particles in the same medium:

$$\frac{z^2}{M} R(T) = \frac{z^2}{M} R\left(T' = T\frac{M}{M'}\right)\,. \tag{3.65}$$

This range scaling law corresponds to the energy loss scaling law of (3.34).

In fig. 3.19 the range of protons and α particles in water, scaled as in (3.65), is plotted against the mass normalised kinetic energy. At not too low energies both curves lie indeed on top of each other and can be described by a common power law:

$$\frac{z^2}{M} R = 0.4 \left(\frac{T}{Mc^2}\right)^{1.75} \mathrm{g\,cm^{-2}} \qquad \text{in the range} \qquad 10^{-3} \lesssim \frac{T}{Mc^2} \lesssim 1\,. \tag{3.66}$$

For electrons the energies for which the range follows a simple power law is much more restricted than the range for heavier particles, because the minimum of the ionisation curves already occurs at $T \approx 600\,\mathrm{keV}$. The ranges for electrons shown in fig. 3.20 for kinetic energies between 10 keV and 1 GeV can indeed be described by a power law up to 600 keV (for water $R \propto T^{1.67}$). Over the entire region shown in fig. 3.20 all curves can be described by a second order polynomial in $\log R$ and $\log T$ with only slightly varying coefficients for media between water and lead:

$$\log_{10} R = a\,(\log_{10} T)^2 + b\,\log_{10} T + c\,. \tag{3.67}$$

With R in $\mathrm{g\,cm^{-2}}$ and T in MeV the coefficients are:

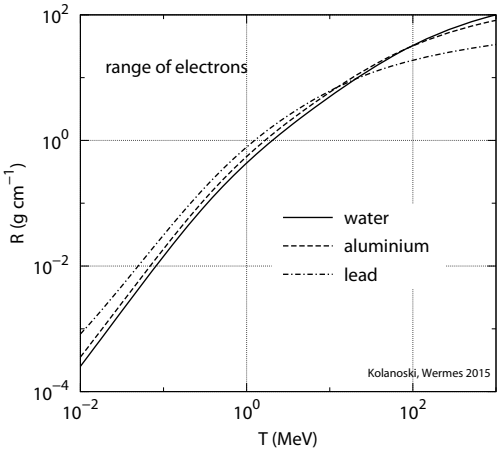


Fig. 3.20 Range of electrons as a function of their kinetic energy in three media (water, aluminum, lead) computed with the programm ESTAR [559]. The curves are very similar in the region between 10 keV and 1 GeV and can be well described by fitting second order polynomials of the logarithms of the variables. The parametrisation of the curves is given in (3.67) and (3.68).

	a	b	c
water	−0.170	1.30	−0.40
aluminium	−0.176	1.25	−0.29
lead	−0.197	1.13	−0.15

(3.68)

In nuclear physics, range measurements are used to determine the energy of protons, α particles, and other nuclei. Figure 3.21 shows tracks of α particles from a radioactive source, measured in a cloud chamber. The tracks have an abrupt ending since they have the same energy. For radiation protection the range of a certain radiation in matter is important to know (table 3.2). Heavier particles reach less far than lighter ones but have a higher ionisation density. As an example, α radiation can damage inner organs only if α emitters are inhaled or otherwise incorporated.

The particularly high energy deposit at the end of the range (see fig. 3.17(a)), just before the particle's stop, is exploited very successfully [89, 858] for radiation treatment of deep lying tumours, first applied with heavy ions already in 1975 at LBL in Berkeley [858]. Modern examples of such developments leading to treatment facilities trace back to dedicated research at the Research Centre GSI in Darmstadt [490],

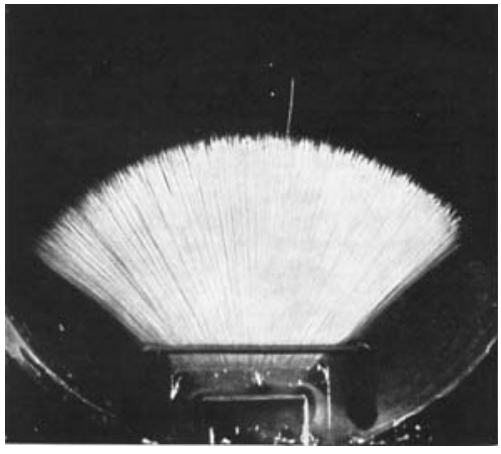


Fig. 3.21 Tracks of α particles from a radioactive decay appearing in a cloud chamber (adapted from [410], reprinted with kind permission by Springer Science+Business Media, see also [781]). The range of the particles has a sharp cut-off, which means that all α particles have the same energy. There is one track with a longer range originating from an excited nuclear state.

Table 3.2 Range of electrons, protons and α particles in air (dry, see level) and water (computed with the interactive programs ESTAR, PSTAR and ASTAR [182]). Given are the CSDA ranges (CSDA = continuous slowing down approximation), obtained by continuously summing up all energy losses according to (3.62). Due to multiple scattering (section 3.4) the projected distances between the start and the end of the track are shorter than the actual length. The reduction factor $f_{proj/csda}$ is given for protons and α particles. For electrons the projected path is ill-defined because of their strong scattering.

Particles	Energy (MeV)	Range (m)			
		air	$f_{proj/csda}$	water	$f_{proj/csda}$
electrons	0.1	1.35×10^{-1}		1.43×10^{-4}	
	1.0	$4.08 \times 10^{+0}$		4.37×10^{-3}	
	10	$4.31 \times 10^{+1}$		4.98×10^{-2}	
protons	0.1	1.53×10^{-3}	0.884	1.61×10^{-6}	0.907
	1.0	2.38×10^{-2}	0.988	2.46×10^{-5}	0.991
	10	$1.17 \times 10^{+0}$	0.998	1.23×10^{-3}	0.998
α particles	0.1	1.38×10^{-3}	0.829	1.43×10^{-6}	0.865
	1.0	5.56×10^{-3}	0.951	5.93×10^{-6}	0.961
	10	1.09×10^{-1}	0.997	1.13×10^{-4}	0.997

see fig. 3.22(a), which has spun off for example the HIT facility at the Deutsches Krebsforschungszentrum DKFZ in Heidelberg [528], or the HIMAC facility at NIRS in Japan [525]. Compared to a treatment with X-rays, the healthy tissue traversed before reaching the tumour is damaged much less (see fig. 3.22(b) and fig. 3.22(c)). In addition, the tumour volume can be scanned precisely by ion beam control and beam energy dependent range variation. Some depth–dose profiles are compared in fig. 3.22(c).

The relative biological effectiveness (RBE) of ions near the Bragg peak increases with their mass and is larger for carbon ions (^{12}C) than for protons or helium ions (^{4}He), for example. For still heavier ions than carbon RBE generally increases, also in the beam entrance region. A therapeutic optimum is reached at medium ion masses such as carbon which is often used in therapies. The physical reasons trace back to less straggling, laterally because for the same range protons must have lower momenta than ^{12}C ions and thus suffer more multiple scattering (see eq. (3.102) on page 67), and longitudinally due to the larger width of the energy-loss distribution (eq. (3.61)). For details we refer to [858]. Cases exist, however, for which protons have turned out to be a very reasonable alternative, not least due to their easier availability (e.g. in applications for eye tumours). In other areas of irradiation of cancerous tissue, irradiations with photons or β rays can be most effective or even just more practicable, for example in the case of leukaemia or if irradiation is needed during an operation.

3.3 Energy loss through bremsstrahlung

So far we have treated the energy loss of charged particles when penetrating matter as caused by their interaction with the atomic shell (excitation and ionisation). Charged particles can, however, also lose energy by radiating electromagnetic quanta, predominantly in the Coulomb field of the nucleus. As shown in fig. 3.23, the process can be

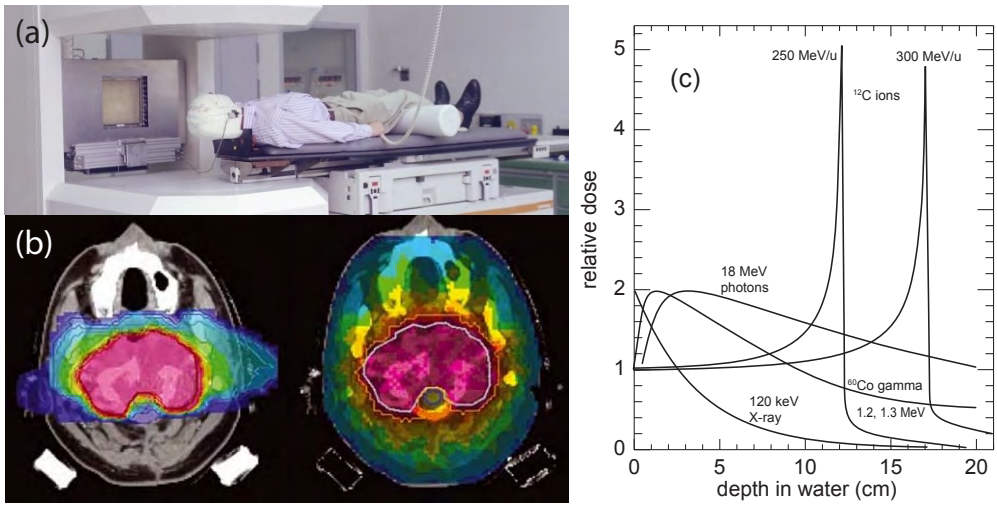


Fig. 3.22 Tumour therapy using ion beams exploits that the fact that maximum energy deposit occurs at the end of the penetration range forming the *Bragg peak*. (a) Patient with brain tumour in the radiation gantry (photo: A. Zschau, GSI). (b) Dose profiles around a brain tumour (the tumour is marked by contour lines): (left) radiation using ^{12}C ions, (right) radiation using gamma rays (source: O. Jäkel, GSI). The colour scale shows the relative radiation dose (blue = minimum, red = maximum). The comparison shows that the surrounding tissue is better spared from damage in ion beams (from two beam directions) than using γ rays (from nine beam directions). Also using γ rays, the applied dose can be better concentrated on the target volume by changing the beam direction ('conformation therapy'), however, the results are by far better for ion radiation. (c) Depth profiles of the dose applied by carbon ions compared to those of X-ray and γ radiation (source: U. Weber, GSI). In ion therapy the dose maximum is distributed over the tumour volume by varying the energy and position of the beam (as is shown in (b)) such that the tumour is irradiated as homogeneously as possible. For irradiation by photons the beam intensity decays exponentially due to absorption. The dose profile, however, shows an initial buildup, the maximum moving more inward with increasing energy. This is because the photon-induced flux of ionising electrons is first built up following the cross section for electron production by photons.

viewed as Rutherford scattering with simultaneous radiation of a photon. In section 3.4 we will discuss that upon scattering off a heavy nucleus the directional change of the projectile particle can no longer be neglected as was the case for scattering off the atomic shell.

3.3.1 Radiation from accelerated charges

In classical electrodynamics the energy emitted per unit time from an accelerated charge is computed as

$$\frac{dW}{dt} = \frac{2}{3}\frac{z^2e^2}{4\pi\epsilon_0 c^3}\,|\ddot{\vec{x}}\,|^2 \;\propto\; \frac{z^4Z^2e^6}{m^2}\,, \tag{3.69}$$

where for the right-hand side $\ddot{x} \propto zZe^2/m^2$ is inserted considering a particle with charge ze at a fixed distance r being decelerated by the field of a nucleus with charge

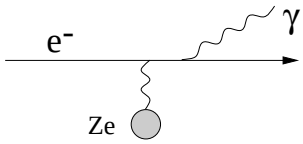


Fig. 3.23 Bremsstrahlung due to the interaction of a charged particle with the Coulomb field of a nucleus.

Ze ($m\ddot{x} = zZe^2/r$). The energy loss hence is inversely proportional to the squared mass of the decelerated particle. A computation employing relativistic quantum electrodynamics yields a proportionality to Z^2E/m^2 of the radiated power (see e.g. [564]). The Z^2 dependence is characteristic for coherent scattering processes off a nucleus, in contrast to ionisation.

The characteristic E/m^2 dependence is the reason that for energies below some 100 GeV energy loss through bremsstrahlung is only significant for electrons and positrons. For electrons in lead bremsstrahlung loss exceeds ionisation energy loss starting already at 7 MeV, in air at about 100 MeV, reflecting the Z^2 dependence. On the other hand this also means that low energy electrons from radioactive sources with energies in the 0.5–2 MeV range mainly lose energy by ionisation and can thus be used to characterise ionisation detectors. For particle momenta occurring in LHC reactions, however, bremsstrahlung also becomes relevant for muons and pions.

In what follows we will focus on bremsstrahlung of electrons (including positrons) with relativistic energies ($E \gg mc^2$). At the end of this section bremsstrahlung of high energy muons will also be discussed.

3.3.2 Energy spectrum and radiation length

3.3.2.1 Energy spectrum of bremsstrahlung

The quantum-electrodynamical analogue to the classical radiation field is the radiation of individual photons with energies E_γ. The quantum mechanical calculation of bremsstrahlung of an electron in the field of a heavy point-like nucleus was first carried out by Bethe and Heitler in 1934 [189]. Later the calculation was improved by different authors. Reviews can be found in [517, 619, 948, 886].

For the exact computation of photon radiation of electrons in realistic materials various approximations must be made:

Charge screening of the nucleus: Since different momentum transfers are involved in the radiation, form factors must be considered for the charge distributions inside the nucleus and for the atomic shell. It turns out that the influence of nuclear form factors can be neglected since the corresponding momentum transfers to the nucleus are small. The electric form factor of the atom, however, turns out to be very important: the shell causes a screening of the nuclear potential which is the stronger the smaller the momentum transfer.[7] The minimum momentum transfer at which the screening effect sets in for a given energy E of the primary electron and energy E_γ of the radiated photon can be estimated [517]. The momentum transfer q is minimal when the momenta of the incoming and outgoing particles are (anti-)parallel. Using the absolute values of the 3-momenta $p,\ p',\ k$ of the in- and outgoing

[7] Note that the influence of the atomic form factor on the spectrum has just the opposite dependence on the momentum transfer than the nuclear form factor. In a classical picture the reason for this is that, when entering the nucleus, the charge seen by the scattering process effectively decreases, whereas upon immersing into the negative charge cloud of the shell the positive charge of the nucleus becomes effectively larger.

electron and the photon one obtains

$$q_{min} = p - p' - k = \frac{1}{c}\left(\sqrt{E^2 - m_e^2c^4} - \sqrt{(E-E_\gamma)^2 - m_e^2c^4} - E_\gamma\right) \approx \frac{m_e^2c^3E_\gamma}{2E\,(E-E_\gamma)}\,. \tag{3.70}$$

We have only considered the 3-momentum transfer q_{min} since the heavy nucleus practically does not receive any energy transfer. In addition, at high energies the momenta p, p' can be replaced by the energies E, $E-E_\gamma$. The approximation on the right-hand side of (3.70) is valid for E and $E-E_\gamma \gg m_ec^2$. Quantum mechanically, the spatial uncertainty corresponding to the momentum transfer, is the formation length or coherence length:

$$l_f = \frac{\hbar}{q_{min}}\,. \tag{3.71}$$

When this length reaches atomic dimensions an increasingly larger part of the entire atomic shell takes part in the interaction and the nucleus is more and more shielded by the shell. As a measure of the amount of screening the ratio of the mean atom radius and the formation length is considered:

$$\eta = \frac{r_{\text{atom}}}{l_f} = \frac{r_{\text{atom}}\,q_{min}}{\hbar}\,. \tag{3.72}$$

In the Thomas–Fermi model (see e.g. [120]) the mean atom radius is given by $r_{\text{atom}} \approx 137\,\hbar/(m_ec\,Z^{\frac{1}{3}})$ and the screening parameter becomes [517]

$$\eta = \frac{137\,m_ec^2\,E_\gamma}{2E\,(E-E_\gamma)\,Z^{1/3}}\,. \tag{3.73}$$

Numbers other than 137 in the expression for η also appear in the literature. Most of the time the two extreme cases 'no screening' ($\eta \gg 1$) and 'complete screening' ($\eta = 0$) are discussed. Exact calculations usually lie between these extremes. The screening effect is more relevant at higher particle energies because the momentum transfers needed for a given radiated energy E_γ are smaller for higher projectile energies as (3.70) shows. In practice one can use the approximation of 'complete screening' starting already at particle energies of some 10 MeV. This is also a valid and good assumption for most cases discussed in this book.

Coulomb correction: In the Born approximation particles are described by in- and outgoing planar waves. The influence on the wave function of the charged particle by the Coulomb fields of the atoms is taken care of by a 'Coulomb correction'.

Dielectric suppression: Bremsstrahlung photons with very small energies are absorbed in the material due to the polarisability of the medium and a resulting loss of coherence [937, 105] leading to an effective cut-off of the low energy part of the photon spectrum.

Scattering off shell electrons: In realistic media one must also consider bremsstrahlung in the field of the shell electrons. It turns out that the influence of the shell can approximately be accounted for by the substitution $Z^2 \to Z(Z+1)$.

LPM effect: At very high energies, roughly above 1 TeV, bremsstrahlung (and also pair production, section 3.5.5) is suppressed by the so-called Landau–Pomeranschuk–Migdal effect (LPM effect). This effect is particularly important for astroparticle

Table 3.3 Numerical values for the radiator functions L_{rad} and L'_{rad}, after [948].

Z	L_{rad}	L'_{rad}
1	5.31	6.144
2	4.79	5.621
3	4.74	5.805
4	4.71	5.924
> 4	$\ln\left(184.15\, Z^{-1/3}\right)$	$\ln\left(1194\, Z^{-2/3}\right)$

physics, but also starts to become important in today's high energy accelerators. We omit this effect in this section, but will describe it in chapter 15 on page 590 and appendix G in the context of the discussion of highly energetic electromagnetic showers.

In the Born approximation—but including screening and Coulomb corrections—the differential cross section for an atom to radiate a photon with energy E_γ, integrated over the emission angle is (see e.g. [948]):

$$\frac{d\sigma}{dE_\gamma} = \frac{\alpha r_e^2}{E_\gamma}\left\{\left(\frac{4}{3} - \frac{4}{3}y + y^2\right)\left[Z^2\left(\phi_1(\eta) - \frac{4}{3}\ln Z - 4f(Z)\right) + Z\left(\psi_1(\eta) - 8\ln Z\right)\right]\right.$$
$$\left. + \frac{2}{3}(1-y)\left[Z^2\left(\phi_1(\eta) - \phi_2(\eta)\right) + Z\left(\psi_1(\eta) - \psi_2(\eta)\right)\right]\right\}. \qquad (3.74)$$

Here r_e is the Bohr electron radius and $y = E_\gamma/E$ the fraction of the electron energy taken by the photon. Other nomenclature used in (3.74) will be defined and commented on below. The Coulomb correction $f(Z)$ is given in [342] as a series expansion in $a = \alpha Z$, with α being the fine structure constant:

$$f(Z) = a^2 \sum_{n=1}^{\infty} \frac{1}{n(n^2+a^2)} \approx a^2\left(\frac{1}{1+a^2} + 0.20206 - 0.0369a^2 + 0.0083a^4 - 0.002a^6\right). \qquad (3.75)$$

For heavy nuclei this correction amounts to up to about 10%. The screening effects are contained in the functions ϕ_1, ϕ_2 for scattering off the nucleus and in ψ_1, ψ_2 for scattering off individual electrons, each with factors Z^2 for the nucleus and Z for the electrons, respectively. These functions have been computed by different authors and for different atomic models used. In what follows, we will give approximations for cross section computations.

In the case of complete screening ($\eta = 0$), that is, for high energies (see page 54) the differential cross section can be given in the following form [948]:

$$\frac{d\sigma}{dE_\gamma} = \frac{4\alpha r_e^2}{E_\gamma}\left\{\left(\frac{4}{3} - \frac{4}{3}y + y^2\right) \times \right.$$
$$\left. \left[Z^2\left(L_{rad}(Z) - f(Z)\right) + ZL'_{rad}(Z)\right] + \frac{1}{9}(1-y)(Z^2+Z)\right\}. \qquad (3.76)$$

Here $\phi_1(0) - \phi_2(0) = \psi_1(0) - \psi_2(0) = 2/3$ has been used. The so-called radiator funtions $L_{rad}(Z)$ and $L'_{rad}(Z)$ are defined as

$$L_{rad} = \frac{1}{4}\phi_1(0) - \frac{1}{3}\ln Z\,,$$

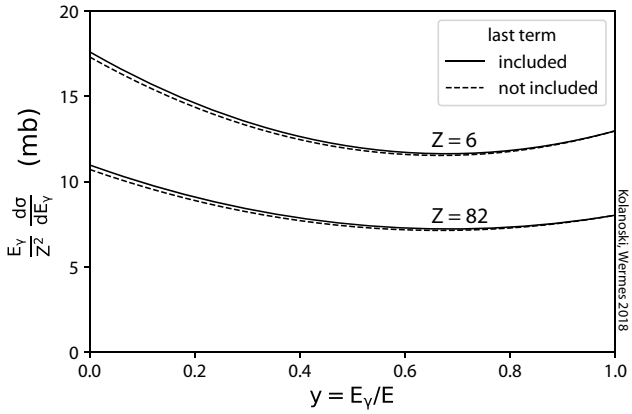


Fig. 3.24 Bremsstrahlung spectra for carbon ($Z=6$) and lead ($Z=82$) according to (3.76), but normalised by a factor E_γ/Z^2. The dashed curves do not include the last term in (3.76).

$$L'_{rad} = \frac{1}{4}\psi_1(0) - 2\ln Z\,.$$

Numerical values are given in table 3.3.

In fig. 3.24 bremsstrahlung spectra according to formula (3.76), multiplied by E_γ/Z^2, are displayed for carbon ($Z=6$) and lead ($Z=82$). The curves are similar in shape, but the fact that they differ in height, even after normalising by a factor E_γ/Z^2, demonstrates that the cross section increases somewhat less than with Z^2. The small difference between the full and the dashed curves means that the last term in (3.76) is quite small (we will learn below that it is ignored for the definition of the radiation length). Note that in the adopted high energy limit the curves do not depend on the primary energy E.

Approximation (3.76) holds for high electron energies with the exception of the kinematic regime of highest photon energies (corresponding to high momentum transfers to the atom). For non-asymptotically high electron energies and not too small photon energies ($E_\gamma > 50\,\text{MeV}$) the cross section (3.74) can be computed using the following approximate formulae [592]:

$$\phi_1(\eta) = \begin{cases} 20.867 - 3.242\,\eta + 0.625\,\eta^2, & \eta \le 1 \\ 21.12 - 4.184\,\ln(\eta + 0.952), & \eta > 1\,, \end{cases} \tag{3.77}$$

$$\phi_2(\eta) = \begin{cases} 20.029 - 1.930\,\eta + 0.086\,\eta^2, & \eta \le 1 \\ \phi_1(\eta), & \eta > 1\,, \end{cases} \tag{3.78}$$

$$\psi_{1,2}(\eta) = \frac{L'_{rad}(Z)}{L_{rad}(Z) - f(Z)}\left(\phi_{1,2}(\eta) - \frac{4}{3}\ln Z - 4f(Z)\right) + 8\ln Z\,. \tag{3.79}$$

The characteristic properties of the bremsstrahlung cross section (3.76) are (a) the Z^2/m_e^2 dependence ($r_e \propto 1/m_e$) which already showed up in the classical formula (3.69), and (b) the (leading order) dependence on the photon energy:

$$\frac{d\sigma}{dE_\gamma} \propto \frac{1}{E_\gamma}\,. \tag{3.80}$$

The divergence at $E_\gamma = 0$ has no practical effect because the lowest frequencies are absorbed by the atoms (by the dielectric suppression mentioned above). In realistic

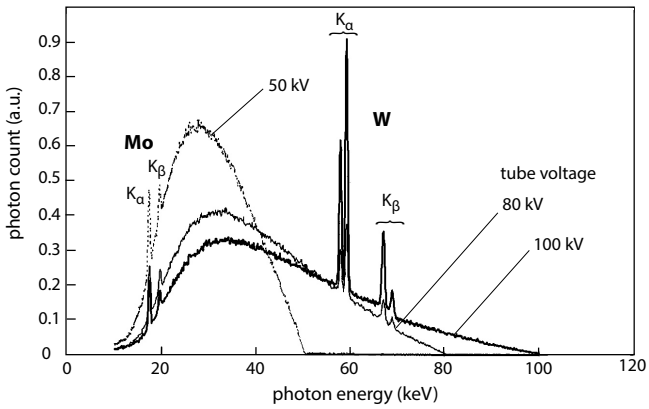


Fig. 3.25 Measured bremsstrahlung spectra of an X-ray tube for different tube voltages, normalised to unit area. X-ray transitions with electrons returning to the K-shell cause lines on top of the continuous bremsstrahlung spectrum, characteristic for the cathode materials, here tungsten with molybdenum traces, filtered with 1.2 mm Al equivalent. Adapted from [193].

simulations the radiation of small photon energies up to a cut-off energy can be treated as a continuous energy loss. Real photons are created only above this cut-off energy (see e.g. [452]).

3.3.2.2 Example: X-ray tube spectra

In X-ray tubes photons with X-ray energies are produced from electron bremsstrahlung emitted from high-Z cathode materials as targets. A typical cathode material is tungsten featuring $Z = 74$, a high melting point and good thermal conductivity for efficient cooling. X-ray photons absorbed still in the cathode can cause electrons to be emitted from inner atomic shells. The vacancies are subsequently filled by electrons from outer-shells emitting characteristic electromagnetic radiation in the X-ray regime corresponding to the levels' energy separation. Figure 3.25 shows measured X-ray tube spectra for three different tube voltages [193]. Superimposed on the (continuous) bremsstrahlung spectrum are *characteristic lines* due to transitions from higher shells (L, M, N) to the K shell. The otherwise linearly falling bremsstrahlung spectrum is at low energies absorbed by filters (1.2 mm Al equivalent), cutting off also the L-lines. Instead of the X-rays emitted as a result of the electron transition the energy can also be transmitted to shell electrons thus being removed from the atom with characteristic energies (Auger electrons). Corresponding electron spectra are discussed in section 3.5.3 and shown in fig. 3.40.

3.3.2.3 Angular distribution of bremsstrahlung

With increasing energy $E = \gamma m_e$ of the incoming electron the photons are emitted into an increasingly narrower cone in the forward direction with a characteristic opening angle

$$\theta_\gamma \approx \frac{1}{\gamma} = \frac{m_e c^2}{E} \,. \tag{3.81}$$

A computation of the angular distribution of the emitted photons is very elaborate and requires to make approximations depending on the kinematical regime considered (see e.g. [180, 619, 592]).

Small-angle approximation. As an example we present here the double differential cross section for bremsstrahlung by electrons as derived by Schiff [861]. The

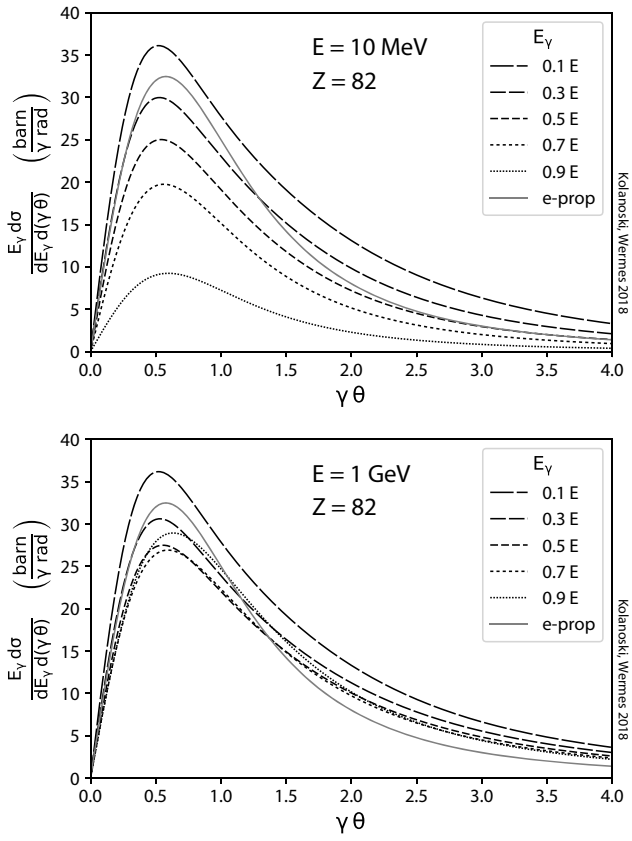


Fig. 3.26 Bremsstrahlung angular distribution of an electron passing through lead ($Z=82$) with energies $E=10\,\text{MeV}$ and $1\,\text{GeV}$. The plotted cross sections (multiplied by E_γ) are differential in the radiated photon energy E_γ and in the 'reduced angle' $\gamma\theta_\gamma$ where θ_γ is the polar angle of the radiated photon with respect to the initial electron direction and $\gamma = E/m_e$ is the Lorentz factor of the electron. Each plot shows curves for different energies E_γ of the radiated photon. In addition the grey curve represents the coarse approximation by the electron propagator term (3.84) with arbitrary normalisation.

calculation was made in the Born approximation with the additional approximations for small polar angles, for high initial electron energies and for complete shielding (which is appropriate for small radiation angles, corresponding to small momentum transfers, see above). In the small-angle approximation, where $\sin\theta_\gamma \approx \theta_\gamma$, the angle appears only in the product $x = \gamma\,\theta_\gamma$. Therefore it is convenient to substitute the variable θ_γ by x. Otherwise we use the variables as defined above with the addition of the outgoing electron's energy $E' = E - E_\gamma$ to write Schiff's differential cross section [861]:

$$\frac{d\sigma}{dE_\gamma\,dx} = \frac{4\alpha\,Z^2\,r_e^2}{E_\gamma}\,x \times \tag{3.82}$$
$$\times\left[\frac{16x^2E'}{(x^2+1)^4E} - \frac{(E+E')^2}{(x^2+1)^2E^2} + \left(\frac{E^2+E'^2}{(x^2+1)^2E^2} - \frac{4x^2E'}{(x^2+1)^4E}\right)\ln M(x)\right]$$
$$\text{with}\qquad \frac{1}{M(x)} = \left(\frac{E_\gamma}{2\gamma\,E'}\right)^2 + \left(\frac{Z^{1/3}}{C\,(x^2+1)}\right)^2 \quad\text{and}\quad C \approx 111\,.$$

In the compendium by Koch and Motz [619] this formula, labelled 2BS, is compared to a collection of other approximations and is also plotted for different electron and photon energies (fig. 6 of [619]).

Schiff's differential cross section (3.82), multiplied by the radiated photon energy E_γ, is plotted for different initial electron energies in fig. 3.26. In each plot the curves

for different E_γ are similar in shape and even of similar height due to the multiplication with E_γ removing the $1/E_\gamma$ dependence. This similarity becomes stronger with increasing electron energy. Most of the photons are radiated under small polar angles of the order of $1/\gamma$. As one can see in fig. 3.26 the most probable angle is found around $\theta_{\gamma,mpv} \approx 1/(2\gamma)$.

The leading dependence of the differential cross section with respect to the radiation angle θ_γ originates from the electron propagator (inner electron line) in the bremsstrahlung diagram of fig. 3.23. The propagator has the form $1/(q_e^2 - m_e^2)$ where q_e is the 4-momentum of the virtual electron. Denoting p'_e and k the 4-momenta of the outgoing electron and the photon, respectively, we get with $q_e = p'_e + k$ and $\beta = |\vec{p}'_e|c/E'$:

$$q_e^2 - m_e^2 = m_e^2 + 2kk_e - m_e^2 = 2E_\gamma E' - 2E_\gamma |\vec{p}'_e| c\cos\theta = 2E_\gamma E'(1 - \beta\cos\theta)\,. \quad (3.83)$$

For small photon energies as compared to the electron energies the angle θ is approximately θ_γ and β is about the value for the incoming electron. The propagator enters the cross section quadratically so that we can approximately write

$$d\sigma \propto \frac{d\cos\theta_\gamma}{(1-\beta\cos\theta_\gamma)^2} \approx \frac{4\gamma^4\,\theta_\gamma\,d\theta_\gamma}{(1+\theta_\gamma^2\gamma^2)^2}\,. \quad (3.84)$$

The right-hand side of (3.84) holds for small angles and $\beta \approx 1$. This propagator contribution is also shown in fig. 3.26. Remarkably, this very coarse approximation reflects quite well the dominant features of the bremsstrahlung angular distribution at small angles.

Large-angle approximation. The large-angle approximation has been calculated in kinematic regimes where shielding can be neglected by Hough [543] (also discussed in the compendium [619]). In [573] numerical values of double-differential cross sections of some selected elements are compiled for angles above 15° at different primary and radiated energies. In order to give a quantitative feeling for the bremsstrahlung contribution at large angles we quote the cross section for carbon integrated over the angles at $E = 500\,\text{MeV}$ and $E_\gamma = 100\,\text{MeV}$. Interpolating the data tabulated in [573] and integrating over the range $\theta_\gamma \geq 15°$ a cross section $d\sigma/dE_\gamma \approx 80\,\text{nb/MeV}$ is obtained. The same cross section integrated over the full angular range is $d\sigma/dE_\gamma \approx 5.4\,\text{mb/MeV}$ according to (3.76) and fig. 3.24. Although the large-angle cross section is thus many orders of magnitude less than the full cross section, in specific experiments, where the forward bremsstrahlung is often excluded by design, the large-angle bremsstrahlung may be of a similar order of magnitude as the processes the experiment is aiming for.

3.3.2.4 Radiation length

Averaging the radiated energies over the bremsstrahlung spectrum in (3.76) yields the mean energy loss per path length:

$$\begin{aligned}\frac{dE}{dx} &= \frac{N_A\,\rho}{A}\int_0^E E_\gamma \frac{d\sigma}{dE_\gamma} dE_\gamma \\ &= 4\alpha r_e^2 \frac{N_A\,\rho}{A}\,E\left[\left(Z^2(L_{rad} - f(Z)) + ZL'_{rad}\right) + \frac{1}{18}(Z^2+Z)\right]\,, \end{aligned} \quad (3.85)$$

where for the upper integration bound $E - m_e c^2 \approx E$ has been used. At high energies, the energy loss per unit path length thus approximately scales with the electron energy. One can then introduce the *radiation length* X_0 as the characteristic length for energy loss through bremsstrahlung by:

$$\left(\frac{dE}{dx}\right)_{rad} = -\frac{E}{X_0}\,. \tag{3.86}$$

Integration of (3.85) yields:

$$E(x) = E_0\, \mathrm{e}^{-\frac{x}{X_0}}\,, \tag{3.87}$$

meaning that after a path length $x = X_0$ an electron on average possesses only $1/e$ of its initial energy; the fraction $1 - 1/e \approx 63\%$ has been radiated off.

In [948] the radiation length X_0 is defined by:

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A\,\rho}{A}\left[Z^2(L_{rad} - f(Z)\,) + ZL'_{rad}\right]\,. \tag{3.88}$$

In this definition the term $\frac{1}{18}(Z^2 + Z)$ in (3.85), taking values between 1% and 1.7% depending on Z, has been neglected.[8] With numerical values for the constants $(4\alpha r_e^2\, N_A)^{-1} = 716.408\,\mathrm{mol\,cm^{-2}}$ we obtain:

$$\rho\, X_0 = \frac{716.408\, A\ \mathrm{mol/g}}{Z^2\left[(L_{rad} - f(Z)) + ZL'_{rad}\right]}\ \frac{\mathrm{g}}{\mathrm{cm}^2}\,. \tag{3.89}$$

Radiation lengths of various materials often occurring in particle physics experiments are listed in table 3.4.

A compact formula, obtained from a fit and deviating by less than 2.5% from the input data for all elements except for He (deviation $\approx 5\%$), is given in [724]:

$$\rho\, X_0 = \frac{716.408\, A\ \mathrm{mol/g}}{Z(Z+1)\,\ln\frac{287}{\sqrt{Z}}}\ \frac{\mathrm{g}}{\mathrm{cm}^2}\,. \tag{3.90}$$

In this representation the dependence on Z can be seen better: in addition to the Z^2 dependence for coherent scattering off the entire nucleus there is the same contribution proportional to Z originating from the scattering off individual shell electrons which has to be summed over.

For mixtures and compounds the radiation length can be calculated from the corresponding contributions X_{0i} of the individual components with weight fraction g_i:

$$\frac{1}{\rho X_0} = \sum_i \frac{g_i}{\rho X_{0i}} \qquad \text{with} \qquad g_i = \frac{\rho_i}{\rho}\,. \tag{3.91}$$

The radiation length X_0 generally serves as an important measure to characterise electromagnetic processes taking place in the Coulomb field of a nucleus, this is, apart from bremsstrahlung, above all e^+e^- pair production (section 3.5.5) and multiple Coulomb scattering (section 3.4). For example, X_0 is used in detector building to characterise the thickness of material traversed on average by a particle. Note, however,

[8] The term is neglected in order to be able to also express the cross section for pair production (section 3.5.5) at high energies by means of the radiation length.

Table 3.4 Radiation length (given as ρX_0 in units mass/area and in units of length at the given density) and critical energy E_c (electrons), E_c^μ (muons), as well as the correspondingly relevant atomic properties for various materials often used in detector applications (adapted from [762], the data for photoemulsions correspond to Ilford G5 [889]).

Material	Z	$\langle Z/A \rangle$ (mol/g)	Density (g/cm^3)	X_0 (g/cm^2)	X_0 (cm)	E_c (MeV)	E_c^μ (GeV)
			elemental solids				
Be	4	0.444	1.85	65.19	35.2	113.7	1328
C (graphite)	6	0.500	2.21	42.70	19.3	81.7	1057
Al	13	0.482	2.70	24.01	8.9	42.7	612
Si	14	0.499	2.33	21.82	9.36	40.2	582
Fe	26	0.466	7.87	13.84	1.76	21.7	347
Cu	29	0.456	8.96	12.86	1.44	19.4	317
Ge	32	0.441	5.32	12.25	2.30	18.2	297
W	74	0.403	19.30	6.76	0.35	8.0	150
Pb	82	0.396	11.35	6.37	0.56	7.4	141
U	92	0.387	18.95	6.00	0.32	6.7	128
			scintillators				
NaI	11,53	0.427	3.67	9.49	2.59	13.4	228
CsI	55,53	0.416	4.51	8.39	1.86	11.2	198
BaF_2	56,9	0.422	4.89	9.91	2.03	13.8	233
$PbWO_4$	82,74,8	0.413	8.30	7.39	0.89	9.6	170
polystyrene	1,6	0.538	1.06	43.79	41.3	93.1	1183
			gases (20 °C, 1 atm)				
H_2	1	0.992	0.084×10^{-3}	63.04	750 500	344.8	3611
He	2	0.500	0.166×10^{-3}	94.32	568 200	257.1	2352
air	7,8	0.499	1.205×10^{-3}	36.62	30 390	87.9	1115
Ar	18	0.451	1.66×10^{-3}	19.55	11 777	38.0	572
Xe	54	0.411	5.48×10^{-3}	8.48	1 547	12.3	232
			other materials				
H_2O (liquid)	1, 8	0.555	1.0	36.1	36.1	78.3	1029
standard rock	11	0.500	2.65	26.5	10.0	49.1	693
photoemulsion			3.82	11.33	2.97	17.4	286

that the definition of the radiation length bares some level of arbitrariness since the published formulae usually rely on some approximations made. Definitions other than those given here and in [746] do exist, based on different approximations, for example in [835]. For the mentioned rough calculations this does usually not play a major role. If more exact computations are needed, Monte Carlo simulations are usually employed which use more precise formulae.

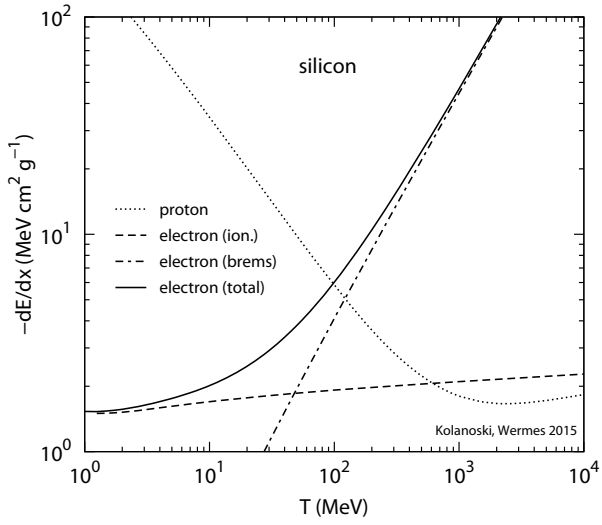


Fig. 3.27 Energy loss by ionisation and by bremsstrahlung for electrons as a function of the electron kinetic energy. Both contributions (dashed lines) intersect at the critical energy $E_c = 47.87\,\text{MeV}$, defined as in (3.92). In the Rossi-PDG definition (3.93) this value is $E_c = 40.19\,\text{MeV}$ [762]. For comparison also the ionisation energy loss for protons is given. All curves have been computed using the interactive programs ESTAR and PSTAR [182].

3.3.3 Critical energy

Energy loss by radiation and by ionisation have different dependencies on the energy E and the mass M of the particle, as well as on the nuclear charge Z of the medium:

$$\begin{aligned} &\text{ionisation:} && \propto\ Z\,\ln E/M \\ &\text{bremsstrahlung:} && \propto\ Z^2\,E/M^2\ . \end{aligned}$$

Due to the different energy dependencies ionisation dominates at low energies, bremsstrahlung at high energies. A *critical energy* E_c is defined as the point in energy at which both energy loss curves intersect (fig. 3.27):

$$\left(\frac{dE}{dx}(E_c)\right)_{ion} = \left(\frac{dE}{dx}(E_c)\right)_{rad}\ . \tag{3.92}$$

Different to (3.92) the Particle Data Group [746] has adopted the definition by Rossi [835] according to which the critical energy is the energy at which the ionisation energy loss after one radiation length equals the energy of the radiating electron:

$$\left(\frac{dE}{dx}(E_c)\right)_{ion} X_0 = -E_c\ . \tag{3.93}$$

This definition merges into (3.92) under the assumption that $(dE/dx)_{rad} \approx -E_c/X_0$. For (3.93) reasonably good approximations are available, differing for solid and liquid compared to gaseous media due to the density effect:

$$\begin{aligned} E_c &\approx \frac{610\,\text{MeV}}{Z+1.24} \quad \text{(solids and liquids)}\,, \\ E_c &\approx \frac{710\,\text{MeV}}{Z+0.92} \quad \text{(gases)}\,. \end{aligned} \tag{3.94}$$

Radiation length X_0 and critical energy E_c are important parameters for the development of electromagnetic showers (see chapter 15). Numerical values for some materials

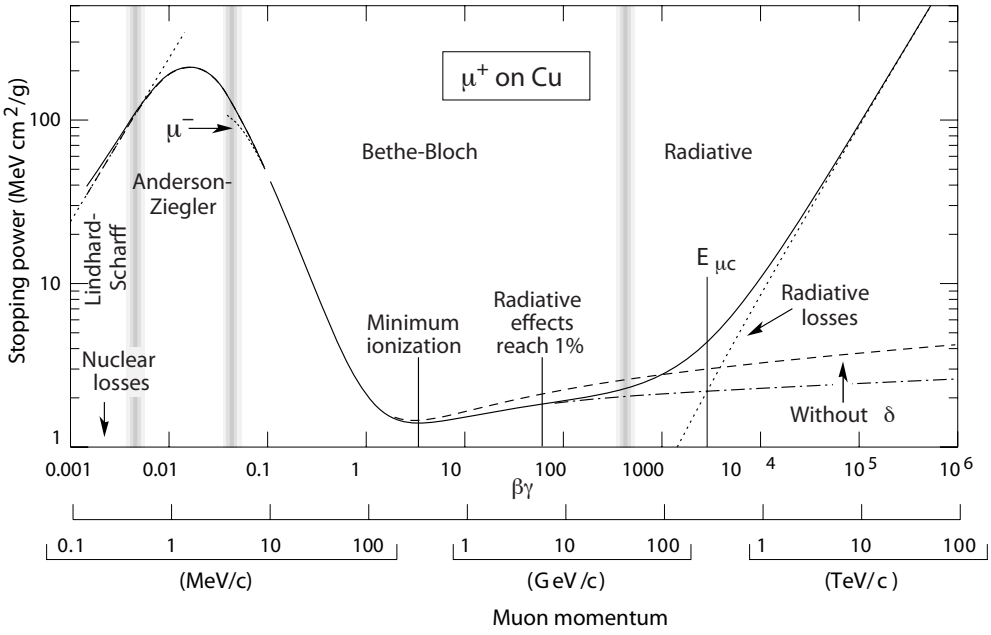


Fig. 3.28 Energy loss for positive muons in copper (taken from [746]; see also detailed description and references therein). The vertical bands separate regions with different theoretical descriptions. The centre marks the validity region of the Bethe–Bloch formula (3.25). At high energies where bremsstrahlung dominates the drawn curves show the total energy loss (solid), the loss by radiation (dotted) as well as by ionisation without (dashed) and with (dash-dotted) the inclusion of the δ term in (3.25). The region at low energies labelled 'Anderson–Ziegler' is described by an empirical fit to measured data. The region below this, labelled 'Lindhard–Scharff', is based on a theoretical model which yields an energy loss dependence roughly proportional to β (dashed line). At the lowest energies non-ionising energy loss caused by elastic nuclear recoil dominates. At low energies the energy loss is smaller for negative than for positive particles ('Barkas effect' [153]) attributed to higher order corrections and indicated here by the curve labelled μ^-.

are given in table 3.4.

3.3.4 Energy loss of high energy muons

The energy loss by bremsstrahlung scales with the inverse squared mass of the radiating particle as $1/m^2$. At very high energies, the radiation contribution to the energy loss hence also plays a role for particles heavier than electrons, first for muons and pions, at still higher energies also for kaons and protons.

In the following we consider the energy loss of high energy muons because of its relevance at modern high energy accelerators in the TeV range (like LHC) and also for the detection of cosmic rays. The energy loss can be represented as a sum of the contributions from ionisation and from bremsstrahlung (fig. 3.28):

$$-\frac{dE}{dx} = a(E) + b(E)\,E\,. \tag{3.95}$$

Here a is the energy loss contribution from ionisation and $b\,E$ the contribution from

Table 3.5 Parameters a and b for the linear approximation (3.95) of the energy loss of muons [309]. The column density is given here in mwe = 'metre water equivalent', that is, the thickness of an equivalent water layer in metres (mwe = $100\,\mathrm{g/cm^2}$). The third column contains the critical energy for muons [762] (for different materials see table 3.4).

Medium	a (GeV/mwe)	b (10^{-3}/mwe)	E_c^μ (GeV)
air	0.281	0.347	1115
ice (≈water)	0.259	0.363	1031
standard rock (see text)	0.223	0.463	693

bremsstrahlung. The critical energy E_c^μ at which both contributions are equal is defined by [482]:

$$a(E_c^\mu) = b(E_c^\mu)\, E_c^\mu \,. \tag{3.96}$$

Values for E_c^μ for different media are given in [762] together with tables containing the energy dependence of a and b. At high energies (above the critical energy) a and b can be taken as constant. This has been shown for example in [309] by fits to simulation results.[9] In table 3.5 fitted parameters for air, ice, and 'standard rock' are listed. 'Standard rock' with the medium properties as given in table 3.4 is used as a reference medium for underground experiments (see section 2.3.3 and chapter 16).

The energy dependence of the energy loss (3.95) allows us to measure the energy of muons above the critical energy without a magnetic field. This is particularly interesting for experiments with a large detector volume that cannot be completely covered magnetically. Examples are neutrino detectors like IceCube [36] detecting muons generated by high energy neutrinos (see section 16.6.5).

Integrating the energy loss (3.95) assuming constant a, b, an energy dependent range for muons with primary energy E can be determined using (3.62):

$$R(E) = \frac{1}{b} \ln\left(1 + \frac{b}{a}\, E\right) \,. \tag{3.97}$$

The range of muons plays an important role for shielding against cosmic radiation in underground experiments. In 'standard rock', for example, muons with an energy of 1 TeV have a range of about 0.9 km and in water or ice of about 2.4 km.

3.4 Multiple Coulomb scattering

Charged particles are scattered in the Coulomb fields of nuclei according to the Rutherford cross section. Relative to this effect, the deflection by interaction with shell electrons are small at energies $E \gg m_e c^2$.

After passage of a particle through a material of thickness Δx the particle has in general undergone multiple scattering processes leading to a statistical distribution of the scattering angle relative to the original direction.

Coulomb scattering off a single nucleus is described by the Rutherford cross section (fig. 3.29):

$$\left.\frac{d\sigma}{d\Omega}\right|_{\text{Rutherford}} = z^2 Z^2 \alpha^2 \hbar^2 \frac{1}{\beta^2 p^2} \frac{1}{4\sin^4(\theta/2)} \,. \tag{3.98}$$

[9] Note, however, that the critical energy cannot be determined using constant a, b in (3.96).

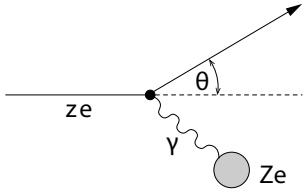


Fig. 3.29 Scattering of a charged particle (charge ze) in the Coulomb field of a nucleus with charge Ze (Rutherford scattering).

Here θ is the scattering angle, z, β and p are charge, velocity and momentum of the scattered particle, and Z is the charge of the nucleus. Given the predominantly small momentum transfers (corresponding to the large reach of electromagnetic interactions) the process is a coherent scattering process off the total charge of the nucleus resulting in the quadratic Z-dependence already mentioned (by contrast the energy loss by ionisation is proportional to Z since all shell electrons contribute as an incoherent sum). If the scattered particle is light compared to the mass of the nucleus it is deflected with only a small energy transfer to the nucleus.

For a not too thin scattering medium, that is, if the number of scatters exceeds about 20, we speak of multiple or Molière scattering. According to the *central limit theorem* of statistics one expects for an infinite number of scatters that the distribution of the scattering angle is Gaussian. The general case of a finite number of scatters has been treated by Molière [710,711]. For most practical cases the distribution obtained by the Molière theory can be approximated well by a Gaussian distribution. One should keep in mind, however, that the Molière distribution predicts larger probabilities for large multiple scattering angles which results from the nature of the underlying Rutherford scattering processes. Figure 3.30 shows a measured scattering angle distribution of protons [198] compared with the exact Molière theory and with the approximation

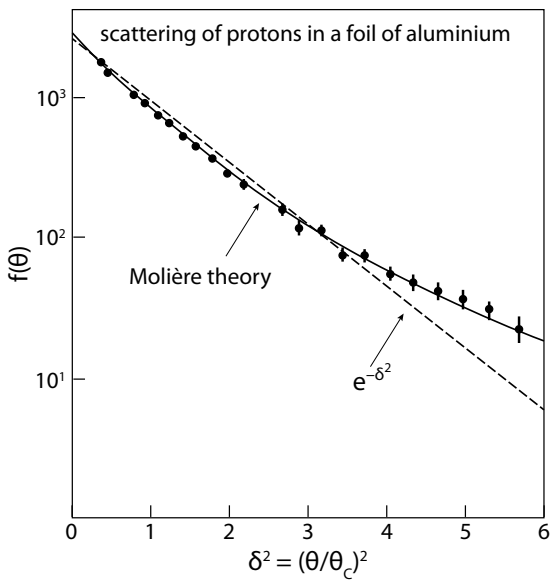
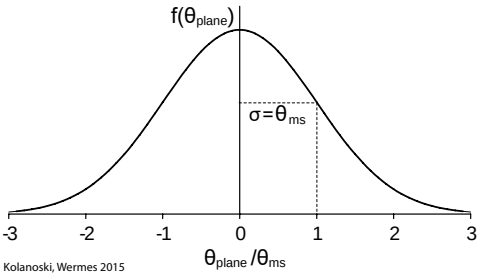
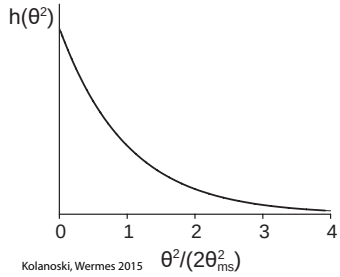


Fig. 3.30 Distribution of the scattering angle of protons with kinetic energy $T = 2.18\,\mathrm{MeV}$ having traversed an aluminium foil with area mass density of $x = 3.42 \times 10^{-3}\,\mathrm{g\,cm^{-2}}$ (i.e. thickness of about 13 µm) [198]. The measurements are compared with the exact Molière theory and also with a Gaussian (data and curves taken from [198]). In this logarithmic representation plotted against the squared scattering angle the Gauss distribution is a straight line. The scattering angle is normalised here to the characteristic angle θ_c appearing in the Molière theory ($\theta_c = \chi_c\sqrt{B}$ with parameters χ_c and B as given in [711], $\theta_c \approx \theta_{ms}\sqrt{2}$). Thus with $\delta = \theta/\theta_c$ the leading dependence of the Molière distribution at small angles is proportional to $\exp(-\delta^2)$.



(a) Distribution of the plane projected scattering angle as defined by (3.103).



(b) Distribution of the squared spatial scattering angle as defined by (3.104).

Fig. 3.31 Gaussian approximation of scattering angle distributions.

of a Gaussian distribution. In this logarithmic representation it is particularly obvious that the tails of the distribution at large scattering angles cannot be described by a Gaussian.

In the small-angle Gaussian approximation the multiple scattering angular distribution is specified by one parameter, the standard deviation θ_{ms} (often denoted θ_0 in the literature) of the angle θ_{plane} projected onto a plane perpendicular to the direction of motion of the incoming particle:

$$f(\theta_{plane})d\theta_{plane} = \frac{1}{\sqrt{2\pi}\,\theta_{ms}} \exp\left(-\frac{\theta^2_{plane}}{2\theta^2_{ms}}\right) d\theta_{plane}\,. \tag{3.99}$$

A simple formula for the parameter θ_{ms} has been estimated by Rossi and Greisen [834] from the variance of the scattering angle distribution as

$$\theta_{ms} \approx \sqrt{\langle\theta^2_{plane}\rangle} \approx \frac{E_s}{\sqrt{2}\,p\,c\,\beta}\sqrt{\frac{x}{X_0}}\,, \tag{3.100}$$

where the scale is specified by

$$E_s = m_e\,c^2\sqrt{\frac{4\pi}{\alpha}} = 21.2\,\text{MeV}\,. \tag{3.101}$$

In (3.100) x is the thickness of the scattering material and X_0 the medium's radiation length. The radiation length appears here again because it characterises processes in the Coulomb field of a nucleus as also the case for bremsstrahlung of electrons in the field of a nucleus (section 3.3) for which X_0 has been defined in (3.88).

A better approximation for the parameter θ_{ms} is obtained by the so-called 'Highland formula' [522] given here in the parametrisation of [679], adapted also by the Particle Data Group [746]:

$$\theta_{ms} = \frac{13.6\,\text{MeV/c}}{p\beta}z\sqrt{\frac{x}{X_0}}\left(1 + 0.038\ln\frac{x}{X_0}\right)\,. \tag{3.102}$$

The spatial scattering angle is the angle between the particle direction before and after the scattering volume and is composed of the two orthogonal projections θ_x, θ_y: $\theta_{space} =: \theta \approx \sqrt{\theta^2_x + \theta^2_y}$ in the small-angle approximation, where x, y are orthogonal to

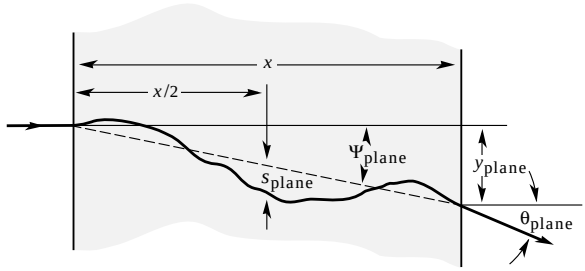


Fig. 3.32 Sketch illustrating multiple scattering of a charged particle in a scattering layer projected onto a plane perpendicular to the incoming particle direction as well as characteristic quantities used to describe this multiple Coulomb scattering process (source: PDG [746]).

the direction of motion. The θ distribution is obtained as the product of the statistically independent distributions for θ_x and θ_y, each following (3.99):

$$g(\theta)d\theta = \frac{1}{\theta_{ms}^2} \exp\left(-\frac{\theta^2}{2\,\theta_{ms}^2}\right)\theta\, d\theta\,. \tag{3.103}$$

Here we used $d\theta_x\, d\theta_y = \theta\, d\theta\, d\phi$ and integrated over ϕ. Note that the spatial scattering angle can only assume positive values, in contrast to the projection of this angle onto a plane (fig. 3.31 and fig. 3.32). It is thus appropriate to examine the probability density of θ^2 which with $2\theta d\theta = d\theta^2$ becomes

$$h(\theta^2)d\theta^2 = \frac{1}{2\theta_{ms}^2} \exp\left(-\frac{\theta^2}{2\,\theta_{ms}^2}\right)d\theta^2\,. \tag{3.104}$$

The distributions (3.103) and (3.104) are normalised to unity in $0 \le \theta < \infty$. The maximum of $h(\theta^2)$ is at $\theta = 0$ and corresponds to the most probable scattering angle in a solid angle interval $\theta\, d\theta\, d\phi$, whereas the average value is at $\langle\theta\rangle = \sqrt{\pi/2}\,\theta_{ms} > 0$. The standard deviation of the spatial scattering angle with respect to the most probably value at $\theta = 0$ is

$$\sqrt{\langle\theta^2\rangle} = \sqrt{2}\theta_{ms}\,. \tag{3.105}$$

The angle $\sqrt{2}\theta_{ms}$ approximately corresponds to the characteristic Molière angle θ_c in fig. 3.30.

For computer simulations of particles traversing a detector the traversed matter is sliced into scattering layers and the relevant scattering quantities are calculated for each layer. The sketch in fig. 3.32 illustrates the quantities describing the effects of multiple scattering after traversing a layer of material with thickness x: the rms[10] averages of the effective scattering angle θ_{plane}, the deflection angle ψ_{plane}, the offset from the original entrance point y_{plane} and the central offset s_{plane}. The relation to the scattering angle parameter θ_{ms} is given by [746]:

$$\theta_{plane}^{rms} = \theta_{ms}\,, \quad \langle\psi_{plane}\rangle = \frac{1}{\sqrt{3}}\,\theta_{ms}\,, \quad \langle y_{plane}\rangle = \frac{1}{\sqrt{3}}\,x\,\theta_{ms}\,, \quad \langle s_{plane}\rangle = \frac{1}{4\sqrt{3}}\,x\,\theta_{ms} \tag{3.106}$$

[10] rms = root-mean-square; square root of the mean squares of a set of numbers x_i, that is, $\sqrt{\langle x^2\rangle}$ (the same as the standard deviation of the x_i if the mean is zero) which is also defined for continuous random variables. Frequently also the terms 'rms deviation' or 'rms error' are used for the rms of the differences between values $\hat{x}_i$ predicted by a model or an estimator and the observed values x_i, that is, $\sqrt{\langle(x-\hat{x})^2\rangle}$.

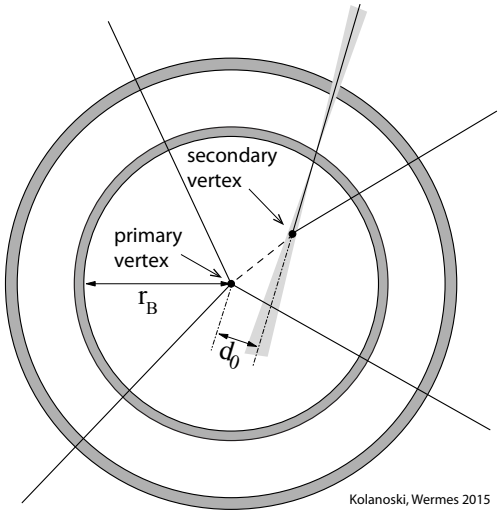


Fig. 3.33 Influence of multiple Coulomb scattering on the reconstruction of secondary vertices. The direction of a track is measured by two layers of a detector (e.g. silicon strip detectors). The scattering at the first detector layer leads to a reconstruction error when extrapolating the track to the primary vertex (indicated by the shaded area).

Table 3.6 Error of the determination of the impact parameter d_0 of a 5 GeV pion assuming a beam pipe with radius 5 cm and a thickness of 1 mm.

Material	X_0 (mm)	x/X_0	θ_{ms} (rad)	Δd_0 (μm)
Al	89	0.011	0.0003	15.0
Be	353	0.003	0.00015	7.5

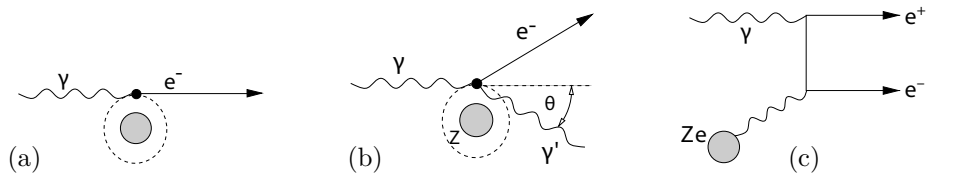
These approximations are precise to about 5% for layer thicknesses of $10^{-3} < x/X_0 < 10$; this is the range of about 5 μm to 50 mm for lead or 0.3 mm to 3 m for air.

Example: Measurement of secondary vertices in a collider experiment. Decays of hadrons with charm and bottom quarks can successfully be detected by identifying their decay vertices which often are sufficiently separated from the collision's primary vertex (see also sections 8.1, 9.4.6 and 14.6.2). Bound states containing b-quarks, for example, have lifetimes of $\tau \approx 1.5$ ps. At a momentum of 10 GeV the average decay length then is $\langle l \rangle = \gamma\beta c\tau \approx 2$ mm.

Whether a track can come from a secondary vertex is indicated by the impact parameter d_0 (see also chapter 14, section 14.6) which is the distance of closest approach to the primary vertex of an extrapolated track in the plane perpendicular to the beam (fig. 3.33). The error of a measurement of d_0 (grey fanned area in fig. 3.33) determines if a track can be separated sufficiently well from the primary vertex.

Let us assume the radius of a beam pipe in which beam collisions take place be $r_B = 50$ mm and the first layer of a very precise detector be directly placed on this beam pipe. A second detector is at a larger radius. Beam pipe and first detector together shall have a thickness x and a joint radiation length X_0, from which we can compute an average effective scattering angle θ_{ms} according to (3.105). Assuming that the spatial resolution of the detectors is much smaller than the uncertainty caused by multiple scattering the error on d_0 becomes:

$$\Delta d_0 = \theta_{ms}\, r_B\,. \tag{3.107}$$



Kolanoski, Wermes 2015

Fig. 3.34 Diagrams for (a) photoelectric effect, (b) Compton effect and (c) pair production.

This means that the distance of the first detector layer to the vertices should be as small as possible and the detector should be as thin as possible in comparison to the radiation length of its material (see also section 9.4.6.2).

In table 3.6 the error of the impact parameter is given for a thickness of the scattering material of $x = 1\,\text{mm}$ and a particle momentum of $p = 5\,\text{GeV/c}$ and $\beta \approx 1$ for the materials aluminium and beryllium. Since the radiation length of Be is about four times larger than for Al the error of the impact parameter is hence a factor of 2 smaller.

In section 9.4 of chapter 9 the conditions for a good vertex resolution are discussed more generally.

3.5 Interactions of photons with matter

For particle detectors the following interactions of photons with matter are of particular importance (fig. 3.34):

- Photoelectric effect: the photon transfers its total energy to an atom which thereupon emits a shell electron.
- Compton effect: the photon is scattered elastically off a shell electron.
- Pair production: the photon converts in the field of a nucleus into an electron–positron pair.

Figure 3.35 shows these processes in a picture of a bubble-chamber event and fig. 3.36 on page 72 shows the various photon cross sections as a function of the photon energy. The three processes dominate at photon energies above the ionisation threshold. At low energies Thomson and Rayleigh scattering also play a role. Thomson scattering is low energy photon scattering off free charges. For electrons it constitutes the low energy Compton limit (see eq. (3.128)). The coherent scattering off a whole atom without shell excitation or ionisation is called Rayleigh scattering.

We discuss all processes entering in fig. 3.36 in sections 3.5.2 to 3.5.5 below. For Rayleigh and Thomson scattering almost no energy transfer to the atom occurs. Hence they are not of importance for photon detection.

3.5.1 Photon absorption

As a result of the above interaction processes, photons are absorbed or—in the Compton case—scattered off from their original direction, with a probability proportional to the path length dx in the medium. In section 3.1 the absorption coefficient μ, specifying the absorption probability per path length, has been defined:

$$-\frac{1}{N}\frac{dN}{dx} = \mu = \rho\,\frac{N_A}{A}\,\sigma = n\,\sigma \qquad (n = \text{target density})\,. \tag{3.108}$$

The reciprocal of μ is the mean free path or *absorption length*:

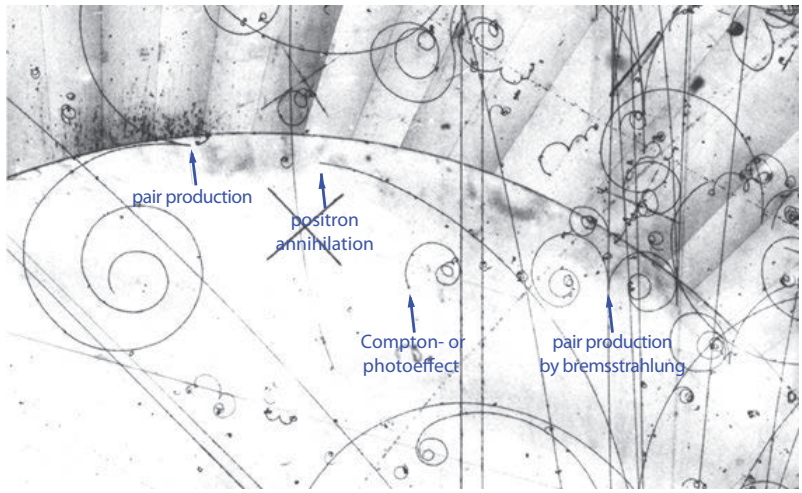


Fig. 3.35 Details of a bubble chamber picture of an electromagnetic shower (see section 15.2) developing from bottom to top (source: CERN [284]). The picture was taken by the '15-foot Bubble Chamber' at Fermilab filled with a mixture of He and Ne. The curvatures of the tracks are due to a magnetic field pointing in this exposure into the plane of the picture (negative tracks are curved clockwise). The crosses are reference points for the measurement of the track parameters. The picture shows examples for the processes sketched by the diagrams in fig. 3.34. Pairs of tracks with opposite curvatures and small opening angle at the vertex are electron–positron pairs, here originating mostly from bremsstrahlung photons. In the picture also the case is indicated in which the photon obviously originates from the annihilation of a positron with a shell electron (inverse pair production). One of the annihilation photons again produces a pair close to the annihilation point at a position lying in the direction of the original positron. Pairs that have a vertex very near a straight line and having a tangential emission topology, likely come from bremsstrahlung of the particle belonging to the straight track. The individual spirals, curved clockwise, are electrons, created by photo or Compton effect.

$$\lambda = \frac{1}{\mu} = \frac{1}{n\,\sigma}\,. \tag{3.109}$$

Again, one usually finds tabulated values normalised to unit density:

$$\frac{\mu}{\rho} = \frac{N_A}{A}\sigma \qquad \text{respectively} \qquad \rho\,\lambda = \left(\frac{\mu}{\rho}\right)^{-1} \tag{3.110}$$

in units cm^2/g or g/cm^2, respectively. Figure 3.37 shows the absorption lengths of various elements as a function of the photon energy. The number of photons in a beam follows an exponential law following from (3.108):

$$N(x) = N_0\,\mathrm{e}^{-\mu x}\,. \tag{3.111}$$

This must be compared with the ionisation interactions affecting charged particles which show a completely different depth dependence: the particles lose energy continuously resulting in an energy dependent but fixed absorption range (see fig. 3.2 on page 25).

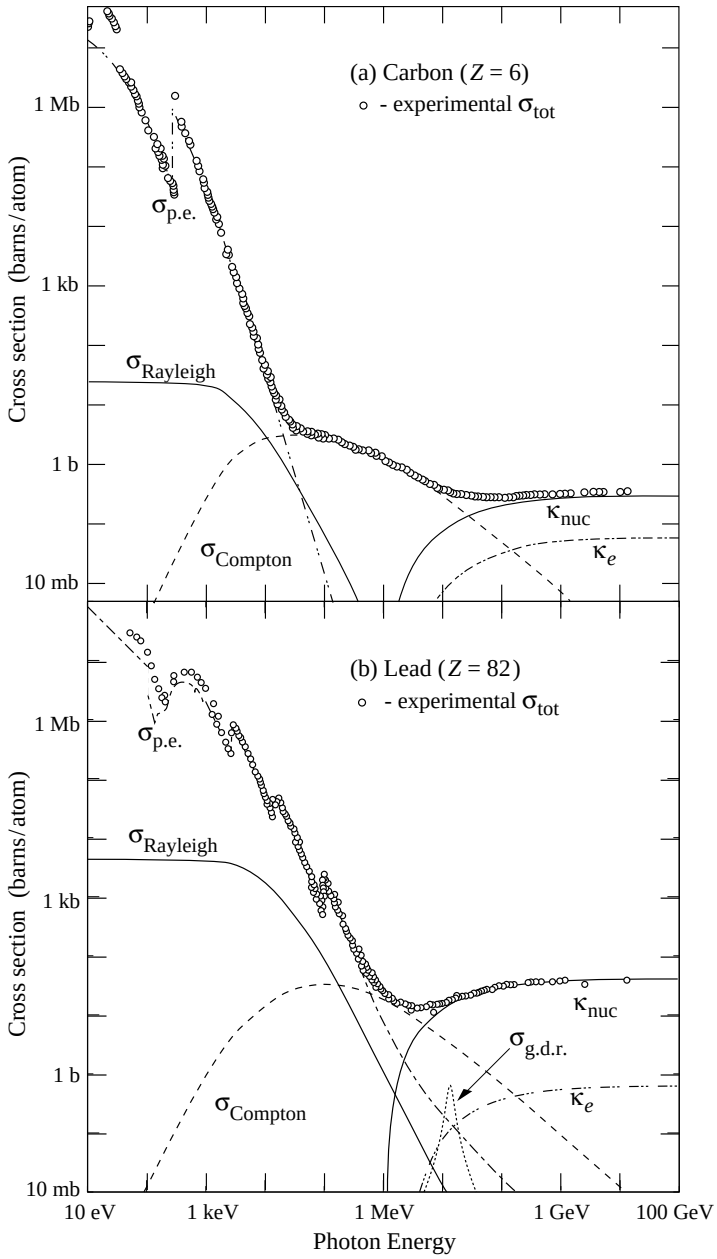


Fig. 3.36 Different contributions to the photon total cross section as a function of energy in carbon and lead (source: PDG [932]).

$\sigma_{p.e.}$ = atomic photoelectric effect (electron ejection, photon absorption)
$\sigma_{Rayleigh}$ = Rayleigh (coherent) scattering; atom neither ionised nor excited
$\sigma_{Compton}$ = incoherent scattering (Compton scattering off an electron)
κ_{nuc} = pair production, nuclear field
κ_e = pair production, electron field
$\sigma_{g.d.r.}$ = photonuclear interactions, most notably the giant dipole resonance.

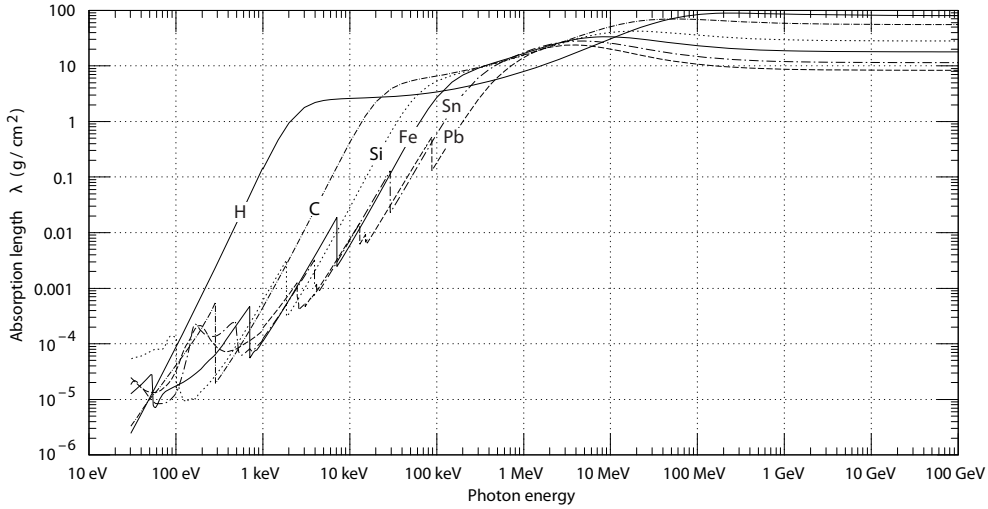


Fig. 3.37 Photon absorption length (or mean free path) $\lambda = 1/(\mu/\rho)$, with $\mu =$ mass attenuation coefficient and $\rho =$ density, for various elements as a function of the photon energy [932].

Figure 3.36 shows the different processes contributing to the interaction cross section of photons for carbon and for lead [746]. The characteristic dependences of the processes on the photon energy and on the nuclear charge Z of the traversed medium are discussed in the following.

3.5.2 Thomson and Rayleigh scattering

The scattering of an electromagnetic wave off a free charge is called Thomson scattering. The scattering occurs because the electric field excites oscillations of the charge in field direction, that is, transverse to the wave direction. The energy is subsequently reemitted in form of dipole radiation. In the photon picture of the electromagnetic wave this leads for unpolarised photons at low energies, where the recoil energy of the charge can be neglected, to an angular distribution of the scattered photon proportional to $1 + \cos^2\theta_\gamma$.

For electrons the differential Thomson cross section is

$$\frac{d\sigma^e_{\mathrm{Th}}}{d\Omega} = \frac{r_e^2}{2}\left(1 + \cos^2\theta_\gamma\right) , \tag{3.112}$$

where θ_γ is the polar angle of the scattered photon with respect to the incoming photon direction. Integration over the solid angle yields the total Thomson cross section,

$$\sigma^e_{\mathrm{Th}} = \frac{8\pi r_e^2}{3} = 0.665\,\mathrm{b}\,, \tag{3.113}$$

with the classical electron radius $r_e \approx 2.8\,\mathrm{fm}$ as defined in (3.26). The Thomson cross section for electrons is often used as a reference for other photon cross sections.

Coherent, elastic photon scattering off a whole atom is called Rayleigh scattering. In the range of visible light, Rayleigh scattering is known to cause the blue sky. Here we use the term more generally for coherent photon scattering up to the X-ray regime as shown in fig. 3.36.

For atoms the above described oscillations become oscillations of shell electrons, determined by the polarisability of the atom. The polarisability depends on the resonances with eigenfrequencies ω_{i0}, which lie for atoms and molecules in the ultraviolet and above. Sufficiently away from the eigenfrequencies one can neglect the widths of the resonances and the scattering cross section can be written as

$$\sigma_R = \sigma^e_{\mathrm{Th}}\,\omega^4 \left(\sum_i \frac{f_i}{\omega^2 - \omega_{i0}^2} \right)^2 . \tag{3.114}$$

The sum goes over all Z dipole oscillators with oscillator strengths f_i, which are normalised as $\sum_i f_i = Z$ (see e.g. [564]).[11] The ω^4 dependence is characteristic of dipole radiation.

For $\omega \ll \omega_0$, which includes the visible light region, this becomes

$$\sigma_R = \sigma^e_{\mathrm{Th}}\,\omega^4 \left(\sum_i \frac{f_i}{\omega_{i0}^2} \right)^2 , \tag{3.115}$$

which is the well-known dependence explaining the blue colour of the sky by the enhanced scattering at higher frequencies.

Above the resonances with $\omega \gg \omega_0$, but the photon wavelength still much larger than the atomic radius, the oscillations become independent of the individual eigenfrequencies and the whole electron shell oscillates with the frequency ω of the driving field. Then the Rayleigh cross section becomes

$$\sigma_R = \sigma^e_{\mathrm{Th}} \left(\sum_i f_i \right)^2 = Z^2\,\sigma^e_{\mathrm{Th}} = \sigma^{\mathrm{atom}}_{\mathrm{Th}} . \tag{3.116}$$

The formula means that the photon sees the Z electrons coherently; it describes the nearly energy independent low energy part of the curves labelled $\sigma_{\mathrm{Rayleigh}}$ in fig. 3.36.

As the wavelength of the photon decreases with increasing energy, the photons begin to resolve the electronic structure of the atom leading to damping of the coherent scattering. The damping can be described by applying a formfactor to the Thomson cross section.

3.5.3 Photoelectric effect

In this section we treat the *photoelectric effect* or just *photoeffect*, both used synonymously in this book. It is the dominant photon interaction process at energies in the lower keV range, typically above 1 keV. Absorption of optical or near optical photons is discussed in chapter 10 together with the corresponding photodetectors.

In the photoelectric effect the photon transfers its complete energy onto an atom which then releases the absorbed energy by emitting an electron into the continuum (fig. 3.34(a)):

$$\gamma + \mathrm{atom} \longrightarrow (\mathrm{atom})^+ + e^-$$

The atom absorbs the recoil momentum and is ionised after the reaction.

[11] The cross section (3.114) is often expressed by the macroscopic parameters of refraction or electric permittivity (as also in [564]). The connection with the microscopic picture of atomic dipoles is given by (11.10).

In order that the photoeffect takes place, the energy E_γ of the photon must exceed the binding energy E_B of the electron ($E_\gamma > E_B$); the energy surplus is transferred as kinetic energy to the electron

$$T = E_\gamma - E_B \,.$$

The recoil energy transferred to the atom can be neglected due to the atom's large mass.

The cross section for photoeffect rapidly drops with increasing photon energy and (at fixed photon energy) strongly increases with increasing Z. For a given photon energy the probability for absorption is largest for the most strongly bound inner shell electrons, for which the energy surpasses the threshold for that shell. This is why jumps in the absorption cross section, so-called absorption edges, are observed when the binding energy of an electron in a particular shell is reached with increasing photon energy: $E_\gamma = E_B^i$, $i = \mathrm{K}, \mathrm{L}, \ldots$. The cross section is highest for shell i, for which $E_\gamma - E_B^i$ is smallest. For carbon (fig. 3.36(a)) one can only see the edge for the K-shell whereas for lead (fig. 3.36(b)) the edges of the K-, L- and M-shells are visible. At still lower energies ($\lesssim 5$ eV) the cross section falls again, a characteristic described in more detail in chapter 10.

Cross section. The computation of the photoelectric effect for all energies is very complex and challenging, requiring approximations, as it involves the entire atom. For all practical applications it is therefore recommended to use measured data which are provided for example by the US National Institute of Standards (NIST) [545] and also as an interactive database [181].

Computations of the cross section for photon absorption in the K-shell only are found in various textbooks, e.g. in [180, 517]. Since these computations reproduce the data fairly well and since K-shell absorption is dominant for energies above the K-edge, the results for this simplified case are presented and discussed in the following.

For photon energies larger than the K-shell ionisation energy, but small compared to the electron mass ($E_\gamma \ll m_e c^2$), the non-relativistic approximation is justified. In this approximation the total absorption cross section in the K-shell per atom is:

$$\sigma_K^{NR} = \sqrt{32}\,\alpha^4\,\epsilon^{-3.5}\,Z^5\,\sigma_{\mathrm{Th}}^e\,f(\xi) \tag{3.117}$$

with $\epsilon = E_\gamma/m_e c^2$, $\alpha = 1/137$ the fine structure constant and $\sigma_{\mathrm{Th}}^e = \frac{8}{3}\pi r_e^2$ the Thomson cross section (3.113), respectively. The expression without the factor $f(\xi)$ corresponds to the Born approximation which assumes a plane wave for the outgoing electron. Practically, however, this approximation does not yield good results for the cross section in any energy range [517]. A reasonable description of the data is only obtained when the factor $f(\xi)$ is included taking care of the fact that the electron must absorb the angular momentum of the photon (see e.g. [180]). The factor $f(\xi)$ depends substantially on the ratio of the binding energy in the K-shell E_K and the kinetic energy transferred to the electron $T = E_\gamma - E_K$:

$$f(\xi) = 2\pi Z\alpha\sqrt{\frac{1}{2\epsilon}}\,\frac{\exp(-4\xi\,\mathrm{arccot}\,\xi)}{1-\exp(-2\pi\xi)}\,, \tag{3.118}$$

$$\text{with} \qquad \xi = \sqrt{\frac{\epsilon_K}{\epsilon-\epsilon_K}}\,, \qquad \epsilon_K = \frac{E_K}{m_e c^2} = \frac{1}{2}Z^2\alpha^2\,.$$

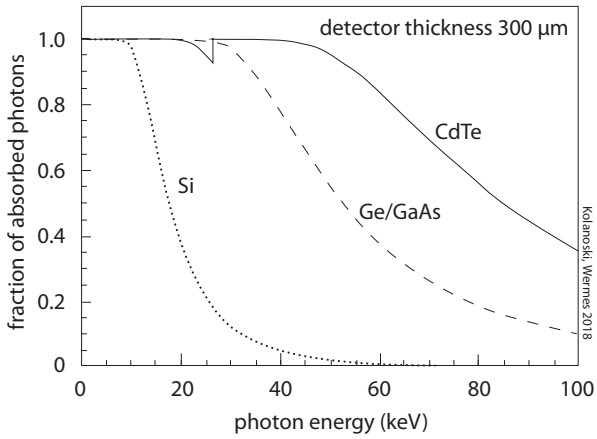


Fig. 3.38 Fraction of photons absorbed by photoeffect in some semiconductor materials, Si (dotted), Ge or GaAs (dashed), and CdTe (solid), for a sensor thickness of 300 µm. Data from NIST [181].

The strong dependence of the cross section on the photon energy and the nuclear charge Z characteristically appears in the Born cross section of (3.117) proportional to $Z^5/E_\gamma^{3.5}$. The correction factor $f(\xi)$, however, mitigates these dependencies such that the Z dependence changes to Z^n with $n = 4-5$ and the E_γ dependence to E_γ^{-m} with $m \lesssim 3.5$.

The non-relativistic approximation (3.117) holds up to values of $\epsilon = 0.5$ according to [517]. At higher energies Dirac wavefunctions have to be used to compute the matrix element of the process. In [517] a relativistic derivation by Sauter is given:

$$\sigma_K = \frac{3}{2}\,\sigma_{\mathrm{Th}}^e\,\alpha^4\,\frac{Z^5}{\epsilon^5}(\gamma^2-1)^{3/2}\left[\frac{4}{3}+\frac{\gamma(\gamma-2)}{\gamma+1}\left(1-\frac{1}{2\gamma\sqrt{\gamma^2-1}}\ln\frac{\gamma+\sqrt{\gamma^2-1}}{\gamma-\sqrt{\gamma^2-1}}\right)\right]. \tag{3.119}$$

Here $\gamma = (T+m_e)/m_e$ is the Lorentz factor of the outgoing electron which for $E_\gamma \gg E_B$ can be written as $\gamma = (E_\gamma + m_e)/m_e$. The non-relativistic Born approximation (3.117) without the correction factor $f(\xi)$ is recovered from (3.119) for small photon energies, that is, $\gamma \to 1$ (but with E_γ still large compared to the threshold E_K). At large photon energies ($E_\gamma \gg m_e$) $\gamma \to \epsilon$ holds, transforming (3.119) in the highly relativistic limit ($\gamma \to \infty$) to the form

$$\sigma_{ph}^{HE} = \frac{3}{2}\,\sigma_{\mathrm{Th}}^e\,\alpha^4\,\frac{Z^5}{\epsilon}\,. \tag{3.120}$$

Thus, at high photon energies the cross section decreases only as $1/E_\gamma$, that is, much more slowly than at low energies near the K-edge. However, in the region where approximation (3.120) is applicable, photoeffect is negligible compared to Compton effect and pair production. Figure 3.38 illustrates this for four semiconductor materials, showing the fraction of X-ray photons absorbed by photoeffect for the same thickness (300 µm).

Angular distribution. The angular distribution of the outgoing electron is given in lowest order as [451, 452]:

$$\frac{d\sigma}{d\cos\theta} \propto \frac{\sin^2\theta}{(1-\beta\cos\theta)^4}\left\{1+\frac{1}{2}\gamma(\gamma-1)(\gamma-2)(1-\beta\cos\theta)\right\} \tag{3.121}$$

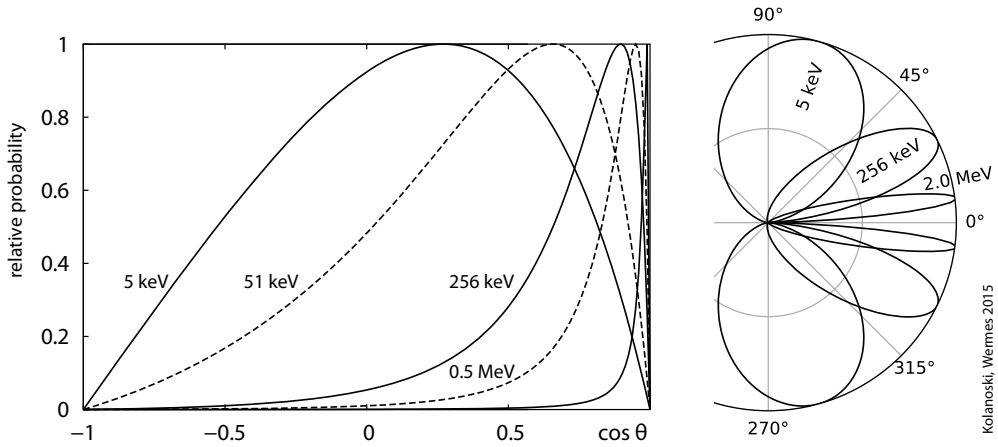


Fig. 3.39 Photoelectric effect: angular distribution (probability per solid angle) of the outgoing electron relative to the direction of the incoming photon (computed with eq. (3.121)) for different kinetic energies of the electron corresponding to Lorentz factors $\gamma = 1.01,\ 1.1,\ 1.5,\ 2.0,\ 5.0$. Left: relative probability plotted versus the cosine of the electron angle; right: the same relative probability plotted as a polar diagram for three electron energies ($\gamma = 1.01,\ 1.5,\ 5.0$).

with θ being the polar angle of the emitted electron relative to the direction of the incoming photon. In higher order additional terms proportional to powers of αZ appear. The distribution is shown in fig. 3.39 for different electron energies. In the non-relativistic limit $\gamma \to 1$ the distribution becomes that of dipole radiation ($d\sigma/d\cos\theta \propto \sin^2\theta$) with preferred emission orthogonal to the direction of the incoming photon. With increasing energy, $\beta \to 1$, the electron is emitted more and more in the direction of the incoming photon.

Energy deposit in a detector. The vacancy created by the photoelectric effect in an inner shell can be filled again by electrons from outer, preferentially nearby shells. The energy released in this process can be re-emitted as a photon or as an electron, so-called *Auger electrons*. Since both processes involve atomic transitions between discrete energy levels, the corresponding photon or electron energies are also discrete, giving rise to characteristic peaks in X-ray and Auger spectra. In X-ray spectra the photon peaks are called *characteristic lines* since their energies are characteristic for an element. The probability for energy release by photon emission increases with Z, whereas Auger emission becomes less probable with increasing Z. The spectroscopy of such X-ray photons (XRF = 'X-ray fluorescence spectroscopy') or of Auger electrons (XPS = 'X-ray photoelectron spectroscopy') is used as a very sensitive method for the analysis of surfaces where either monochromatic X-rays or electrons bombard a surface to be examined. An example for an electron spectrum is given in fig. 3.40. Characteristic lines observed in the spectra of X-ray tubes are shown in fig. 3.25 on page 58.

If the total energy of the incoming photon is detected, a so-called *photopeak* is observed in the measured energy spectrum. If, on the other hand, a secondary photon with discrete energy leaves the detector, a second peak (escape peak) can develop. For tests of gas-filled or semiconductor detectors the radioactive isotope ^{55}Fe is often used

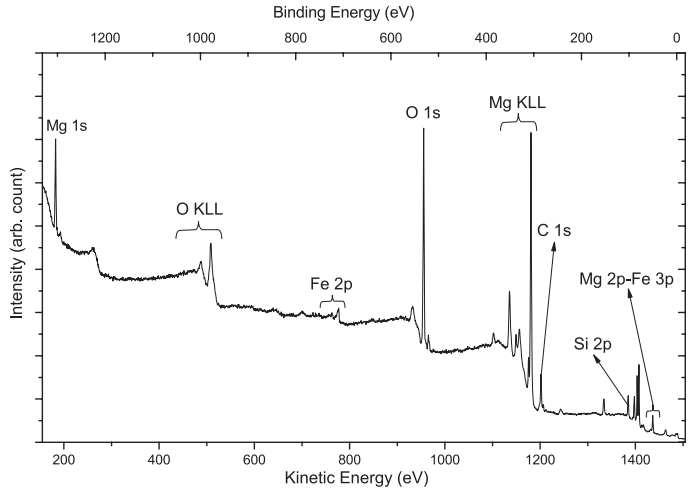


Fig. 3.40 Electron spectrum of an olivine mineral surface bombarded by a Al-K_α X-ray source ($h\nu = 1486.6$ eV) (from [268]). The photopeaks are labelled by the corresponding element and the bound state, from which the electron was emitted. In addition two Auger lines from oxygen and magnesium (O KLL, Mg KLL) can be seen.

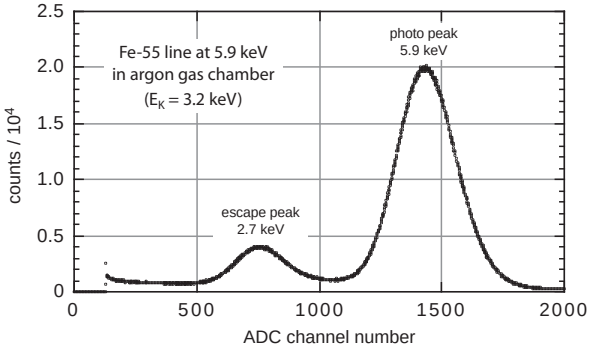


Fig. 3.41 Photo and escape peaks of the 5.9-keV γ line of a ^{55}Fe isotope measured in argon [191].

which has a γ line at 5.9 keV (see table A.1 in the appendix). If ^{55}Fe photons are for example absorbed in the argon gas filling of a proportional chamber (section 7.6.1) one can—in addition to the photopeak at 5.9 keV—observe an escape peak at about 2.9 keV (fig. 3.41). The energy that has left the detector is predominantly the energy difference between the L- and the K-shell, carried outside the detector by photon or electron emission. Auger electrons are with some probability completely absorbed in the detector and their deposited energy hence also contributes to the photopeak.

3.5.4 Compton effect

The Compton effect is the scattering of a photon off a free or quasi-free electron (fig. 3.34(b)). A shell electron is termed 'quasi-free' in this context when the energy of the incoming photon is substantially larger than the electron's binding energy ($E_\gamma \gg E_B$). We call a scattering process Compton scattering if the electron is kicked out of the atom, carrying away the recoil momentum. There is also the possibility that, even

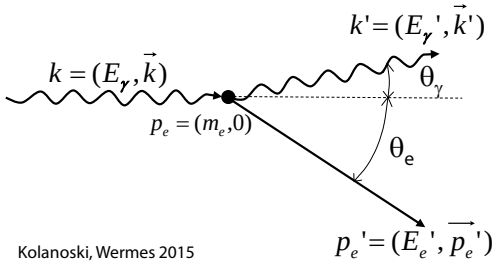


Fig. 3.42 Kinematics of the Compton process. The electron is considered as 'quasi free'.

at high photon energies, the electron remains in the atom shell. In this case the whole atom takes the recoil ('coherent photon scattering') and the scattered photon's energy is nearly the same as the initial energy, independent of the scattering angle. Coherent photon scattering becomes strong at lower energies. At photon energies of visible light, where the corresponding wavelengths are much larger than the atom dimensions, this is known as Rayleigh scattering (see fig. 3.36). The Compton effect is dominant in an extended energy region around 1 MeV. The extent of this region is much larger at low Z of the medium than at high Z (see fig. 3.49).

Kinematics. Since the photon is scattered quasi-elastically off the electron, energy and angle of the scattered photon are not independent. To quantify this relationship we use the 4-momenta defined in fig. 3.42: $k = (E_\gamma, \vec{k}c)$ and $p_e = (m_e c^2, 0)$ are the 4-momenta of photon and electron (at rest) before the scattering, and $k' = (E'_\gamma, \vec{k}'c)$ and $p'_e = (E'_e, \vec{p}'_e c)$ are the 4-momenta after the scattering. The angle of the scattered photon with respect to the direction of the incoming photon we call θ_γ, that of the electron θ_e. From energy-momentum conservation

$$k + p_e = k' + p'_e\,, \tag{3.122}$$

we find:

$$\begin{aligned} (k - k')^2 = (p'_e - p_e)^2 \quad &\Rightarrow \quad -k\,k' = m_e^2 c^4 - p'_e\,p_e \\ \Rightarrow \qquad E_\gamma\,E'_\gamma(1 - \cos\theta_\gamma) &= m_e c^2\,(E'_e - m_e c^2) = m_e c^2\,(E_\gamma - E'_\gamma)\,. \end{aligned} \tag{3.123}$$

The right-hand side of the last equation uses the kinetic energy of the electron

$$T = E'_e - m_e c^2 = E_\gamma - E'_\gamma\,, \tag{3.124}$$

which follows from the energy part of (3.122). The energy of the scattered photon as a function of the scattering angle results from (3.123):

$$E'_\gamma = \frac{E_\gamma}{1 + \epsilon\,(1 - \cos\theta_\gamma)} \tag{3.125}$$

with $\epsilon = E_\gamma / m_e c^2$.

Cross section. The differential cross section per (free) electron, known as the Klein–Nishina formula [609], is calculated using methods of quantum electrodynamics:

$$\frac{d\sigma_C}{d\Omega_\gamma} = \frac{r_e^2}{2\,[1 + \epsilon(1 - \cos\theta_\gamma)]^2} \left(1 + \cos^2\theta_\gamma + \frac{\epsilon^2(1 - \cos\theta_\gamma)^2}{1 + \epsilon(1 - \cos\theta_\gamma)}\right)\,. \tag{3.126}$$

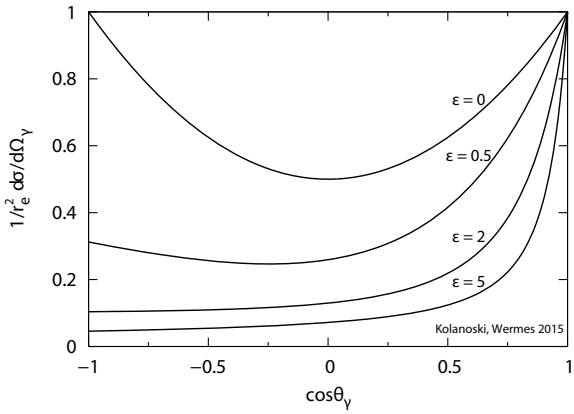


Fig. 3.43 Angular distribution of the scattered photon for Compton scattering according to the Klein–Nishina formula.

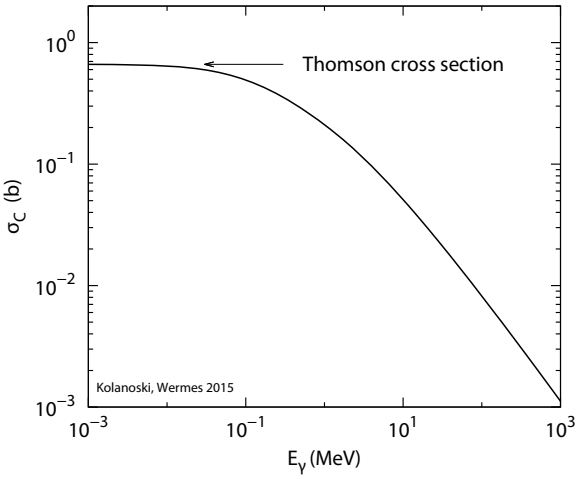


Fig. 3.44 Compton cross section per electron as a function of the photon energy. According to equation (3.128) σ_C approaches the Thomson cross section (3.113) for small energies. For very large energies σ_C develops a dependence roughly like $1/E_\gamma$.

For small photon energies, $\epsilon \to 0$, the differential cross section for classical Thomson scattering results, as given in (3.112) in section 3.5.2:

$$\frac{d\sigma^e_{\mathrm{Th}}}{d\Omega} = \frac{r_e^2}{2}\left(1+\cos^2\theta_\gamma\right) .$$

The angular distribution for Compton scattering is shown in fig. 3.43. There is a strong rise in the forward direction which is steeper for larger photon energies. Integration of the Klein–Nishina formula (3.126) over the solid angle yields the total Compton cross section per electron (fig. 3.44):

$$\sigma_C = 2\pi\, r_e^2 \left[\frac{1+\epsilon}{\epsilon^2}\left(\frac{2(1+\epsilon)}{1+2\epsilon} - \frac{1}{\epsilon}\ln(1+2\epsilon)\right) + \frac{1}{2\epsilon}\ln(1+2\epsilon) - \frac{1+3\epsilon}{(1+2\epsilon)^2}\right] . \quad (3.127)$$

For very small, respectively very large photon energies the following approximations hold:

$$\sigma_C \approx \begin{cases} \dfrac{8\pi\, r_e^2}{3}(1-2\epsilon) & \text{for } \epsilon \ll 1 \qquad \text{(Thomson limit)}\,, \\ \dfrac{\pi\, r_e^2}{\epsilon}\left(\ln(2\epsilon)+\dfrac{1}{2}\right) & \text{for } \epsilon \gg 1 \qquad \text{(highly relativistic)}\,. \end{cases} \tag{3.128}$$

Compton scattering off electrons in atomic shells. The Compton cross section (3.127) has been derived for free electrons. An electron bound in an atom can only be considered quasi-free if the photon energy is well above the electron's binding energy. As the photon energy increases, more and more shell electrons become free in this sense and hence the Compton cross section per atom approaches proportionality to Z with the individual electrons contributing incoherently:

$$\sigma_C^{\text{atom}} = Z\,\sigma_C\,. \tag{3.129}$$

Here σ_C is the Klein–Nishina cross section for a single free electron. Figure 3.36 shows that the Compton cross section drops towards lower energies where coherent scattering (Rayleigh scattering) off the whole atom (leaving the electron shell intact) becomes dominant (see section 3.5.2).

Recoil energy. For detectors it is important that the kinetic energy of the recoil electron $T = E_\gamma - E'_\gamma$ (3.124) can be detected. Reformulating the Klein–Nishina formula (3.126) one obtains the differential dependence of the Compton cross section on the kinetic energy T of the recoil electron:

$$\frac{d\sigma}{dT} = \frac{\pi\, r_e^2}{m_e c^2 \epsilon^2}\left[2 + \frac{t^2}{\epsilon^2(1-t)^2} + \frac{t}{1-t}\left(t - \frac{2}{\epsilon}\right)\right]\,, \tag{3.130}$$

where $t = T/E_\gamma$ (fig. 3.45). Because the scattering process is elastic, as for the photon also for the electron a one-to-one relation between energy and angle θ_e results:

$$\cos\theta_e = \frac{T(E_\gamma + m_e c^2)}{E_\gamma\sqrt{T^2 + 2m_e c^2 T}} = \frac{1+\epsilon}{\sqrt{\epsilon^2 + 2\epsilon/t}}\,. \tag{3.131}$$

The maximum energy transfer to the electron is obtained from (3.125) for backward scattering of the photon ($\theta_\gamma = 180°$) corresponding to forward scattering of the electron ($\theta_e = 0°$). The electron kinetic energy becomes maximal in this case: $T \to T_{max}$. In the measured energy spectrum this leads to the so-called 'Compton edge' at

$$T_{max} = E_\gamma \frac{2\epsilon}{1+2\epsilon}\,, \tag{3.132}$$

lying somewhat below the photopeak (fig. 3.45). The energy difference between photopeak and Compton edge $E'_\gamma(\theta = \pi)$ decreases with increasing E_γ and reaches

$$E'_\gamma(\theta = \pi) \approx \frac{m_e c^2}{2} \quad \text{for} \quad E_\gamma \gg m_e c^2\,. \tag{3.133}$$

Inverse Compton effect. The scattering of low energy photons off high energy electrons is called 'inverse Compton effect' (IC). Inverse Compton effect occurs when

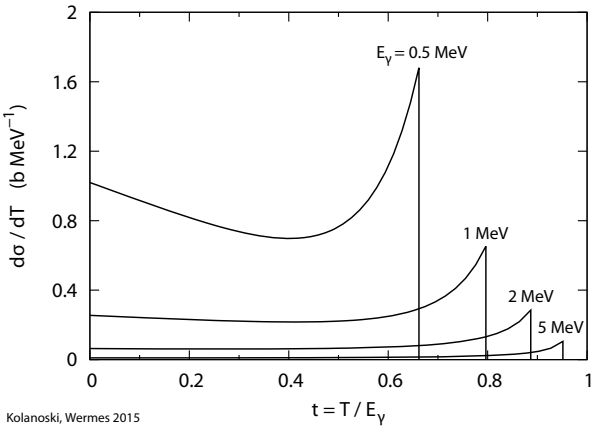


Fig. 3.45 Dependence of the Compton cross section on the normalised kinetic energy of the electron $t = T/E_\gamma$ for different energies of the primary photon. The cut-off at high t is called 'Compton edge' and corresponds to the energy which can be maximally transferred to the electron. The Compton edge always lies below the photopeak.

electrons are accelerated and then transfer their energy to photons. This is a way to produce high energy photon beams via Compton scattering off laser photons from a high energy electron beam. In astrophysics this IC effect plays an important role in the generation of high energy (TeV) gamma rays.

For the energy of the scattered photon one obtains

$$E'_\gamma = E_\gamma \frac{1+\beta\cos\alpha}{1-\beta\cos\theta_\gamma^{IC}+\frac{E_\gamma}{E_e}\left(1+\cos\left(\theta_\gamma^{IC}+\alpha\right)\right)} \tag{3.134}$$

with energies E_γ, E'_γ and E_e of the in- and outgoing photon and incoming electron, respectively, the angle θ_γ^{IC} of the scattered photon against the direction of the incoming electron[12], and the angle α defining the deviation of the incoming photon direction from that of a head-on collision.

For the collinear case, that is, $\alpha = 0$, the maximum scattered photon energy results for $\theta_\gamma^{IC} = 0$ as (γ is the Lorentz factor of the electron):

$$E'_{\gamma,max} = E_\gamma \frac{1+\beta}{1-\beta+2\,\frac{E_\gamma}{E_e}} = \gamma^2 E_\gamma \frac{(1+\beta)^2}{1+2\,\frac{E_\gamma}{m_e c^2}\gamma(1+\beta)}\,. \tag{3.135}$$

For small incoming photon energies we have approximately:

$$E'_{\gamma,max} \approx 4\gamma^2 E_\gamma \quad \text{for} \quad E_\gamma \ll \frac{m_e\,c^2}{4\gamma}\,. \tag{3.136}$$

The quadratic increase of the maximum photon energy with the energy of the electron shows that the energy transfer of IC scattering is very effective. The spectrum of IC photons has a strong enhancement at the maximum energy of the scattered photons; the average energy is about $1/3\,E'_{\gamma,max}$.

3.5.5 Pair production

In the Coulomb field of a charge a photon can convert into an electron–positron pair according to the diagram of fig. 3.34(c). The energy of the photon must be larger than

[12]The angle θ_γ^{IC} differs from the previously defined scattering angle θ_γ: $\theta_\gamma^{IC} = \pi - \theta_\gamma$.

(a) Bremsstrahlung. (b) Pair production.

Fig. 3.46 Diagrams of the two Bethe–Heitler processes: (a) bremsstrahlung and (b) pair production.

twice the electron mass plus the recoil energy transmitted to the field producing charge. Except for very light elements, pair production occurs predominantly in the Coulomb field of the nucleus.[13] For nuclei the recoil energy is practically always negligible such that the threshold energy for pair production is

$$E_\gamma \approx 2m_e c^2 \qquad \text{(threshold)}\,. \tag{3.137}$$

Photon emission by an electron (bremsstrahlung) and pair production are closely related, as is evident from fig. 3.46. If one modifies in the graph for bremsstrahlung (fig. 3.46(a)) the outgoing to an incoming photon and the incoming electron into an outgoing positron one obtains the graph for pair production (fig. 3.46(b)). The corresponding matrix elements are related, at least in lowest order. Therefore both processes are treated together in the fundamental work of Bethe and Heitler [189], hence also called 'Bethe–Heitler processes'. Reviews can be found for example in [342, 517, 719, 948].

The cross section for pair production shows a similar dependence on the screening effect caused by the atomic shell, as we have discussed for bremsstrahlung in section 3.3.2. The screening parameter corresponding to the one in (3.73) is here [517]

$$\eta = \frac{137\, m_e c^2 E_\gamma}{2E_+\, E_-\, Z^{1/3}}\,, \tag{3.138}$$

where E_+, E_- are the energies of the positron and the electron, respectively. Full screening is obtained for small values of η.

The differential cross section for the creation of an electron–positron pair, where the electron receives the fraction $x = E_-/E_\gamma$ of the photon's energy, is in the Born approximation given by [948]:

$$\begin{aligned}\frac{d\sigma}{dx} = \alpha r_e^2 \Bigg\{ \left(\frac{4}{3}x(x-1)+1\right) \left[Z^2(\phi_1(\eta) - \frac{4}{3}\ln Z - 4f(Z)\,) + Z(\psi_1(\eta) - \frac{8}{3}\ln Z)\right] \\ + \frac{2}{3}x(1-x)\left[(Z^2(\phi_1(\eta)-\phi_2(\eta)) + Z(\psi_1(\eta)-\psi_2(\eta))\right] \Bigg\}\,.\end{aligned} \tag{3.139}$$

In this approximation the cross section is symmetric with respect to electrons and positrons, that is, for the transition $x \to 1-x$. The Coulomb correction $f(Z)$ and the screening functions $\phi_{1,2}$ and $\psi_{1,2}$ are defined as for bremsstrahlung in section 3.3.2.1. With the approximate formulae (3.77)–(3.79) for these functions, curves of the differential cross section for various photon energies have been computed in fig. 3.47.

[13] Pair production can also occur with an electron being the scattering partner.

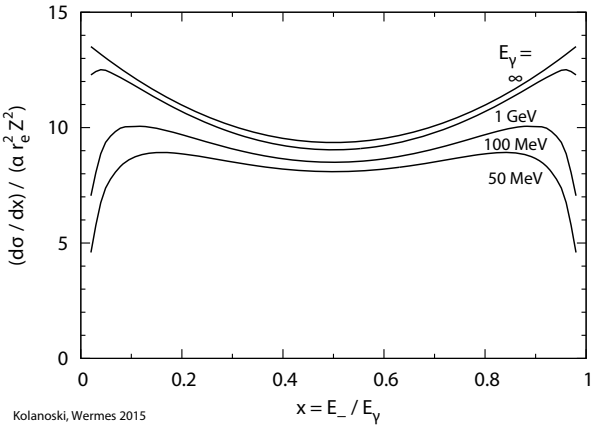


Fig. 3.47 Differential cross sections for pair production in units of $\alpha\, r_e^2\, Z^2/E_\gamma$ as a function of the normalised electron energy $x = E_-/E_\gamma$. The curves for $E_\gamma = 50\,\text{MeV}$, 100 MeV and 1 GeV are computed with variable screening (corresponding to (3.139)), the asymptotic curve is computed assuming full screening.

For the case of full screening the cross section can be written similar to the one for bremsstrahlung [948]:

$$\frac{d\sigma}{dx} = 4\alpha r_e^2 \left\{ \left(\frac{4}{3}x(x-1) + 1 \right) [\, Z^2(L_{rad} - f(Z)\,) + ZL'_{rad}\,] + \frac{1}{9}x(1-x)(Z^2+Z) \right\}. \tag{3.140}$$

The functions L_{rad} and L'_{rad} have already been introduced in the context of bremsstrahlung in section 3.3.2.1. Integration of (3.140) yields the total pair production cross section in high energy approximation:

$$\sigma_{pair} = 4\alpha r_e^2 \frac{7}{9} \left\{ \left[Z^2(L_{rad} - f(Z)\,) + ZL'_{rad} \right] + \frac{1}{42}(Z^2+Z) \right\}. \tag{3.141}$$

The equation shows that the pair production cross section in this approximation is independent of the photon energy and is proportional to the mean energy loss by bremsstrahlung given in (3.85). The latter is the case if one neglects in both cases the last two terms, that is, here the term $\frac{1}{42}(Z^2+Z)$. Consequently, one can express the pair production cross section in terms of the radiation length defined in (3.88):

$$\sigma_{pair} \approx \frac{7}{9}\,\frac{1}{X_0}\,\frac{A}{N_A\,\rho}\,. \tag{3.142}$$

Hence the photon absorption length for pair production becomes

$$\lambda_\gamma = \frac{1}{\sigma_{pair}\, n_{atoms}} = \frac{9}{7} X_0\,, \tag{3.143}$$

where $n_{atoms} = \rho N_A/A$ specifies the density of scattering centres. For pair production the radiation length is the distance over which pair production takes place with a probability of

$$P(e^+e^-) = 1 - e^{-\frac{7}{9}} = 54\%\,.$$

In comparison, after one radiation length an electron has lost the fraction $1 - 1/e = 63.2\,\%$ of its energy by bremsstrahlung.

The field of the nucleus is formed by the coherent sum of Z nucleon charges leading to the Z^2 dependence of the cross section for pair production. Only for light nuclei

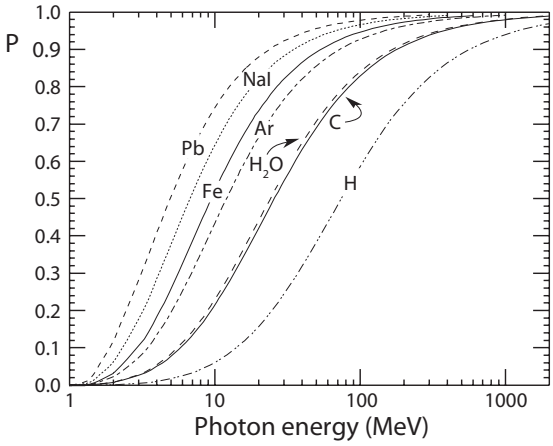


Fig. 3.48 Probability P for different media to create an electron—positron pair in a photon interaction (from [746]). In the energy range shown, apart from pair production only Compton scattering contributes significantly to photon interactions. Therefore the probability for Compton scattering is approximately $1 - P$.

is pair production off electrons of a comparable magnitude as pair production off nuclei (κ_e and κ_N in fig. 3.36). One should, however, be aware that for example for hydrogen both processes contribute about equally at high photon energies. Electron or nucleus, respectively, must absorb the recoil momentum in the process. Since a recoil electron receives a much larger energy transfer than a nucleus due to its much lighter mass, the threshold for pair production off electrons is moved up in energy and the near threshold rise is flatter (fig. 3.36). The threshold behaviour for different nuclei is shown in fig. 3.48.

Even at large momentum transfers Δp to the nucleus, the energy transfer $(\Delta p)^2/2M$ remains small due to the large masses of nuclei. The remaining energy after pair creation is transferred as kinetic energy to the e^+ and the e^-. The development of electromagnetic showers at high energies (section 15.2) is largely determined by the interplay between bremsstrahlung and pair production. Since both processes depend on the radiation length, the thickness of detector layers is often measured in multiples of their radiation length (see table 3.4 on page 62).

3.5.6 Dependence of photon processes on energy and atomic number

The following list provides a summary of approximate dependencies of the discussed photon interactions on the powers of E_γ and the atomic number Z:

Process	Equation	Z	E
photoeffect	(3.117)	$Z^n, \quad n \approx 4-5$	$E_\gamma^{-k}, \quad k \lesssim 3.5$
photoeffect, $E_\gamma \to \infty$	(3.120)	Z^5	$E_\gamma^{-k}, \quad k \to 1$
Compton effect	(3.127), (3.129)	Z	$1/E_\gamma$ for $E_\gamma \gg m_e c^2$
pair production	(3.140)	Z^2	$\sim$ const for E_γ large

For the photoeffect the first row is only valid in a restricted region closely above the K-edge.

Figure 3.49 shows regions in the (E_γ, Z)-plane in which a process is dominant in each case. For a given Z value the strong decline of the photoeffect with energy leads to the fact that the Compton effect becomes dominant from a certain energy onwards. It is then taken over by pair production starting at its threshold and becoming constant at high energies. The stronger dependence on Z of photoeffect and pair production

compared to Compton scattering is responsible for the fact that the region in which the Compton effect dominates becomes smaller with increasing Z. For silicon (Z = 14), for example, the Compton effect dominates roughly between 60 keV and 15 MeV, whereas for lead it is dominant between about 0.6 MeV and 4 MeV. For media with high Z, the Compton effect only dominates in a relatively narrow window around 1 MeV. In this region around 1 MeV photon absorption generally reaches a minimum. Electromagnetic radiation of this energy is therefore difficult to shield.

3.6 Interactions of hadrons with matter

In interactions of hadrons (p, n, π, K, ...) with matter, besides ionisation, hadronic interactions play an important role. A theoretical treatment of hadronic interaction is very challenging due to the multitude of possible reactions which are generally also less well computable than electromagnetic processes. Here we only introduce some terms and definitions used to describe the interactions of high energetic hadrons in detectors and defer a more detailed treatment to chapter 15 where hadron calorimeters are described.

In analogy to the radiation length X_0 for electromagnetic processes, for high energy hadrons a (hadronic) *absorption length* λ_a is defined. If N_0 hadrons initially enter an absorber, after a distance x a number

$$N(x) = N_0 \, \mathrm{e}^{-x/\lambda_a} \tag{3.144}$$

still remain. The absorption length can be deduced from the inelastic hadronic cross section being about constant for high energy hadrons (≈ 45 mb per nucleon, see eq. (3.9)):

$$\lambda_a = \frac{A}{N_A \, \rho \, \sigma_{inel}} \propto \mathcal{A}^{-\frac{2}{3}} \, . \tag{3.145}$$

The approximate proportionality to $\mathcal{A}^{-2/3}$ results from (3.9) and the assumption that ρ is roughly proportional to A (remember that we denote the atomic mass by A and the mass number by $\mathcal{A}$, see footnote 1 on page 24). Again, in tables one usually finds quoted values for $\rho \, \lambda_a$ instead. For larger Z the absorption length λ_a is generally much larger than the radiation length X_0 (see table 15.3 on page 599). This is why hadron calorimeters are always much larger than electromagnetic calorimeters (see chapter 15). Instead of the absorption length an *interaction length* λ_{int} is also defined

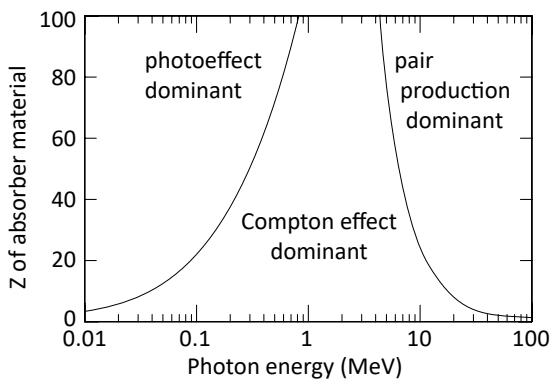


Fig. 3.49 Regions in the (E_γ, Z) plane in which the different photon absorption processes are dominant (adapted from [382], data from [181]).

in which the sum of the elastic and the inelastic cross sections enters. In many cases the relevant quantity, however, is the absorption length λ_a because elastically scattered particles usually do not transfer much energy to the medium and are rarely strongly deflected from their original direction.

Charged hadrons lose energy continuously by ionisation of the traversed medium, as described by the Bethe–Bloch formula (3.25). Hadronic interactions are the cause that not all particles travel up to the range defined by ionisation only (eq. (3.64)) but are rather absorbed on their way. With increasing energy, when the ionisation range becomes much larger than the absorption length, hadronic interactions start to dominate in an extended medium and hadronic showers develop (section 15.3).

3.7 Simulation of interactions in detectors

In order to extract numerical values for quantities like reaction probability, cross section, or lifetime from measurements of particles by a detector, the detection efficiency must be known. Mathematically the determination of the detection efficiency for a certain particle reaction is the solution of an integral over all possible final states of a reaction weighted with the probability to detect a given final state. Usually, such high dimensional integrals cannot be solved analytically, not least because the functions to be integrated do not exist in analytic form. A standard method for numerical integration of complex integrals is the so-called 'Monte Carlo method'. In this method discrete final states (normally many-particle systems) are generated at random and the individual particles are traced through the detector. In the tracing process the various interactions the particles undergo with the media, like energy loss, absorption or the creation of secondaries, are simulated (again by random number generation). Finally, it is decided whether a particular final state of the reaction is accepted or is rejected, on the basis of satisfying certain criteria for its detection. Then the efficiency ϵ for the detection of the reaction is calculated as the number N_{acc} of all accepted final states divided by the number N_{gen} of generated states:

$$\epsilon = \frac{N_{acc}}{N_{gen}} \,. \tag{3.146}$$

The program package Geant4 [452] is often used as a framework to simulate responses of detectors or detector parts. In Geant4 detector geometries can be defined, the different detector materials can be described and a number of different programs can be invoked, specialised to simulate various particle reactions. In general one would like to describe all processes exactly on a micro-physical level. Often it is however necessary to find a compromise between accuracy, computing time and adaptability.

In the simulations the individual particles are generally traced progressively in small steps and energy loss, absorption, or the creation of secondaries are computed according to their cross section probabilities. At high energies in particular, featuring collisions with high particle multiplicities, correspondingly large computing times arise. Hence desired precision and CPU-time investment must be balanced against each other. The process step sizes need to be optimised and cut-off energies must be defined up to which particles shall be traced before depositing their rest energy locally. Step sizes and cut-off energies depend on the specific requirements and are empirically adapted.

4

Movement of charge carriers in electric and magnetic fields

4.1 Introduction

The detection methods for charged particles most often exploit the ionisation of sensitive detector layers by particles passing through. An electric field moves the created charges towards the electrodes which sense the signals induced by this movement. In most detectors the charges are generated in gases (chapter 7), in liquids (e.g. in liquid-argon calorimeters, chapter 15) or in semiconductors (chapter 8). In gases and liquids the charge carriers are electrons and ions; in semiconductors they are electrons and 'holes'. Holes are missing electrons in the valence energy band which behave like (moving) positive charges.

For charge carriers moving in electric and magnetic fields one can distinguish ordered and unordered motions which are superimposed (fig. 4.1):

- A random, unordered motion with a velocity distribution which—in thermal equilibrium and without external fields—is classically described by the Maxwell–Boltzmann distribution[1]. An external electric field can 'heat up' the carrier motion and hence change the distribution.
- A drift motion whose direction is determined by the external electric and magnetic fields.

If a concentration gradient exists, for example when charges are created locally, the unordered motion leads to diffusion of the charge carriers causing a dispersion of the local charge distribution.

[1]The Maxwell–Boltzmann distribution is the classical approximation for both the Bose–Einstein distribution for bosons and the Fermi–Dirac distribution for fermions in the case that quantum effects are not dominating, that is, at low particle densities and/or high temperatures. The classical conditions are usually fulfilled for charge transport in gases. In semiconductors the (intrinsic) charge densities are higher and quantum effects play a role in interactions with the lattice. However, often the classical approximation can also be used for semiconductors, as for the charge carrier distribution in the conduction band (see eq. (8.13)).

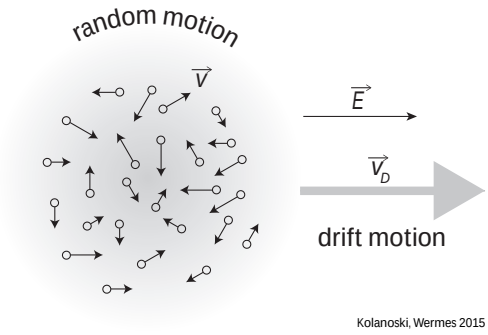


Fig. 4.1 Schematic illustration of the random, unordered motion of particles in a charge cloud and the superimposed drift motion in the direction of the electric field.

The drift velocity $\vec{v}_D$ results from an equilibrium between the accelerating electric force and a damping or friction force arising from collisions with surrounding atoms and molecules. In most cases the drift velocity is much smaller than the mean unordered velocity $\langle v \rangle$:

$$|\vec{v}_D| \ll \langle v \rangle \,. \tag{4.1}$$

In all cases considered in this chapter the velocities are much smaller than the velocity of light and the kinematics can be treated non-relativistically.

In section 4.2 the Boltzmann transport equation for charge carriers in a medium with external fields is introduced and solved under simplifying assumptions. With these solutions expressions for the drift velocities of charge carriers under the influence of external electric and magnetic fields are computed in section 4.3. In section 4.4 the dispersion of a charge cloud due to diffusion caused by the unordered motion will be investigated. The solutions of the Boltzmann transport equation will be discussed in sections 4.5 and 4.6 for ions and electrons in gases, respectively. For the description of the motion of electrons and holes in semiconductors lattice effects must be taken into account, as discussed in section 4.7.

The subject of this chapter is described in textbooks on statistical physics (e.g. [874, 544]) where, however, transport phenomena are often not sufficiently detailed. For example, the Boltzmann transport equation is usually only treated in the context of the kinetic gas theory covering the motion of ions in gases. For the more complex phenomena related to the movement of electrons in gases we will refer to the special literature within the respective text passages. The application to transport phenomena in semiconductors is described, for example, in [402].

4.2 Charge carrier transport: Boltzmann transport equation

The Boltzmann transport equation describes the evolution of the microscopic position and velocity distributions of charge carriers in a medium in dependence on external forces, like electric and magnetic fields, and other parameters like temperature and pressure. We consider in the following an ensemble of charged particles which can interact with each other and with the medium. From the solution of the equation, which needs as input the microscopic scattering cross sections and inelasticities, drift and diffusion are determined as collective properties of the distributions which are macroscopic observables of the charge carriers.

Solving the transport equation is very involved and in general only possible numerically. Therefore, in the following we shall make simplifying assumptions which nevertheless should permit a discussion of the characteristics of drift and diffusion.

In practice, more accurate results are obtained by numerically evaluating the transport equation. This is usually done for the electron motion in gases which exhibits particularly complex dependences on the mixing ratios of gases and on the external fields.

4.2.1 Boltzmann transport equation

In the following we consider a charge cloud in a medium with the phase space distribution $f(\vec{r},\vec{v},t)$. The probability dp to find a charge carrier at the time t in the phase space interval $[(\vec{r},\vec{v}),(\vec{r}+d\vec{r},\vec{v}+d\vec{v})]$ is

$$dp(\vec{r},\vec{v},t) = f(\vec{r},\vec{v},t)\, d^3\vec{r}\, d^3\vec{v}\,. \tag{4.2}$$

If the total number of charge carriers is conserved the integrated probablity can be normalised in the full phase space (PS):

$$\int_{\mathrm{PS}} f(\vec{r},\vec{v},t)\, d^3\vec{r}\, d^3\vec{v} = 1\,. \tag{4.3}$$

The evolution of the distribution in the phase space is described by the Boltzmann transport equation

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{d\vec{r}}{dt}\,\vec{\nabla}_{\vec{r}} f + \frac{d\vec{v}}{dt}\,\vec{\nabla}_{\vec{v}} f = \left.\frac{\partial f}{\partial t}\right|_{coll}\,. \tag{4.4}$$

The total time derivative df/dt on the left side is expanded in terms of the partial derivatives in the middle part. The first term of the partial derivatives occurs in the case of explicit time dependence of f and vanishes in the stationary case. The two next terms describe the changes of f by motions in phase space where the first one describes the motions in position space, corresponding to the diffusion, and the second one the motions in velocity space. The acceleration $d\vec{v}/dt$ in the second term originates from the external electric and magnetic fields. The expression on the right-hand side is called the 'collision integral' and is a functional which contains integrals over f, thus turning the Boltzmann transport equation into an integro-differential equation. The collision integral accounts for the friction due to the collisions with the atoms or molecules of the medium, which are described by measured or calculated cross sections for the microscopic elastic or inelastic scattering processes. The solutions are the microscopic distributions in phase space from which the macroscopic quantities, like drift and diffusion, are derived. On the other hand, the microscopic cross sections and inelasticities can be inferred from the measured macroscopic quantities.

Acceleration term. Let the acceleration be given by the Lorentz force,

$$m\frac{d\vec{v}}{dt} = q\left(\vec{E} + \vec{v}\times\vec{B}\right)\,, \tag{4.5}$$

where the force acts on a particle with mass m and charge q ($q = \pm e$ or multiples thereof, where e is the elementary charge). The coordinate system can be chosen such that the fields lie in a plane spanned by the 2- and 3-axes, as displayed in fig. 4.2:

$$\vec{E} = (0,0,E) \qquad \text{and} \qquad \vec{B} = (0,B_2,B_3)\,. \tag{4.6}$$

Then (4.5) yields

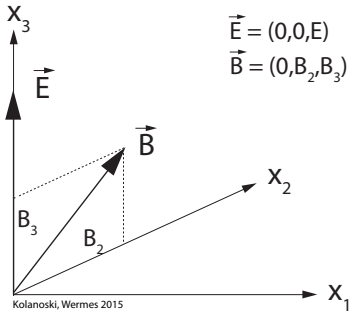


Fig. 4.2 Coordinate system defined by the fields $\vec{E}$ and $\vec{B}$. The 3-axis is aligned along the E-field and the 1-axis along the cross product $\vec{E} \times \vec{B}$. The 2-axis completes a right-handed orthogonal system.

$$\frac{d\vec{v}}{dt} = \frac{q}{m} \begin{pmatrix} 0 \\ 0 \\ E \end{pmatrix} + \begin{pmatrix} v_2\omega_3 - v_3\omega_2 \\ -v_1\omega_3 \\ v_1\omega_2 \end{pmatrix} , \tag{4.7}$$

where

$$\omega_i = \frac{q}{m} B_i \qquad (i = 1, 2, 3) \tag{4.8}$$

are the cyclotron frequencies.

4.2.2 Collision integral

The Boltzmann transport equation (4.4) means that in the collision-free case, which implies that the collision integral vanishes, the density in the phase space does not change thus yielding $df/dt = 0$. This follows from Liouville's theorem, which is also valid for systems on which external forces act as long as those forces are conservative (see e.g. [874]). Only the interactions of the particles among each other, described by the collision integral, can alter the phase space density.

The collision integral describes the scattering of particles into a phase space element and out of it. If the interactions occur elastically by two-body collisions $1+2 \leftrightarrow 3+4$, the integral reads (setting $\vec{v}_4 = \vec{v}$ for the particle under consideration):

$$\left.\frac{\partial f}{\partial t}\right|_{\text{coll}} = \int W(\vec{v}_1, \vec{v}_2; \vec{v}_3, \vec{v}) \left\{ f(\vec{r}, \vec{v}_1, t) f(\vec{r}, \vec{v}_2, t) - f(\vec{r}, \vec{v}_3, t) f(\vec{r}, \vec{v}, t) \right\} \mathrm{d}\vec{v}_1 \mathrm{d}\vec{v}_2 \mathrm{d}\vec{v}_3 \, . \tag{4.9}$$

The integral specifies how at a given time t and position $\vec{r}$ the phase space density $f(\vec{r}, \vec{v}, t)$ changes due to the in and out scattering of particles with velocity $\vec{v}$. The function $W(\vec{v}_1, \vec{v}_2; \vec{v}_3, \vec{v})$ gives the probability that the particles 1 and 2 which meet in space and time at $\vec{r}$, t scatter into the second pair of velocities and vice versa. The term $W(\vec{v}_1, \vec{v}_2; \vec{v}_3, \vec{v}) f(\vec{r}, \vec{v}_1, t) f(\vec{r}, \vec{v}_2, t)$ specifies the increase which follows from the scattering of particle 1 and 2 yielding a particle in the considered phase space element around $\vec{v}$, while the other particle obtains the velocity $\vec{v}_3$ (fig. 4.3(a)). Correspondingly, the other (negative) term describes the decrease of the density in the considered phase space element. That means, the particle with velocity $\vec{v}$ leaves the element by scattering with a particle of velocity $\vec{v}_3$ while the scattering partners assume the velocities $\vec{v}_1$ and $\vec{v}_2$ (fig. 4.3(b)).

The scattering probability $W(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v})$ contains the differential cross sections and the constraints due to energy and momentum conservation. The integral in (4.9) has to be taken over all possible scattering partners 1, 2, 3 in the initial and final states.

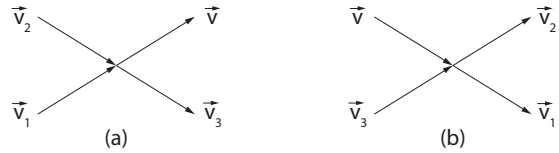


Fig. 4.3 Two-body scattering processes at the position $\vec{r}$ and time t. (a) A particle is scattered into the velocity element around $\vec{v}$ and (b) a particle is scattered out of the velocity element around $\vec{v}$.

In (4.9) it is assumed that only one particle type is involved and that the scattering processes are reversible, yielding

$$W(\vec{v}_1, \vec{v}_2; \vec{v}_3, \vec{v}_4) = W(\vec{v}_3, \vec{v}_4; \vec{v}_1, \vec{v}_2)\,. \tag{4.10}$$

In general, one has to sum over all particle types and one has to account for possible inelastic interactions where the energy can be redistributed between the particle types or the energy can be exchanged with the surroundings. Additionally, particle sources and sinks can occur, like for example the generation and absorption of secondary charge carriers. However, in the following simplified consideration we shall assume particle number conservation.

In practice the evaluation of the collision integral is the real challenge for solving the Boltzmann transport equation. As mentioned before, due to the collision integral it becomes an integro-differential equation and can usually only be solved numerically. In the following we discuss essential properties of velocity distributions using approximations.

4.2.3 Special solutions of the Boltzmann transport equation

4.2.3.1 Maxwell–Boltzmann distribution

The stationary, position-independent and force-free case implies $f(\vec{r}, \vec{v}, t) = f(\vec{v})$ and the expressions with partial derivatives in (4.4) vanish. Then one obtains for the collision integral:

$$\left.\frac{\partial f}{\partial t}\right|_{\text{coll}} = 0\,. \tag{4.11}$$

For the elastic scattering considered in (4.9) a sufficient condition to obtain an equilibrium solution f_0, which fulfils (4.11), obviously is

$$f_0(\vec{v}_1) f_0(\vec{v}_2) = f_0(\vec{v}_3) f_0(\vec{v}_4)\,. \tag{4.12}$$

One can show that in this case it is also a necessary condition, see for example [874]. The velocities on the left and the right side of (4.12) are linked by energy-momentum conservation (as also in the integral (4.9)). In statistical physics (e.g. [874]) the solutions are shown to always take the form of the Maxwell–Boltzmann equation:

$$f_0(\vec{v}) = C\,\exp(-Av^2)\,, \tag{4.13}$$

where C and A are constants. One can easily see that this function solves (4.12):

$$C^2\,\exp(-A(v_1^2 + v_2^2)) = C^2\,\exp(-A(v_3^2 + v_4^2))\,. \tag{4.14}$$

The equality follows from energy conservation: $v_1^2 + v_2^2 = v_3^2 + v_4^2$.

4.2.3.2 Relaxation approximation

To give a not too complicated example for a solution of the Boltzmann equation in the case of a non-vanishing $\left.\frac{\partial f}{\partial t}\right|_{\text{coll}}$ we resort to the so-called relaxation approximation.

In a closed system the interactions take care that the energy and momentum distributions are stationary on a statistical average. Fluctuations in these distributions are equalised with a time constant τ, the 'relaxation time', during which the system returns to the equilibrium state. Since the reason for the relaxation is the interactions of the particles, the time τ is related to the mean time between two collisions, the 'collision time' (see section 4.3.).

Let us assume that the perturbation, for example by external fields, is switched off at $t = 0$ and that the relaxation is proportional to the deviation of the fluctuation from the equilibrium distribution, the *effective* collison integral in the relaxation approximation can be set to

$$\frac{\partial f}{\partial t} = \left.\frac{\partial f}{\partial t}\right|_{\text{coll}} = -\frac{f - f_0}{\tau}\,. \tag{4.15}$$

This equation has the solution

$$f(t) = f_0 + (f - f_0)\,\mathrm{e}^{-t/\tau}, \tag{4.16}$$

describing that the perturbed system returns to its equilibrium with a time constant τ.

The relaxation approximation can be used for systems which are stationary on a statistical average. This is often fulfilled in the cases of charge transport in detector media, that is, in the cases relevant for this book. The examples which we give in the following yield quite realistic results and can be used for simulations of charge transport in media.

4.2.4 Perturbation by external fields

We now consider a system which is shifted from the original equilibrium distribution $f_0(\vec{v})$ to a new position and time independent distribution $f(\vec{v})$ by applying external fields. With the Lorentz force in (4.5) the Boltzmann transport equation (4.4) has the form

$$\frac{\partial f}{\partial t} + \frac{q}{m}\left(\vec{E} + \vec{v} \times \vec{B}\right) \vec{\nabla}_{\vec{v}} f = \left.\frac{\partial f}{\partial t}\right|_{coll}\,. \tag{4.17}$$

For a stationary distribution f, on a time-scale much larger than the relaxation time, the time derivative of f vanishes.

4.2.4.1 With electric field only

To begin with, we assume that the external force is given by an electric field only. If the field is sufficiently weak so that the equilibrium state is not substantially disturbed, we can apply the relaxation approximation:

$$\frac{\partial f}{\partial t} + \frac{q}{m}\vec{E}\,\vec{\nabla}_{\vec{v}} f \approx \frac{q}{m}\vec{E}\,\vec{\nabla}_{\vec{v}} f_0 = -\frac{f - f_0}{\tau}\,. \tag{4.18}$$

Here the additional approximations of applying the differential operators to f_0 instead of f and of neglecting $\partial f/\partial t$ have been used.

For solving the equation we express the absolute value of the velocity of a particle through its kinetic energy:

$$\epsilon = \frac{1}{2} m v^2 \,. \tag{4.19}$$

The equilibrium distribution f_0 depends only via the energy ϵ on the velocity. Therefore the velocity gradient in (4.18) can be written as

$$\vec{\nabla}_{\vec{v}} f_0 = \frac{\partial f_0}{\partial \epsilon} \vec{\nabla}_{\vec{v}} \epsilon = m \vec{v} \frac{\partial f_0}{\partial \epsilon} \,. \tag{4.20}$$

Given this equation, the solution of (4.18) is

$$f = f_0 - \frac{q}{m} \tau \vec{E} \vec{\nabla}_{\vec{v}} f_0 = f_0 - q \vec{E} \tau \vec{v} \frac{\partial f_0}{\partial \epsilon} = f_0 - q E \tau v_3 \frac{\partial f_0}{\partial \epsilon} \,. \tag{4.21}$$

On the right-hand side we used the definition of the coordinate system as in (4.6) yielding $\vec{v}\vec{E} = v_3 E$. Obviously the resulting distribution f is anisotropic with a preference in field direction (3-axis).

4.2.4.2 Generalisation to the full Lorentz force

We now consider the Boltzmann transport equation in the form of (4.17) including the full Lorentz force in the relaxation approximation:

$$\frac{\partial f}{\partial t} + \frac{q}{m} \left(\vec{E} + \vec{v} \times \vec{B} \right) \vec{\nabla}_{\vec{v}} f = -\frac{f - f_0}{\tau} = -\frac{\delta f}{\tau} \,. \tag{4.22}$$

In this case, however, we cannot employ the same approximation for the velocity gradient as in (4.20) because the then arising triple product $\vec{v} \cdot (\vec{v} \times \vec{B})$ vanishes. As a result the magnetic field would have no effect on the distribution which, however, does not correspond to the observation. Therefore the effect of the B-field in (4.22) has to be accounted for by evaluating the gradient in higher order containing the anisotropy due to the E-field. Accordingly, we write the gradient as

$$\vec{\nabla}_{\vec{v}} f = \vec{\nabla}_{\vec{v}} (f_0 + \delta f) = m \vec{v} \frac{\partial f_0}{\partial \epsilon} + \vec{\nabla}_{\vec{v}} \delta f \,. \tag{4.23}$$

Inserting this into (4.22) and using the approximation $\partial f / \partial t \approx \partial f_0 / \partial t = 0$ yields

$$q \vec{E} \vec{v} \frac{\partial f_0}{\partial \epsilon} + \frac{q}{m} \left(\vec{v} \times \vec{B} \right) \vec{\nabla}_{\vec{v}} \delta f = -\frac{\delta f}{\tau} \,. \tag{4.24}$$

Here again for the E-field the velocity gradient is applied to the f_0 approximation while for the B-field this contribution vanishes and only the δf term contributes.

For the solution of (4.24) we make the ansatz that an effective E-field $\vec{A}$ exists with which a solution can be constructed corresponding to (4.21):

$$\delta f = -\frac{q}{m} \tau \vec{A} \vec{\nabla}_{\vec{v}} f_0 = -q \tau \vec{v} \vec{A} \frac{\partial f_0}{\partial \epsilon} \,. \tag{4.25}$$

From this equation the gradient of δf is calculated as

$$\vec{\nabla}_{\vec{v}} \delta f = -q \tau \vec{A} \frac{\partial f_0}{\partial \epsilon} \,. \tag{4.26}$$

Inserting δf and $\vec{\nabla}_{\vec{v}}\,\delta f$ into (4.24) yields

$$\vec{v}\vec{E} - \frac{q}{m}\tau \underbrace{\left(\vec{v}\times\vec{B}\right)\vec{A}}_{=(\vec{B}\times\vec{A})\vec{v}} = \vec{v}\vec{A} \qquad \Rightarrow \qquad \vec{E} - \frac{q}{m}\tau\left(\vec{B}\times\vec{A}\right) = \vec{A}\,. \tag{4.27}$$

From the right equation one obtains a relation between $\vec{E}$, $\vec{B}$ and $\vec{A}$ which is independent of the velocity vector. The solution for $\vec{A}$ can be represented by the components in the directions of $\vec{E}$, $\vec{B}$ and $\vec{E}\times\vec{B}$, hence as a linear combination $\vec{A} = a\,\vec{E} + b\,\vec{B} + c\,(\vec{E}\times\vec{B})$. Inserting this expression into (4.27) yields by equating the coefficients of $\vec{E}$, $\vec{B}$, $\vec{E}\times\vec{B}$:

$$\vec{A} = \frac{1}{1+\omega^2\tau^2}\left(\vec{E} + \frac{(\vec{E}\,\vec{B})\vec{B}}{B^2}\omega^2\tau^2 + \frac{\vec{E}\times\vec{B}}{B}\omega\tau\right), \tag{4.28}$$

where $\omega = qB/m$ is the cyclotron frequency. Inserting $\vec{A}$ into the expression for δf in (4.25) yields the approximated solution for the distribution $f = f_0 + \delta f$:

$$f = f_0 - q\,\tau\,\vec{v}\,\vec{A}\,\frac{\partial f_0}{\partial\epsilon} = f_0 - qE\,\tau\frac{1}{1+\omega^2\tau^2}\,\vec{v}\begin{pmatrix} -\omega_2\tau \\ \omega_2\omega_3\tau^2 \\ 1+\omega_3^2\tau^2 \end{pmatrix}\frac{\partial f_0}{\partial\epsilon}\,. \tag{4.29}$$

The right-hand side uses the definitions of the field components and the cyclotron frequencies in (4.6) to (4.8) and fig. 4.2.

4.3 Drift velocity

One of the macroscopic observables which we will derive from the treatment of charge transport using the Boltzmann transport equation is the drift velocity of the charge carriers resulting from the external fields. The drift velocity $\vec{v}_D$ is the mean value of the velocity vectors of the particle ensemble with respect to the distribution function f:

$$\vec{v}_D = \langle\vec{v}\,\rangle = \int \vec{v}\,f(\vec{v})\,d^3\vec{v}\,. \tag{4.30}$$

A non-vanishing drift velocity results only if $f(\vec{v})$ has an asymmetry in $\vec{v}$ space. In order to describe such an asymmetry in terms of angular distributions we express the velocity components using spherical coordinates:

$$v_1 = v\sin\theta\cos\phi, \qquad v_2 = v\sin\theta\sin\phi, \qquad v_3 = v\cos\theta\,. \tag{4.31}$$

In addition we express the absolute value of the velocity by the kinetic energy:

$$\epsilon = \frac{1}{2}mv^2 \quad \Rightarrow \quad v = \sqrt{\frac{2\epsilon}{m}}, \quad dv = \frac{d\epsilon}{\sqrt{2m\epsilon}}, \quad v^2dv = \sqrt{\frac{2\epsilon}{m}}\frac{1}{m}d\epsilon\,. \tag{4.32}$$

Then the differential velocity distribution becomes

$$f(\vec{v})\,dv_1\,dv_2\,dv_3 = f(\vec{v})\,v^2\,dv\,d\cos\theta\;d\phi = f(\epsilon,\cos\theta,\phi)\sqrt{\frac{2\epsilon}{m^3}}\,d\epsilon\,d\cos\theta\,d\phi\,. \tag{4.33}$$

Here $f(\epsilon, \cos\theta, \phi)$ is the original function f with the velocity components expressed by the variables $\epsilon,\ \cos\theta,\ \phi$ corresponding to (4.31) and (4.32). The velocity distribution f, which we mostly refer to in this chapter, has to be distinguished from the energy–angle distribution F which is defined by

$$f(\vec{v})\, v^2\, dv = f(\epsilon, \cos\theta, \phi) \sqrt{\frac{2\epsilon}{m^3}}\, d\epsilon = F(\epsilon, \cos\theta, \phi)\, d\epsilon\,. \tag{4.34}$$

Since both the velocity distribution and the energy–angle distribution are probability densities they are normalised:

$$1 = \int_\Omega \int_0^\infty f(\vec{v})\, v^2 dv\, d\Omega = \int_\Omega \int_0^\infty f(\epsilon, \cos\theta, \phi) \sqrt{\frac{2\epsilon}{m^3}}\, d\epsilon\, d\Omega = \int_\Omega \int_0^\infty F(\epsilon, \cos\theta, \phi)\, d\epsilon\, d\Omega\,. \tag{4.35}$$

4.3.1 Drift velocity with E-field only

With the approximate distribution function (4.21) the drift velocity becomes according to (4.30):

$$v_{D,i} = \int_\Omega \int_0^\infty v_i\, f(\vec{v})\, v^2 dv\, d\Omega = \int_\Omega \int_0^\infty v_i\, f_0\, v^2 dv\, d\Omega \; - \int_\Omega \int_0^\infty q\, E\, \frac{\partial f_0}{\partial \epsilon}\, \tau\, v_i v_3\, v^2 dv\, d\Omega\,. \tag{4.36}$$

The first term on the right-hand side, which represents the averaging of the velocity component v_i over the isotropic part f_0 of the distribution, vanishes for all components i. Since v_i and v_3 for $i = 1, 2$ are uncorrelated the second term also vanishes for $i = 1, 2$ but remains finite for $i = 3$. This can be seen when averaging the terms $v_i\, v_3$, which contain the angular dependences of $f(\vec{v})$, over the space angle. Expressing $v_i\, v_3$ by spherical coordinates according to (4.31) one finds

$$\int_\Omega v_i\, v_3\, d\Omega = \int_\Omega v^2 \sin\phi \sin\theta \cos\theta\, d\Omega = \int_\Omega v^2 \cos\phi \sin\theta \cos\theta\, d\Omega = 0 \qquad (i = 1, 2)\,, \tag{4.37}$$

$$\int_\Omega v_i\, v_3\, d\Omega = \int_\Omega v^2 \cos^2\theta\, d\Omega = \frac{4\pi}{3} v^2 \qquad (i = 3)\,. \tag{4.38}$$

Thus the drift velocity has only the one component in E-field direction resulting from the anisotropy of the distribution $f(\vec{v})$ with respect to this direction:

$$v_{D,3} = -\frac{4\pi}{3} \int_0^\infty q\, E\, \frac{\partial f_0}{\partial \epsilon}\, \tau\, v^4\, dv\,. \tag{4.39}$$

The remaining integration over the velocity can be turned into an integration over the energy using (4.32):

$$v_{D,3} = -\frac{4\pi}{3} \frac{qE}{m} \int_0^\infty \tau \left(\frac{2\epsilon}{m}\right)^{3/2} \frac{\partial f_0}{\partial \epsilon} d\epsilon = -4\pi \frac{2}{3} \frac{qE}{m^2} \int_0^\infty \lambda\, \epsilon\, \frac{\partial f_0}{\partial \epsilon} d\epsilon\,. \tag{4.40}$$

On the right side of the equation, τ has been substituted by $\lambda = \tau v$, which is the mean free path between collisions (see (3.7)),

$$\lambda = \frac{1}{n\sigma}\,, \tag{4.41}$$

with the particle density n and the interaction cross section σ. Depending on the context the drift velocities are either expressed using the mean free path or the collision time τ. For charge transport in gases, often λ is used and in semiconductors mostly τ.

4.3.1.1 Drift velocity dependence on the mean free path

By partial integration of (4.40) one finds

$$\begin{aligned} v_{D,3} &= -4\pi\frac{2}{3}\frac{qE}{m^2}\int_0^\infty \lambda\,\epsilon\frac{\partial f_0}{\partial\epsilon}d\epsilon = 4\pi\frac{2}{3}\frac{qE}{m^2}\left(\int_0^\infty f_0\,\frac{\partial}{\partial\epsilon}(\lambda\epsilon)d\epsilon - \underbrace{\Big[f_0\,\lambda\epsilon\Big]_0^\infty}_{=0}\right) \\ &= 4\pi\frac{2}{3}\frac{qE}{m^2}\int_0^\infty f_0\,\left(\frac{\partial\lambda}{\partial\epsilon}\epsilon + \lambda\right)d\epsilon = 4\pi\frac{2}{3}\frac{qE}{m^2}\left(\int_0^\infty \lambda\,f_0\,d\epsilon + \int_0^\infty f_0\,\epsilon\,\frac{\partial\lambda}{\partial\epsilon}d\epsilon\right)\,. \end{aligned} \tag{4.42}$$

The last term in the first row vanishes because the energy distribution $f_0(\epsilon)$ always goes to zero for $\epsilon \to \infty$ (mostly exponentially). The integrals in the last term of the second row can be rewritten as averages of λ/v and $d\lambda/dv$ with respect to the distribution $f_0(\vec{v})$:

$$\begin{aligned} \left\langle\frac{\lambda}{v}\right\rangle &= \int\frac{\lambda}{v}f_0\,v^2dv\,d\Omega = 4\pi\int_0^\infty\frac{\lambda}{v}f_0\,\frac{v}{m}d\epsilon = \frac{4\pi}{m}\int_0^\infty\lambda f_0\,d\epsilon\,, \\ \left\langle\frac{\partial\lambda}{\partial v}\right\rangle &= \int\frac{\partial\lambda}{\partial v}\,f_0\,v^2dv\,d\Omega = 4\pi\int_0^\infty\frac{2\epsilon}{v}\frac{\partial\lambda}{\partial\epsilon}\,f_0\,\frac{v}{m}d\epsilon = 2\,\frac{4\pi}{m}\int_0^\infty\epsilon\frac{\partial\lambda}{\partial\epsilon}\,f_0\,d\epsilon\,, \end{aligned} \tag{4.43}$$

where

$$\epsilon = \frac{1}{2}mv^2, \qquad \epsilon\frac{\partial\lambda}{\partial\epsilon} = \frac{1}{2}v\frac{\partial\lambda}{\partial v} \qquad \text{and} \qquad d\epsilon = m\,v\,dv$$

have been used.

Replacing the integrals in (4.42) by the averages (4.43) the drift velocity becomes

$$v_{D,3} = \frac{qE}{m}\left(\frac{2}{3}\left\langle\frac{\lambda}{v}\right\rangle + \frac{1}{3}\left\langle\frac{d\lambda}{dv}\right\rangle\right) = \pm\mu E\,. \tag{4.44}$$

The right-hand side of (4.44) introduces the mobility μ which in isotropic media and without magnetic field is generally defined as a positive quantity by the equation

$$\vec{v}_D = \pm\mu\vec{E}\,. \tag{4.45}$$

The sign in (4.45) has to be chosen according to that of the charge q of the drifting particle.

If the cross section does not vary too strongly with energy the first term in (4.44) dominates, for example for ions in gases. With the mean collision time

$$\langle\tau\rangle = \left\langle \frac{\lambda}{v} \right\rangle \tag{4.46}$$

one obtains

$$v_{D,3} = \frac{2}{3}\frac{qE}{m}\langle\tau\rangle = \pm\mu E\,, \qquad \mu = \frac{2}{3}\frac{|q|}{m}\langle\tau\rangle\,. \tag{4.47}$$

The prefactor 2/3 results when the acceleration by the electric field acts on an isotropic thermal velocity distribution (see the detailed derivation in section 8 of [943]). Since the assumptions (e.g. isotropy, $v_D \ll \langle v\rangle$ or velocity dependences of the cross sections) can differ, the prefactors can have different values. In addition, note that in the approximation (4.15) τ originally had the meaning of a relaxation time, but now it is interpretated as collision time. The interpretation of τ as collision time corresponds to the assumption that on average at each collision the velocities are isotropically redistributed and the energy gained by a particle from the field is dispersed, hence the system is brought back to its respective equilibrium. If for example, the mean collision time τ instead of the mean free path λ is assumed to be constant, (4.44) with $d\lambda/dv = \tau$ yields

$$v_{D,3} = \frac{qE}{m}\tau = \pm\mu E\,, \qquad \mu = \frac{|q|}{m}\tau \qquad (\tau = const)\,. \tag{4.48}$$

4.3.1.2 Drift velocity dependence on the collision time (energy weighted average)

In semiconductor physics often an 'energy weighted average' of the collision time τ is introduced instead of (4.46) (see e.g. [402]). To define such an average we go back to (4.40):

$$v_{D,3} = -\frac{4\pi}{3}\frac{qE}{m}\int_0^\infty \tau\left(\frac{2\epsilon}{m}\right)^{3/2}\frac{\partial f_0}{\partial\epsilon}\,d\epsilon\,. \tag{4.49}$$

By partial integration and assuming again that f_0 approaches zero for $\epsilon\to\infty$, we find an integral relation between f_0 and $\partial f_0(\epsilon)/\partial\epsilon$:

$$\int_0^\infty f_0(\epsilon)\epsilon^{1/2}d\epsilon = -\frac{2}{3}\int_0^\infty \frac{\partial f_0(\epsilon)}{\partial\epsilon}\epsilon^{3/2}d\epsilon\,. \tag{4.50}$$

Since f_0 is the unperturbed equilibrium distribution we assume it to be normalised as in (4.35) for f, which means treating the perturbation term as being small. Combining the normalisation of f_0 with the relation (4.50) yields

$$1 = \frac{4\pi}{m}\sqrt{\frac{2}{m}}\int_0^\infty f_0(\epsilon)\,\epsilon^{1/2}d\epsilon = -\frac{2}{3}\frac{4\pi}{m}\sqrt{\frac{2}{m}}\int_0^\infty \frac{\partial f_0(\epsilon)}{\partial\epsilon}\epsilon^{3/2}d\epsilon\,. \tag{4.51}$$

We can now divide (4.49) by the expression on the right-hand side of (4.51) which is unity and thus does not change the relation for the drift velocity in (4.49). This yields a modified expression for the drift velocity:

$$v_{D,3} = \mu\,E = \frac{qE}{m}\frac{\int_0^\infty \tau\,\epsilon^{3/2}\frac{\partial f_0}{\partial\epsilon}d\epsilon}{\int_0^\infty \epsilon^{3/2}\frac{\partial f_0}{\partial\epsilon}d\epsilon} =: \frac{qE}{m}\langle\tau\rangle_\epsilon \qquad \Longrightarrow \qquad \mu = \frac{|q|}{m}\langle\tau\rangle_\epsilon\,. \tag{4.52}$$

The equation defines the 'energy weighted averaging' which we distinguish from the usual averaging (over the distribution f) by a subscript ϵ on the average symbol. Since

(4.52) has been directly derived from (4.49) both equations lead to identical results for the drift velocity.[2] The introduction of an 'energy weighted averaging' is helpful if f_0 can be approximated by the Maxwell–Boltzmann distribution and τ depends on the energy by a power law. In this case (4.52) can be solved analytically (see e.g. [893,402] and also eq. (4.142) on page 126).

In general the 'energy weighted average' of a function $g(\epsilon)$ is defined by

$$\langle g \rangle_\epsilon = \frac{\int_0^\infty g(\epsilon)\, \epsilon^{3/2}\, \frac{\partial f_0}{\partial \epsilon} d\epsilon}{\int_0^\infty \epsilon^{3/2}\, \frac{\partial f_0}{\partial \epsilon} d\epsilon} \,. \tag{4.53}$$

4.3.2 Drift velocities with E- and B-fields

With the external E- and B-field components as given in (4.6) the distribution function (4.29) yields with a calculation analogous to (4.36):

$$v^B_{D,1} = -\frac{4\pi}{3}\frac{qE}{m}\int_0^\infty \tau \frac{\omega_2 \tau}{1+\omega^2\tau^2}\left(\frac{2\epsilon}{m}\right)^{3/2}\frac{\partial f_0}{\partial \epsilon} d\epsilon = \frac{qE}{m}\left\langle \tau \frac{-\omega_2\tau}{1+\omega^2\tau^2}\right\rangle_\epsilon , \tag{4.54}$$

$$v^B_{D,2} = \frac{4\pi}{3}\frac{qE}{m}\int_0^\infty \tau \frac{\omega_2\omega_3 \tau^2}{1+\omega^2\tau^2}\left(\frac{2\epsilon}{m}\right)^{3/2}\frac{\partial f_0}{\partial \epsilon} d\epsilon = \frac{qE}{m}\left\langle \tau \frac{\omega_2\omega_3\tau^2}{1+\omega^2\tau^2}\right\rangle_\epsilon , \tag{4.55}$$

$$v^B_{D,3} = \frac{4\pi}{3}\frac{qE}{m}\int_0^\infty \tau \frac{1+\omega_3^2\tau^2}{1+\omega^2\tau^2}\left(\frac{2\epsilon}{m}\right)^{3/2}\frac{\partial f_0}{\partial \epsilon} d\epsilon = \frac{qE}{m}\left\langle \tau \frac{1+\omega_3^2\tau^2}{1+\omega^2\tau^2}\right\rangle_\epsilon . \tag{4.56}$$

The right-hand side displays an energy weighted average of a more complex function of the collision time using (4.53). Simple conclusions that are independent of the averaging procedure are as follows:

(1) Without an E-field there is no drift motion: from $E = 0$ follows $v^B_{D,i} = 0$ for all i.
(2) $E = E_3 \neq 0$ always yields $v^B_{D,3} \neq 0$, independent of the B-field.
(3) For $B_2 = 0$ ($\vec{B}$ parallel to $\vec{E}$) yields only $v^B_{D,3} \neq 0$ (drift as for $B = 0$).
(4) $B_3 = 0$ ($\vec{B}$ perpendicular to $\vec{E}$) yields

$$v^B_{D,1},\ v^B_{D,3} \neq 0 \quad \text{and} \quad v^B_{D,2} = 0 \tag{4.57}$$

(the drift motion takes place in the plane perpendicular to the B-field).

4.3.2.1 B perpendicular to E

For perpendicular fields $\vec{E}$ and $\vec{B}$, hence $B_3 = \omega_3 = 0$, $B_2 \neq 0$, $\omega_2 = \omega \neq 0$, we want to estimate the ratio of the absolute values of the drift velocities with and without magnetic field. For that purpose we use the right-hand sides of (4.54) to (4.56) without averaging, which means that τ is considered to be an effective collision time:

$$\frac{v^B_D}{v^0_D} = \frac{\sqrt{(v^B_{D,1})^2 + (v^B_{D,3})^2}}{|v^0_{D,3}|} = \frac{1}{\sqrt{1+\omega^2\,\tau^2}} \qquad (<1)\,. \tag{4.58}$$

We see that the drift velocity is smaller with magnetic field than without. The explanation is that the distance between collisions projected onto the drift direction is shorter than the actual curved path of the particles in the magnetic field (see fig. 4.4).

[2] The term 'energy weighted averaging' is somewhat misleading because the integration does not imply the distribution function itself and the energy factor only formally looks like a weight.

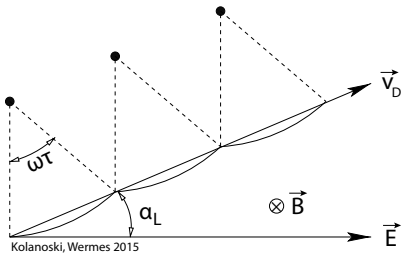


Fig. 4.4 Descriptive explanation of the Lorentz angle. The electron moves in the magnetic field on a curved orbit, here presented by a segment of a circle. During the collision time τ the electron has turned its direction by the angle $\omega\tau$. After each collision follows a similar segment of a circle if one assumes that the velocity vector vanishes on average after a collision. In contrast to this simplified picture, however, the segments are stochastically distributed. The whole sequence of such microscopic orbit segments yields a macroscopic change in the drift direction with respect to the direction of the electric field by the Lorentz angle α_L.

Figure 4.4 also shows that between the drift velocity and the accelerating E-field an angle emerges, the 'Lorentz angle':

$$\tan\alpha_L = \frac{|v^B_{D,1}|}{v^B_{D,3}} = \omega\tau\,. \tag{4.59}$$

According to fig. 4.4 the Lorentz angle $\alpha_L \approx \omega\tau$ can easily be understood as the average deflection angle of an electron during the (effective) collision time.

In the approximation for low magnetic fields, $1+\omega^2\tau^2 \to 1$, the drift components in (4.54) to (4.56) become by energy weighted averaging:

$$v^B_{D,1} = -\frac{qE}{m}\omega_2 \left\langle\tau^2\right\rangle_\epsilon\,, \qquad v^B_{D,2} = 0\,, \qquad v^B_{D,3} = \frac{qE}{m}\left\langle\tau\right\rangle_\epsilon\,. \tag{4.60}$$

Characteristic are the dependences of the drift component in the E-field direction, $v^B_{D,3}$, on τ and that of the drift component generated by the B-field, $v^B_{D,1}$, on τ^2. Then the Lorentz angle can be calculated as

$$\tan\alpha_L = \omega\frac{\left\langle\tau^2\right\rangle_\epsilon}{\left\langle\tau\right\rangle_\epsilon} = \omega\left\langle\tau\right\rangle_\epsilon\, r_H\,. \tag{4.61}$$

On the right side we introduced the so-called 'Hall factor' r_H which is used in semiconductor physics to describe the ratio of the mobilities measured by the Hall effect to the mobility given by the conductivity (see section 4.7.3). For small magnetic fields r_H is given by

$$r_H = \frac{\left\langle\tau^2\right\rangle_\epsilon}{\left\langle\tau\right\rangle^2_\epsilon}\,. \tag{4.62}$$

4.3.2.2 B parallel to E

This case is given by $\omega_2 = 0$ and $\omega = \omega_3 \neq 0$, and the drift velocity has only the component in the direction of both fields. From (4.56) one concludes that the drift velocity is in this case the same as without magnetic field, that is to say

$$\vec{v}^B_D = \vec{v}^0_D\,. \tag{4.63}$$

While the drift velocity does not change the transverse diffusion (perpendicular to the fields) becomes smaller, as further discussed in section 4.6.4.5.

4.4 Diffusion without external fields

A spatially inhomogeneous distribution of the charge carriers will be equalised by diffusion. In the force-free case the evolution towards a stationary, equalised distribution is described by the Boltzmann transport equation (4.4) without a force term:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{d\vec{r}}{dt}\vec{\nabla}_{\vec{r}} f = \left.\frac{\partial f}{\partial t}\right|_{coll} . \tag{4.64}$$

The particles diffuse along the spatial gradients with a diffusion speed determined by the collision integral.

The diffusion equations known as Fick's laws can be derived from the Boltzmann transport equation (shown in books on statistical physics, e.g. [544]). In the following we want to start from these laws without retracing their derivation.

4.4.1 Diffusion equation

We consider a spatially localised charge cloud with a time dependent distribution $f(\vec{r}, \vec{v}, t)$ in phase space, here without external fields. Such a localised charge cloud will spread out due to the thermal motions, which means the charges diffuse into the surroundings. The spatial number density results from the integral over the volume in velocity space:

$$\rho(\vec{r}, t) = N_0 \int f(\vec{r}, \vec{v}, t) d^3\vec{v} \, . \tag{4.65}$$

Here N_0 is the number of particles in the charge cloud at time zero, and

$$N(t) = \int \rho(\vec{r}, t) d^3\vec{r} \tag{4.66}$$

is this number at an arbitrary time t. In general, $N = N(t)$ is a function of time. It will only be constant if no (free) charges are lost, for example by recombination or attachment, and if no new charges are created. In this case the continuity equation holds:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}\vec{j}_D = 0 \, . \tag{4.67}$$

According to Fick's first law the diffusion current $\vec{j}_D$ flows in the direction of the density gradient with the diffusion coeffcient D as a proportionality constant:[3]

$$\vec{j}_D = -D\vec{\nabla}\rho \, . \tag{4.68}$$

Inserting this into (4.67) yields the diffusion equation, referred to as Fick's second law:

$$\frac{\partial \rho}{\partial t} - D\Delta\rho = 0 \, . \tag{4.69}$$

This differential equation can be solved by a spatial Gaussian distribution:

[3] The diffusion coefficient D is a (scalar) constant only if the diffusion is isotropic. In general D enters into (4.68) as a tensor (see e.g. [325, 217]).

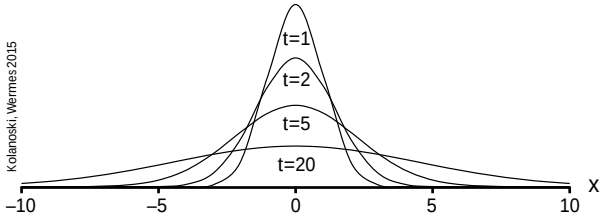


Fig. 4.5 Time evolution of diffusion in one dimension for particles that were created at the origin at the time $t = 0$ (in arbitrary units).

$$\rho(\vec{r},t) = \frac{N}{(4\pi Dt)^{3/2}} \exp\left(-\frac{\vec{r}^{\,2}}{4Dt}\right) . \tag{4.70}$$

The mean quadratic width of this distribution along each coordinate x_i ($i = 1, 2, 3$) grows linearly with time:

$$\langle x_i^2 \rangle = \frac{1}{N} \int x_i^2 \, \rho(\vec{r},t) \, d^3\vec{r} = 2Dt \, . \tag{4.71}$$

In this equation the integral goes over the full space, that is, from $-\infty$ to $+\infty$ for all three coordinates. The mean quadratic radius of the charge cloud becomes after the time t:

$$\langle r^2 \rangle = \langle x_1^2 \rangle + \langle x_2^2 \rangle + \langle x_3^2 \rangle = 6Dt \, . \tag{4.72}$$

The widths along the three coordinate axes are

$$\sigma_1 = \sigma_2 = \sigma_3 = \sqrt{2Dt} \, . \tag{4.73}$$

Since the solution (4.70) turns into a δ function for $t = 0$,

$$\lim_{t \to 0} \rho(\vec{r},t) = N \, \delta(\vec{r}) \, , \tag{4.74}$$

it follows that a point-like distribution evolves as a result of diffusion into a Gaussian-shaped cloud expanding with the square-root of time. This means that a charge that has been created at the time $t = t'$ at the position $\vec{r}\,'$ has a probability density to be seen at the time $t > t'$ at position $\vec{r}$ given by a Gaussian distribution (fig. 4.5):

$$\rho(\vec{r},t;\vec{r}\,',t') = N \left(\frac{1}{4\pi D(t-t')}\right)^{3/2} \exp\left(-\frac{(\vec{r}-\vec{r}\,')^2}{4D(t-t')}\right) . \tag{4.75}$$

In general charges are generated at different points and different times, as is usually the case for the passage of ionising particles through a detector medium. The distribution of the charge cloud $\rho(\vec{r},t)$ at time t is obtained by integrating over the charge generation as given by the distribution $u(\vec{r}\,',t')$ of the positions and times of generation points:

$$\rho(\vec{r},t) = N \int u(\vec{r}\,',t') \left(\frac{1}{4\pi D(t-t')}\right)^{3/2} \exp\left(-\frac{(\vec{r}-\vec{r}\,')^2}{4D(t-t')}\right) d^3\vec{r}\,' dt' \, . \tag{4.76}$$

Example. In the case of a line-shaped distribution along the trajectory of an ionising particle, u takes the form:

$$u(\vec{r}\,',t') = \frac{1}{t_1' - t_0'} \, \delta\left(\vec{r}\,' - [\vec{r}_0 + \vec{c}\,(t' - t_0')]\right) \qquad (t_0' \le t' \le t_1') \, . \tag{4.77}$$

The vector $\vec{r}\,'(t) = \vec{r}_0 + \vec{c}\,(t' - t'_0)$ represents the linear equation of the particle trajectory passing between t'_0 and t'_1 through the medium with the velocity $\vec{c}$. The particle enters the medium at $\vec{r}\,' = \vec{r}_0$. Choosing the particle direction as the x_3 axis one obtains the one-dimensional distribution

$$u(x'_3, t') = \frac{1}{t'_1 - t'_0}\,\delta\left(x'_3 - [x_{3,0} + c\,(t' - t'_0)]\right) \qquad (t'_0 \le t' \le t'_1)\,. \tag{4.78}$$

We now use this to execute the integral (4.76):

$$\rho(\vec{r}, t) = N \left(\frac{1}{\sqrt{4\pi D(t - t'_0)}}\right)^3 \exp\left(-\frac{x_1^2 + x_2^2}{4D(t - t'_0)}\right) \times \frac{1}{t'_1 - t'_0} \int_{t'_0}^{t'_1} \exp\left(-\frac{(x_3 - [x_{3,0} + c\,(t' - t'_0)])^2}{4D(t - t'_0)}\right) dt'\,. \tag{4.79}$$

In the expression for the width of the charge cloud the time was assumed to be constant, $t' \approx t'_0$, because the duration of the particle passage is short compared to the relevant diffusion and drift times. Then (4.79) yields a two-dimensional diffusion distribution perpendicular to the particle trajectory, described by the coordinates x_1 and x_2, with a relative amplitude depending on x_3 via the integral over dt' (second row in (4.79)). The solution of this integral is

$$\frac{1}{t'_1 - t'_0} \int_{t'_0}^{t'_1} \exp\left(-\frac{(x_3 - [x_{3,0} + c\,(t' - t'_0)])^2}{4D(t - t'_0)}\right) dt' = \frac{\sqrt{\pi D(t - t'_0)}}{c(t'_1 - t'_0)} \left[\operatorname{erf}\left(\frac{(x_3 - x_{3,0})}{\sqrt{4D(t - t'_0)}}\right) - \operatorname{erf}\left(\frac{x_3 - x_{3,0} - c(t'_1 - t'_0)}{\sqrt{4D(t - t'_0)}}\right)\right]\,. \tag{4.80}$$

Here we used the Gaussian error function erf (see e.g. [976]). The expression $c(t'_1 - t'_0)$ gives the length of the line charge created along the particle path. The diffusion leads to an elongated charge distribution whose cross section at fixed x_3 follows a two-dimensional Gaussian distribution. In the case of very long charge line distributions, $x_{3,0} \ll x_3 \ll x_{3,0} + c(t'_1 - t'_0)$, the integral in (4.80) becomes independent of x_3 (because of $\operatorname{erf}(\pm\infty) = \pm 1$) and thus also the diffusion cloud at any time t.

4.4.2 Determination of the diffusion coefficient

For a particle which scatters off the other particles of a medium, the probability density for the path length between two collisions is (see also eq. (3.6))

$$p(r) = \frac{1}{\lambda} \mathrm{e}^{-r/\lambda}\,, \tag{4.81}$$

where $\lambda = \langle r \rangle$ is the mean free path of the particle in the medium as used before in (4.41). In (3.7) we had shown that λ is inversely proportional to the particle density and to the collision cross section.

In the following we consider a particle originating from $\vec{r} = 0$ which diffuses thermally. The diffusion expansion until the kth collision is described by the mean quadratic

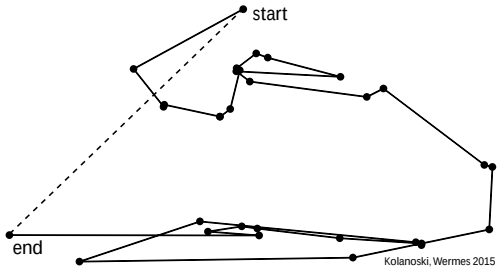


Fig. 4.6 Simulated example for a random walk of a particle which freely moves by a path length following the distribution (4.81) and will then be scattered isotropically. In the limit of a large number of scatters, the length of the connecting line between start and end (dashed) follows a normal distribution.

distance $\langle r^2\rangle_k$ from the origin (see fig. 4.6). The positions of the collisions are given by the vectors $\vec{r}_k$.

– At the first collison:

$$\langle r^2\rangle_1 = \int_0^\infty r^2 p(r)dr = \frac{1}{\lambda}\int_0^\infty r^2 \mathrm{e}^{-\frac{r}{\lambda}}dr = 2\lambda^2\,.$$

– At the second collison:

$$\langle r^2\rangle_2 = \langle(\vec{r}_1+\vec{r}_2)^2\rangle = \langle r_1^2 + r_2^2 + 2r_1r_2\cos\theta\rangle\,.$$

Additionally the right-hand side has to be averaged over the angle θ between $\vec{r}_1$ and $\vec{r}_2$. The assumption of an isotropic direction distribution corresponds to a uniform distribution in $\cos\theta$:

$$\begin{aligned}\langle r^2\rangle_2 &= \frac{1}{2}\int_{-1}^{+1} d\cos\theta\,\langle r_1^2 + r_2^2 + 2r_1r_2\cos\theta\rangle \\ &= \langle r_1^2 + r_2^2\rangle = \langle r_1^2\rangle + \langle r_2^2\rangle = 2\,\langle r^2\rangle_{1,2} = 2\,(2\lambda^2)\,.\end{aligned} \tag{4.82}$$

– At the kth collison:

$$\langle r^2\rangle_k = k\,2\,\lambda^2\,.$$

At the kth collison, on average the time

$$\delta t = \frac{k\,\lambda}{v}$$

has passed. Therefore the time dependence of $\langle r^2\rangle$ follows with (4.71) and (4.72):

$$\frac{\partial\langle r^2\rangle}{\partial t} \approx \frac{\langle r^2\rangle_k}{\delta t} = 2\lambda v = 6D\,. \tag{4.83}$$

After averaging over the velocities and the mean free path λ, which in general depend on v, one obtains for the diffusion coefficient:

$$D = \frac{\langle\lambda v\rangle}{3}\,. \tag{4.84}$$

The averaging is carried out over the distribution function f which is usually determined by numerically solving the Boltzmann transport equation.

As a measure for the spread of the charge cloud during the drift motion the 'characteristic energy' ϵ_k is defined by the ratio of the diffusion coefficient D to the mobility of a charge q (introduced in eq. (4.45)):

$$\epsilon_k = \frac{qD}{\mu} \geq kT\,. \tag{4.85}$$

A small value of ϵ_k is favourable for the localisation of a charge cloud. A lower limit for ϵ_k is stated by the *Nernst–Townsend–Einstein relation*[4] as $\epsilon_k = kT$, which is the mean thermal energy of a particle in equilibrium with a medium at temperature T (k is the Boltzmann constant). The limit arises because a perturbation, for example by an electric field, can only heat up but not cool the particles. For $\epsilon_k \approx kT$ the velocity distribution corresponds approximately to the Maxwell–Boltzmann distribution. Deviations from this 'thermal limit' prevail in particular for the electron drift in gases (see section 4.6.4.3).

4.5 Motion of ions in gases

The velocity distribution of ions in drift gases practically does not change with an applied electric field for field strengths which are usual for ionisation detectors. Hence the energy dependence of cross sections does not play a role and the mobilities (4.45) become independent of the electric field.

The ion motion in gases plays an important role in the development of signals on the anode of a wire chamber if gas amplification takes place near the anode (see section 5.3.2). Relevant parameters for the motion of ions in their own gas are listed in table 4.1. In addition, values of mobilities for some combinations of ionised organic molecules in different drift gases are given in table 4.2.

4.5.1 Example: gas parameters for ideal gases

In the following we will compute the diffusion coefficient, the mobility and the characteristic energy for the molecules of an ideal gas. The velocities follow a Maxwell–Boltzmann distribution:

$$f(\vec{v})\,4\pi\,v^2 dv = 4\pi \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2\,\exp\left(-\frac{mv^2}{2kT}\right)\,dv\,. \tag{4.86}$$

On the left hand side is the function $f(\vec{v})$ as defined in (4.33) integrated over the solid angle. Using the relationships of (4.32) yields the corresponding distribution of the kinetic energy ϵ:

$$F(\epsilon)\,4\pi\,d\epsilon = 2\pi \left(\frac{1}{\pi kT}\right)^{\frac{3}{2}} \sqrt{\epsilon}\,\exp\left(-\frac{\epsilon}{kT}\right)\,d\epsilon\,. \tag{4.87}$$

As compared to (4.34) and (4.35) the angular dependence of $F(\epsilon, \cos\theta, \phi)$ has been integrated out assuming an isotropic distribution. Hence the distributions (4.86) and (4.87) are normalised such that the integral over v or ϵ, respectively, is unity.

In the following we will use some mean values with respect to the distribution (4.86) or (4.87) (obviously both distributions yield the same results):

$$\langle v \rangle = \sqrt{\frac{8kT}{\pi m}} \qquad \text{and} \qquad \left\langle \frac{1}{v} \right\rangle = \sqrt{\frac{2m}{\pi kT}}\,. \tag{4.88}$$

[4] Also called Nernst–Townsend relation or Einstein relation.

Table 4.1 Gas parameters for ions moving in their own gas under normal conditions (from [849]): mean free path, mean (unordered) velocity, diffusion coefficient and mobility.

Gas	λ (cm)	$\langle v \rangle$ (cm/s)	D (cm^2/s)	μ^+ (cm^2/V/s)
H_2	1.8×10^{-5}	2×10^5	0.34	13.0
He	2.8×10^{-5}	1.4×10^5	0.26	10.2
Ar	1.0×10^{-5}	4.4×10^4	0.04	1.7
0_2	1.0×10^{-5}	5.0×10^4	0.06	2.2
H_20	1.0×10^{-5}	7.1×10^4	0.02	0.7

Table 4.2 Mobilities of some ions in gases (from [849]). The listed organic gases are methane (CH_4), isobutane (i-C_4H_{10}) and methylal ($(OCH_3)_2CH_2$), which are often used as 'quenchers' (see section 7.5).

Gas	Ion	μ^+ (cm^2/V/s)
Ar	$(OCH_3)_2CH_2^+$	1.51
i-C_4H_{10}	$(OCH_3)_2CH_2^+$	0.55
$(OCH_3)_2CH_2$	$(OCH_3)_2CH_2^+$	0.26
Ar	i-$C_4H_{10}^+$	1.56
i-C_4H_{10}	i-$C_4H_{10}^+$	0.61
Ar	CH_4^+	1.87
CH_4	CH_4^+	2.26
Ar	CO_2^+	1.72
CO_2	CO_2^+	1.72

In an ideal gas the cross section σ for elastic collisions, and thus also the mean free path λ, are energy independent:

$$\lambda = \frac{1}{n\,\sigma} = \frac{kT}{p\sigma}\,, \qquad \text{with} \quad n = \frac{p}{kT}\,. \tag{4.89}$$

On the right-hand side the particle density $n = dN/dV$ is expressed in terms of pressure p and temperature using the ideal gas equation $pV = NkT$.

With these preparations we can now determine the gas parameters diffusion coefficient D, mobility μ and characteristic energy ϵ_k. With (4.89) the diffusion coefficient can be computed according to (4.84):

$$D = \frac{\langle \lambda v \rangle}{3} = \frac{\lambda \langle v \rangle}{3} = \frac{1}{3\sigma p}\sqrt{\frac{8(kT)^3}{\pi m}} \propto \frac{T^{3/2}}{p}\,. \tag{4.90}$$

Equation (4.44) yields for the mobility (with $d\lambda/dv = 0$ because the scattering is energy independent for ideal gases):

$$\mu = \frac{2}{3}\frac{q}{m}\lambda \left\langle \frac{1}{v} \right\rangle = \frac{q}{3\sigma p}\sqrt{\frac{8kT}{\pi m}} \propto \frac{T^{1/2}}{p}\,. \tag{4.91}$$

Inserting (4.90) and (4.91) into the definition of the characteristic energy in (4.85) yields for ideal gases the Nernst–Townsend–Einstein relation:

$$\epsilon_k = \frac{qD}{\mu} = kT\,. \tag{4.92}$$

Hence for ions in gases (at not too high field strengths) the characteristic energy reaches the lower limit kT as given in (4.85), corresponding to the minimally possible dispersion of the charge cloud during the drift motion. This result is expected since we assumed thermal equilibrium with no external forces.

4.5.2 Blanc rule for gas mixtures

The Blanc rule, see [212], states that the inverse mobility of ions in a gas mixture is given by the sum of the inverse mobilities of the components, μ_i, weighted by the respective relative concentrations:

$$\frac{1}{\mu} = \sum_i \frac{f_i}{\mu_i}\,. \tag{4.93}$$

The rationale for this rule is that the mobilities are inversely proportional to the product of cross section and particle density which have to be added for gas mixtures. A corresponding equation also holds for the diffusion coefficients because they have the same dependence on the cross section and the particle density. Deviations from this rule are observed if charge exchange between the drifting ions and the molecules of a gas component with a different mobility becomes important.

4.6 Motion of electrons in gases

The motion of electrons in gases is much more complicated than that of ions. The reason is that for electrons the cross sections and inelasticities are in most cases strongly energy dependent due to the numerous electronic excitation levels of the gas atoms. The discussion in this section will therefore be more extensive, also accounting for the particular importance for gases as detector media. We follow here the manifold literature on electron motion in gases,[5] for example [535, 550, 753, 752, 805, 866].

In the following we want to find a stationary, position independent solution of the Boltzmann transport equation (4.17) for electron motion in gases under the influence of external fields. As in section 4.2.4 we start in lowest order with an isotropic distribution f_0 and derive the anisotropies from the perturbations of f_0 induced by the electric and magnetic fields. Employing the obtained distributions we will determine the detector-relevant quantities drift velocity, diffusion coefficient and Lorentz angle. As already noted, in practice this determination is carried out by numerical computations, for example using the program Magboltz [196, 195].

4.6.1 Parametrisation of the collision integral

The evaluation of the collision integral requires the energy dependent cross sections σ, or instead the mean free paths λ for a particle density n, and the corresponding inelasticities of the scattering cross sections. In general, the parameters are functions of the energy ϵ:

– $\lambda(\epsilon) = \frac{1}{n\sigma(\epsilon)}$: mean free path,

[5] The essentials of electron motion in gases have already been clarified with the classical work of J. Townsend and others at the beginning of the last century, see for example [943].

– $\tau(\epsilon) = \frac{\lambda(\epsilon)}{v}$: collision time, the mean time between two collions for a particle with energy ϵ,

– $\Lambda(\epsilon) = \frac{\Delta\epsilon}{\epsilon}$: inelasticity, the relative energy loss by a scattering ('cooling').

In elastic scattering processes the elasticity condition connects the initial velocity v with the velocity v' after the scattering and the scattering angle ψ:

$$v' = v\left[1 - \frac{m}{M}(1 - \cos\psi)\right] . \tag{4.94}$$

After averaging over the scattering angle distribution, assumed to be isotropic, the energy loss due to the recoil for $m \ll M$ becomes $\Lambda_{\text{recoil}}(\epsilon) = 2m/M$. Adding the inelastic contributions from j excitation levels with excitation energies ϵ_i and cross sections $\sigma_i(\epsilon)$ $(i = 1, \ldots, j)$ yields the mean inelasticity:

$$\Lambda(\epsilon) = 2\frac{m}{M} + \sum_{i=1}^{j} \frac{\sigma_i(\epsilon)\,\epsilon_i}{\sigma(\epsilon)\,\epsilon} . \tag{4.95}$$

Here $\sigma(\epsilon)$ is the total cross section for scattering of electrons off gas molecules.

4.6.2 Cross sections and inelasticities

In fig. 4.7 the cross sections for some gases, which are often used for detectors, are displayed as a function of the electron energy. The elastic cross section is in each case plotted as 'momentum-transfer (MT) cross section' which is the cross section weighted by the momentum transfer. For all gases the elastic cross section exhibits maxima and minima which arise from quantum-mechanical interferences of the electron wavefunction when scattering on the potential well of the atom or molecule (see e.g. [862]). This quantum-mechanical phenomenon is called the Ramsauer effect or Ramsauer–Townsend effect [807, 395]. The so-called *Ramsauer minimum*, which is particularly distinct for argon and the other heavy noble gases, plays a special role for the energy dependence of the drift velocity. For example, the elastic electron cross section in pure argon in fig. 4.7 shows the Ramsauer minimum at $\epsilon \approx 0.3\,\text{eV}$ and a maximum at $\epsilon \approx 10\,\text{eV}$.

Argon has its first excitation level at 11.6 eV, below which the cross section is purely elastic, with a small damping by the recoil energy transferred onto the atom according to (4.95). The polyatomic molecules CH_4, CO_2 and CF_4 in fig. 4.7 have vibration and rotation levels already in the eV region and below. These gases are employed for the 'cooling' of the random electron velocities. For example, the energy distributions in argon in fig. 4.8(a) are clearly shifted downwards by adding 10% methane (CH_4), as can be seen in fig. 4.8(b) (see the discussion of the figure in the following section).

4.6.3 Approximate solutions of the Boltzmann transport equation

In order to determine the parameters for drift and diffusion of electrons in gases we use in the following the relaxation approximation for the velocity distribution as given by (4.29). In contrast to ions which, at not too high electric fields, exhibit thermal velocities following a Maxwell–Boltzmann distribution (see section 4.5), electrons are in general heated up by an electric field. Therefore, the isotropic steady-state part f_0, to which the field dependence is added as a perturbation in the approximate solution (4.29), has first to be determined. In the literature cited at the beginning of this

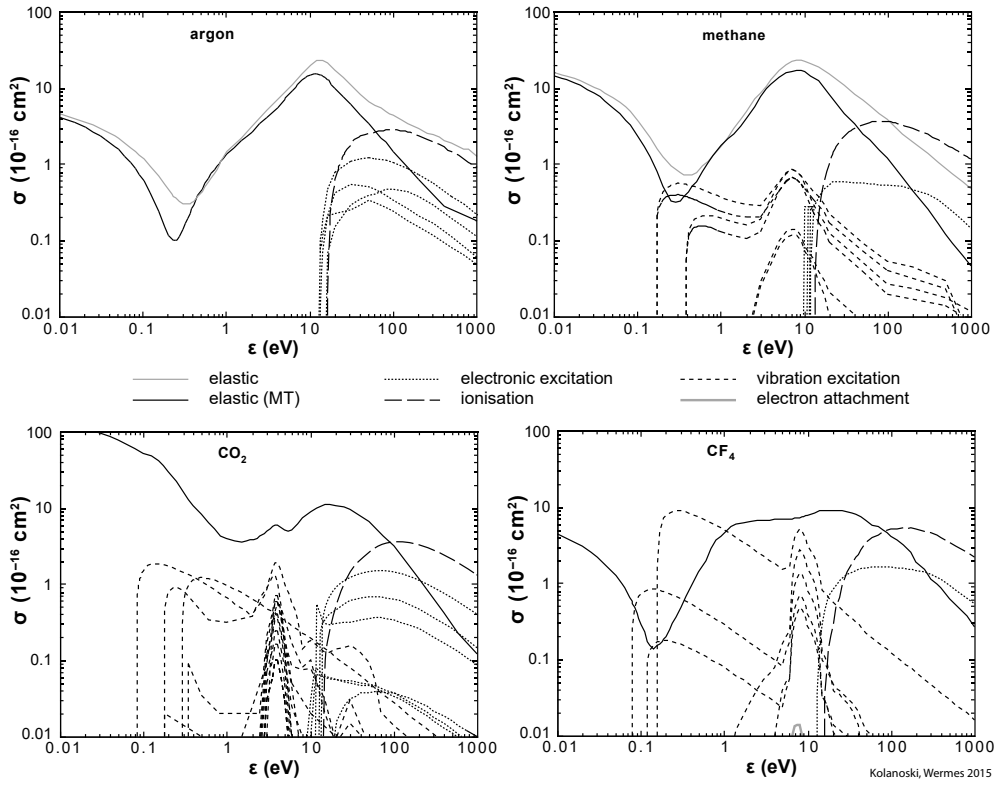


Fig. 4.7 Cross sections for elastic and inelastic electron scattering in different gases (MT denotes the 'momentum transfer cross section' see text). The plots have been created interactively on the web page [195] (the data were retrieved in 2013). The legend explains the line patterns for the different types of cross sections.

section (page 108) one usually starts with the Boltzmann equation (4.4) for a stationary distribution ($\partial f/\partial t = 0$) in a phase space continuum without spatial structure ($\vec{\nabla}_{\vec{r}} f = 0$), which means there is no diffusion. The thus reduced equation expresses that the force term, which causes a movement in velocity space, has to be balanced by the collision integral:

$$\frac{d\vec{v}}{dt}\,\vec{\nabla}_{\vec{v}} f = \left.\frac{\partial f}{\partial t}\right|_{coll} . \tag{4.96}$$

The solutions offered in the literature exploit the steady-state condition that the kinetic energy an electron gains by the work of the field over the mean free path λ, that is, $qE\lambda$, has on average to be compensated by the mean energy loss in a collision, $\Delta\epsilon = \Lambda\,\epsilon$. In the following we just quote the solution without a derivation which would be beyond the scope of this book. For the derivation and more details we again refer to the literature cited on page 108.

4.6.3.1 Isotropic part of the distribution

A solution of (4.96) for the isotropic part of the energy distribution of electrons, corresponding to the distribution f_0 in section 4.2.4, is given by:[6]

$$f_0(\epsilon) = C \, \exp\left[- \int_0^\epsilon \frac{3\epsilon' \Lambda(\epsilon') w_B}{\left(qE\lambda(\epsilon')\right)^2 + 3\epsilon' \Lambda(\epsilon') w_B \, kT} d\epsilon' \right] . \tag{4.97}$$

Before we discuss the different terms in this equation let us note that the function $f_0(\epsilon)$, corresponding to $f(\epsilon, \cos\theta, \phi)$ of (4.33) in the isotropic case, is actually a velocity distribution with the velocity expressed by the kinetic energy. The normalisation of $f_0(\epsilon)$ is given by (4.35) (the factor 4π is due to the integration over the full solid angle):

$$4\pi \int_0^\infty f_0(\epsilon) \sqrt{\frac{2\epsilon}{m^3}} \, d\epsilon = 1 \, . \tag{4.98}$$

This normalisation fixes also the constant C in (4.97). Instead of f_0 often the energy distribution F_0, analogously defined according to (4.34), is used leading to the energy distribution corresponding to (4.97):

$$F_0(\epsilon) = C' \, \sqrt{\epsilon} \, \exp\left[- \int_0^\epsilon \frac{3\epsilon' \Lambda(\epsilon') w_B}{\left(qE\lambda(\epsilon')\right)^2 + 3\epsilon' \Lambda(\epsilon') w_B \, kT} d\epsilon' \right] . \tag{4.99}$$

The normalisation of $F_0(\epsilon)$, which also fixes the constant C', is given by

$$4\pi \int_0^\infty F(\epsilon) \, d\epsilon = 1 \, . \tag{4.100}$$

In (4.97) and (4.99) the quantity w_B, which accounts for the effect of the magnetic field (oriented as in (4.6)), is defined as

$$w_B = w_B(\vec{B}, \tau) = \frac{1 + \omega^2 \, \tau^2}{1 + \omega_3^2 \, \tau^2} \, , \qquad \text{with} \quad \omega^2 = \omega_2^2 + \omega_3^2 = \left(\frac{q}{m} B \right)^2 . \tag{4.101}$$

The variable $\tau = \lambda / v$ is the collision time. The numerator of the integrand in (4.97) and (4.99) contains the energy loss per collison and the first term in the denominator contains the work of the field over the mean free path, $qE\lambda$. The second term in the denominator has been added by hand to account for the thermal motion of the gas molecules (see the explanation in [871]).

If the term $\left(eE\lambda(\epsilon')\right)^2$ in the denominator of the integrand, which describes the energy gain from the electric field, becomes small as compared to the second term, which contains the inelastic damping, then the integral becomes ϵ / kT and $F_0(\epsilon)$ turns into the Maxwell–Boltzmann distribution:

$$F_0(\epsilon) \, 4\pi \, d\epsilon = 2\pi \left(\frac{1}{\pi kT} \right)^{\frac{3}{2}} \sqrt{\epsilon} \, \exp\left(- \frac{\epsilon}{kT} \right) \, d\epsilon \, . \tag{4.102}$$

[6] The formulae for a general orientation of the magnetic field relative to the electric field are taken from [866]. In the case that the magnetic field is orthogonal to the electric field the solution coincides with corresponding ones in the literature, for example in [805].

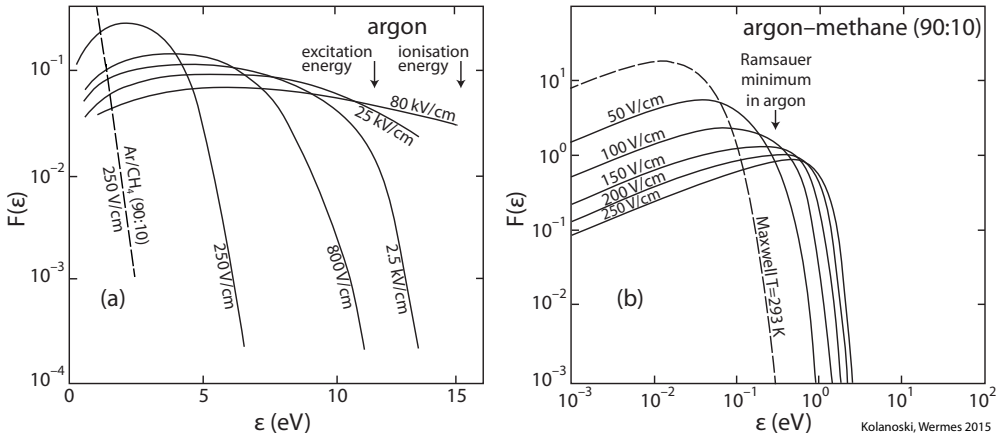


Fig. 4.8 Computed energy distributions of electrons in argon and in a mixture of argon (90%) and methane (10%) for different field strengths at normal pressure and temperature (101 325 Pa, 293.15 K = 20 °C). The energy distributions $F(\epsilon)$ are defined as in (4.34). The curves originate from plots in [752] (argon) and [864] (argon–methane). (a) Pure argon: the energy distributions shift with a higher electric field towards higher energies. The curve for $E = 2.5$ kV/cm clearly shows the downward bending when reaching the lowest excitation energy (11.6 eV). At much higher field strengths the distribution continues up to the ionisation energy (15.8 eV) where secondary ionisation sets in. (b) Argon–methane (90:10): the curve with the field strength 250 V/cm can be compared to the corresponding one for pure argon (in (a) the tail of the distribution is shown as a dashed line). The maximum shifts from about 2.5 eV in pure argon to about 0.6 eV in the mixture with methane. The methane admixture damps the electron energies due to inelastic scattering. For comparison also the Maxwell–Boltzmann distribution for $T = 293.15\,\mathrm{K} = 20\,^\circ\mathrm{C}$ ($kT \approx 0.025$ eV) is shown.

With increasing $(eE\lambda(\epsilon'))^2$ the electron distribution will be 'heated' as demonstrated in fig. 4.8 using the example of argon (a) and of an argon–methane mixture (b). In argon the damping is very low, permitting long tails of the distribution towards high energies, while an admixture of a gas with polyatomic molecules damps the energies. See also the related discussion in section 4.6.2.

4.6.3.2 Distribution with anisotropies due to external fields

In the relaxation approximation the complete distribution function $f(\vec{v})$ is set up according to (4.29) as

$$f(\epsilon, \cos\theta, \phi) = f_0(\epsilon) - \frac{qE}{m}\tau\frac{1}{1+\omega^2\tau^2}\,\vec{v}\begin{pmatrix} -\omega_2\tau \\ \omega_2\omega_3\tau^2 \\ \omega_3^2\tau^2 \end{pmatrix} m\,\frac{\partial f_0}{\partial\epsilon}\,. \tag{4.103}$$

This function contains the anisotropies due to the external fields,[7] which provide absolute value and direction of the drift velocity according to the equations in section 4.3.

[7] In the literature the distribution $f(\epsilon, \cos\theta, \phi)$ is often expanded into a series of spherical harmonics $Y_{lm}(\theta, \phi)$, see for example [805]. The representation (4.103) corresponds to an expansion up to $l = 1$. This becomes obvious when in (4.103) the components of $\vec{v}$ are expressed in spherical coordinates.

In the following we will use the distribution function (4.103) to derive the parameters for electrons moving in a gas.

4.6.4 Determination of the macroscopic gas parameters

4.6.4.1 Dependences of the drift velocity

In the following we want to analyse the dependence of the drift velocity on the electron cross sections in the drift gas which enter through the mean free path $\lambda(\epsilon)$ and the inelasticity $\Lambda(\epsilon)$ into the formula for the drift velocity (4.44). For this purpose we evaluate the drift velocity with the distribution as given in (4.97). In order to obtain an easily understandable relation, we assume $\omega_B = 1$ ($\vec{B} = 0$) and that the kT term in the denominator can be neglected. Then the derivative of f_0 in (4.40) becomes

$$\frac{\partial f_0}{\partial \epsilon} = -f_0 \frac{3\,\epsilon\,\Lambda}{(qE\lambda)^2}\,. \tag{4.104}$$

If we additionally assume that the change of λ with energy is small, hence $\partial\lambda/\partial\epsilon \approx 0$, then follows from (4.40) with (4.104):

$$v_D = 4\pi \frac{1}{qE}\sqrt{\frac{2}{m}}\int_0^\infty \frac{\Lambda}{\lambda}\,\epsilon^{\frac{3}{2}}\,f_0(\epsilon)\sqrt{\frac{2\,\epsilon}{m^3}}\,d\epsilon = \sqrt{\frac{2}{m}}\,\frac{1}{qE}\left\langle \frac{\Lambda}{\lambda}\,\epsilon^{\frac{3}{2}}\right\rangle . \tag{4.105}$$

The averaging on the right-hand side has to be done with respect to the f_0 distribution using for the integration (4.33). From (4.105) one obtains

$$qEv_D = \left\langle \Lambda\epsilon\frac{v}{\lambda}\right\rangle . \tag{4.106}$$

The right-hand side is the mean energy loss per time because $\Delta\epsilon = \Lambda\epsilon$ is the energy loss in one collision and v/λ is the number of collisions per time. The left side is the mean work per time done by the field. Therefore (4.106) states that in the stationary case the work delivered by the field will be compensated by the energy loss due to the collisions. Hence we recover the assumptions made for solving the Boltzmann equation (4.96).

4.6.4.2 A simplified model for the gas parameters

For the illustration of the influence of the parameters E, λ, Λ—that is, of the electric field and the cross sections—on the gas parameters v_D, D, ϵ_k we follow a discussion in [753]. There the (unordered) velocity v is assumed to be constant or to have a narrow distribution. Then the two equations (4.106) and (4.44) can be equated without averaging over v:

$$v_D = \frac{1}{qE}\Lambda\epsilon\frac{v}{\lambda} = \frac{2}{3}\frac{qE}{m}\frac{\lambda}{v}\,. \tag{4.107}$$

Using in addition $\epsilon = mv^2/2$ this yields v, v_D, D, ϵ_k as functions of E, λ, Λ:

$$v = \sqrt{\frac{2}{\sqrt{3\Lambda}}\frac{qE}{m}\lambda}\,, \tag{4.108}$$

$$\epsilon = \frac{qE\lambda}{\sqrt{3\Lambda}}\,, \tag{4.109}$$

$$v_D = \sqrt{\frac{2}{3}\sqrt{\frac{\Lambda}{3}}\frac{qE}{m}\lambda}\,, \tag{4.110}$$

$$D = \frac{\lambda v}{3} = \frac{1}{3}\sqrt{\frac{2}{\sqrt{3\Lambda}}\frac{qE}{m}\lambda^3}\,, \tag{4.111}$$

$$\epsilon_k = \frac{qD}{\mu} = \frac{qED}{v_D} = \frac{qE\lambda}{\sqrt{3\Lambda}} = \epsilon\,. \tag{4.112}$$

Hence in a narrow distribution approximation $\epsilon_k = \epsilon$, revealing the meaning of the characteristic energy. The typical dependences of the motion parameters on E, λ, Λ are as follows:

- v, $v_D \propto \sqrt{\lambda}$: Both velocities increase with an increase of the mean free path because the acceleration phase between the collision increases. The acceleration heats up the unordered motion.
- $D \propto \lambda^{3/2}$: The diffusion increases with the mean free path.
- v, D, ϵ_k decrease with increasing Λ (the inelasticity causes a 'cooling' of the electrons).
- v_D increases with increasing Λ (v small $\Rightarrow$ $\tau = \lambda/v$ large: the acceleration phase becomes longer).

The gas parameters v_D and ϵ_k are functions of $E\lambda$. Because of the relation

$$E\lambda = \frac{E}{n\sigma} \propto \frac{E}{p}\frac{1}{\sigma}\,, \tag{4.113}$$

these quanties only depend on the so-defined 'reduced field strength' E/p or E/n. In the older literature these quantities were often given in the following units:

- E/p: 1 $\mathrm{V\,cm^{-1}\,Torr^{-1}} \mathrel{\widehat{=}} 760$ V/cm at NTP,
- E/n: 1 Townsend = 1 Td = $10^{-17}\ \mathrm{V\,cm^2} \mathrel{\widehat{\approx}} 250$ V/cm at NTP.

The abbreviation NTP denotes normal conditions for temperature and pressure as defined in section 2.5.2 on page 20 (1 atm = 101 325 Pa = 760 Torr (mmHg), 293.15 K = 20 °C).

Example. As an example we consider the energy distributions of electrons in an argon–methane mixture (90:10), as depicted in fig. 4.8(b) for different field strengths. The distributions are shifted to smaller energies as compared to the distributions in pure argon since the inelastic cross section of methane, which sets in at about 0.1 eV, damps the higher energies. Therefore, up to about 200 V/cm the maxima of the distributions lie below the Ramsauer minima of both gas components. Using our rough approximation of a narrow energy distribution we assume that the energy lies in the declining region of the cross section below the Ramsauer minimum. In this region we approximate the cross section (dominated by argon, see fig. 4.7) by a power law in energy and a constant inelasticity:

$$\sigma(\epsilon) = \sigma_0\,(\epsilon/\mathrm{eV})^{-2} = 6.5 \times 10^{-19}\,\mathrm{cm^2}\,(\epsilon/\mathrm{eV})^{-2}, \tag{4.114}$$

$$\Lambda(\epsilon) \approx 0.25\,. \tag{4.115}$$

The choice for the inelasticity Λ is somewhat arbitrary because below the Ramsauer minimum Λ is very small, as for pure argon. In our simple model, however, some

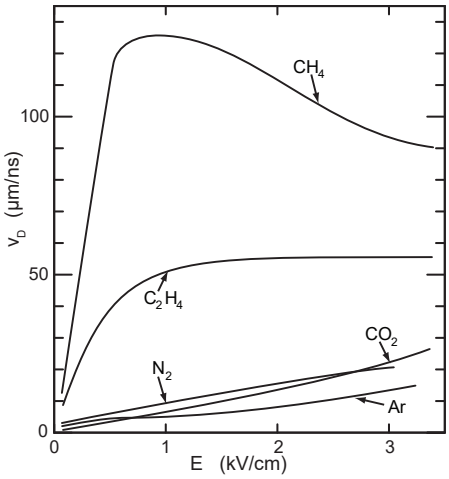


Fig. 4.9 Behaviour of the electron drift velocity as a function of the field strength for different gases (from [849]). The saturation of the drift velocity, possibly followed by a decrease at high fields, is characteristic for hydrocarbons, see the corresponding discussion in the text.

inelasticity is required because the characteristic energy, which is taken as the mean value of the narrow distribution, has to be moved from higher to lower energies by these inelasticities.

With these values for σ and Λ and under normal conditions an energy of, for instance, $\epsilon = 0.1\,\text{eV}$, hence a value below the Ramsauer minimum, is obtained applying a field strength $E = 140\,\text{V/cm}$. This field and the resulting drift velocity $v_D = 55\,\mu\text{m/ns}$ are typical for the operation of drift chambers (section 7.10). Under these conditions the ions' drift velocity is slower by a factor of 1000 or more. For example, the drift velocity of the Ar^+ ions at the quoted field strength is $v_D = 2.4\,\mu\text{m/}\mu\text{s}$.

4.6.4.3 Drift velocities in various gases and gas mixtures

For complex molecules, like the hydrocarbons methane, ethane and propane or as well carbon dioxide, which are often used in drift chambers, the energy dependent inelastic excitations are important (examples in fig. 4.7). In mixtures of noble gases (mostly argon) with such gases the energy distribution of the electrons is shifted towards the Ramsauer minimum (fig. 4.8), thus into the region of strongly varying cross sections.

In drift chambers (section 7.10) the admixture of hydrocarbons is employed for adjustment of the 'working point' which besides the setting of drift field also includes the chosen gas amplification, see section 7.4. Gases that reach a maximum of the drift velocity at drift fields typical for wire chamber operation (some kV/cm) are called 'saturating' gases (fig. 4.9). The gas saturation, in particular with a flat maximum, can be advantageously exploited in order to obtain an approximately constant drift velocity even for inhomogeneous fields. The adjustment options for argon–methane and argon–ethane are shown in fig. 4.10. Depending on the application other criteria can also be essential for the choice of the mixture, for example little diffusion, stable gas amplification, only minor 'ageing' of the detector and others (see section 7.5). For argon–ethane a typical mixing ratio is 50:50 which requires a drift field of somewhat less than 1 kV/cm in order to operate in the regime of saturated drift velocities. To reach this regime with argon–methane in the ratio 90:10 only about 120 V/cm is required, which makes this gas attractive for long drift paths with a very high total voltage (see e.g. section 7.10.10). A disadvantage of argon–methane (90:10) is the larger diffusion as compared to argon–ethane (50:50).

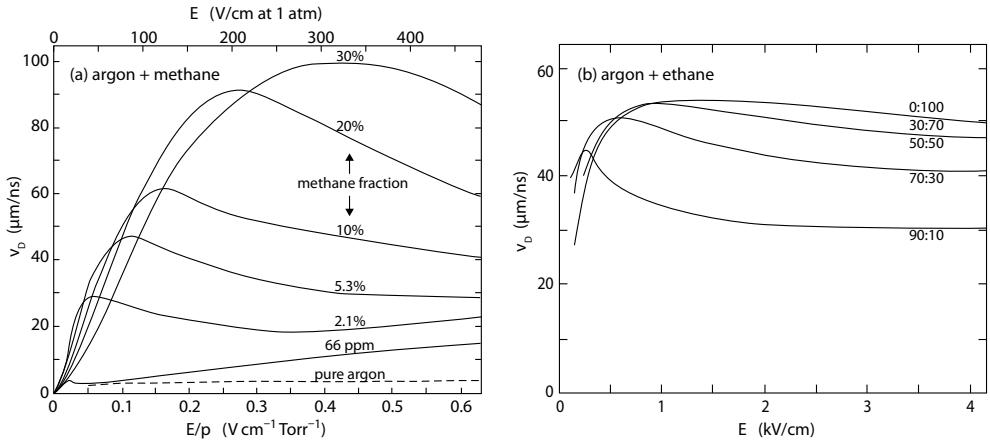


Fig. 4.10 Measured drift velocities of electrons in different gases and gas mixtures as a function of the electric field. The measurement curves have been taken from [433] and [570], respectively. (a) Argon–methane: the plot shows that already a very small admixture of methane causes strong changes of the drift velocities. Varying the mixing ratio the maxima can be adjusted in a wide range of (reduced) field values. (b) Argon–ethane: for these gas mixtures the drift velocities at the maxima are more similar than for the methane admixtures.

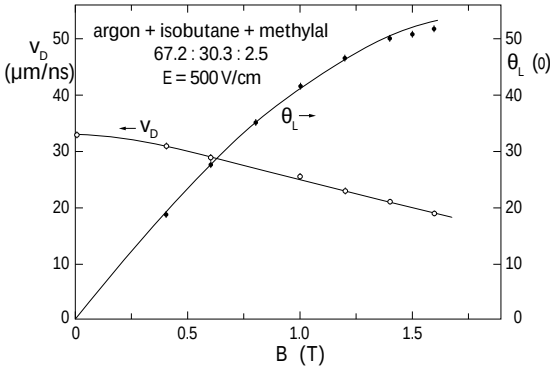


Fig. 4.11 Drift velocity (left axis, open points) and Lorentz angle (right axis, full points) for electrons in the specified gas mixture as a function of the applied magnetic field at fixed drift field [249]. The curves have been calculated using a simple model as described in the text.

Because of the strong energy dependence of the cross sections and inelasticities there exists no simple formula for drift velocities of electrons in gas mixtures that would correspond to the Blanc rule for ions given in (4.93).

4.6.4.4 Lorentz angle

Figure 4.11 shows the drift velocities and Lorentz angles (section 4.3.2.1) as a function of the applied magnetic field (perpendicular to the E-field). The measurements are compared with curves calculated using the formulae (4.58) and (4.59), which are based on a simplified model assuming that the collision times with and without B-field are equal, hence $\tau(E, B) = \tau(E, 0)$. The calculation matches very well the measurements for the plotted range of field strengths; at higher fields larger deviations arise [249].

Arrangements with crossed E- and B-fields are often employed in storage ring experiments with cylindrical drift chambers in solenoidal fields (see section 7.10, fig. 7.38).

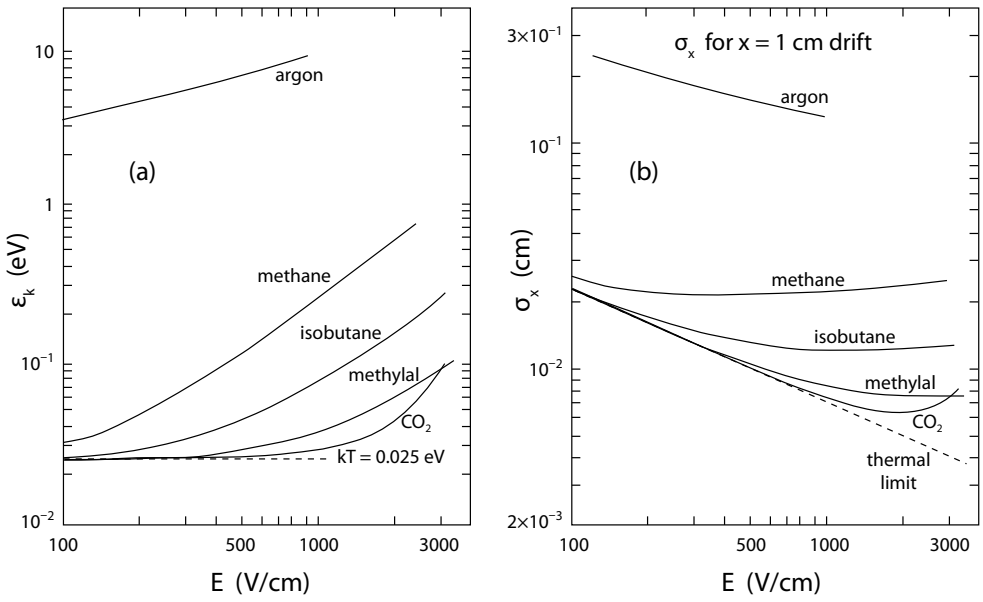


Fig. 4.12 (a) Characteristic energy and (b) width of a diffusion cloud after 1 cm drift distance (using eq. (4.117)) plotted against the electric field strength for different gases (adapted from [871], with kind permission of Elsevier).

4.6.4.5 Diffusion

In isotropic media the diffusion coefficient is determined using (4.84):

$$D = \frac{\langle \lambda v \rangle}{3} = \int \frac{\lambda v}{3} f(v, \cos\theta, \phi)\, v^2 dv d\cos\theta\, d\phi = 4\pi \int \frac{\lambda v}{3} f_0(v)\, v^2\, dv\,. \tag{4.116}$$

The right-hand side is only valid if λv has no directional dependence, see the related discussion below.

Employing (4.73) and (4.85), a constant drift velocity v_D results in a width of the charge cloud of electrons after traversing the distance x:

$$\sigma_x = \sqrt{2Dt} = \sqrt{\frac{2Dx}{v_D}} = \sqrt{\frac{2\epsilon_k\, x}{q\, E}}\,. \tag{4.117}$$

The diffusion is minimal in the thermal limit (4.92), corresponding to the characteristic energy $\epsilon_k = kT$, yielding

$$\sigma_{x,th} = \sqrt{\frac{2kT\, x}{q\, E}}\,. \tag{4.118}$$

Examples for the field dependence of the characteristic energy and the diffusion are shown in fig. 4.12. Minimal diffusion are exhibited by carbon dioxide (CO_2) and methylal ($(OCH_3)_2CH_2$), which stay close to the thermal limit up to relatively high field strengths. Also the hydrocarbons methane and isobutane show relatively small diffusion because the random motions are well damped. The strongest diffusion in this figure is shown by pure argon because of the absence of damping by vibrational and rotational excitations of the gas molecules (see fig. 4.7). For all examples in fig. 4.12

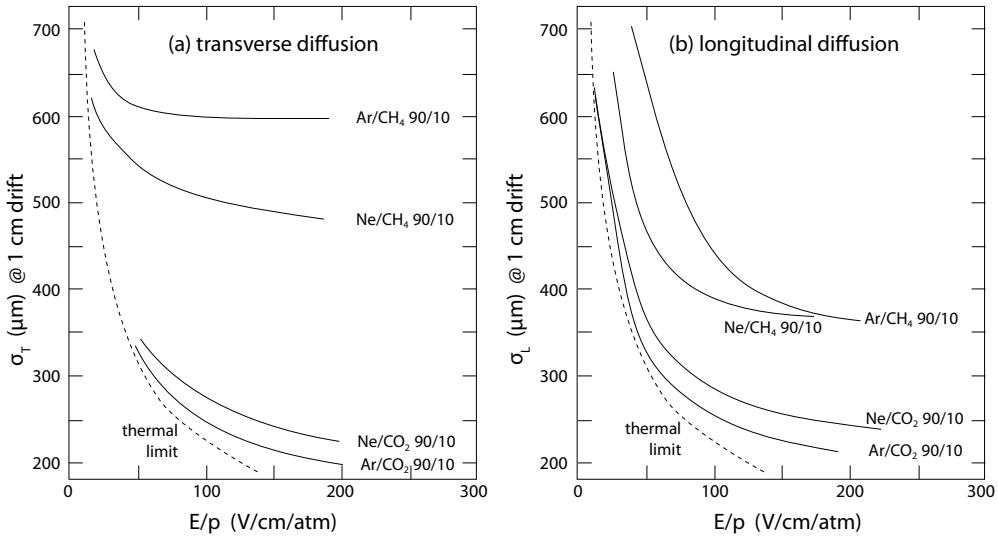


Fig. 4.13 Transverse and longitudinal diffusion widths (for 1 cm drift distance) as a function of the drift field for different gases without magnetic field (adapted from [254]).

the diffusion initially decreases with the field followed by a flattening and potentially an increase again. According to (4.117) an increase comes about if the characteristic energy increases more strongly than proportionally to E. This behaviour can be observed in fig. 4.12 for methane and isobutane and at very high fields also for CO_2.

In general, the velocity distributions and the mean free paths are direction dependent, which has to be taken into account for the calculation of the diffusion coefficients according to (4.116). Without magnetic field an asymmetry relative to the electric field direction arises. Then in general, transversal and longitudinal diffusion with coefficients D_T and D_L, respectively, have to be distinguished. Examples are shown in fig. 4.13. Superimposed magnetic fields have a stronger effect on the diffusion, such that the diffusion coefficient has in general to be written as a tensor (see e.g. [217]).

An important effect in the case of parallel E- and B-fields is the reduction of the transverse diffusion, which is given by

$$D_T = \frac{D}{1+\omega^2\tau^2}\,, \tag{4.119}$$

where τ is the mean collision time.[8] The magnetic field causes a winding of the transversely moving electrons around the electric field. This field configuration is applied for drift chambers of the type 'time projection chamber' (TPC) with very long drift paths (see section 7.10).

4.7 Charge carrier transport in semiconductors

What has been discussed for gases in the previous sections of this chapter also applies, very generally speaking, to semiconductors. However, lattice phenomena must be taken

[8] In this form, formula (4.119) only holds for small magnetic fields, up to about 0.5 T. For larger fields $1+\omega^2\tau^2$ approaches again a linear behaviour as a function of τ^2, however, with a smaller slope (see e.g. [217]).

into account. A specific account on the motion of charge carriers in semiconductors is given in the following sections.

In a semiconductor the charge carriers are electrons and holes (see chapter 8). In pure semiconductor crystals the 'intrinsic' (without doping) electron and hole densities are equal (in silicon at room temperature about $10^{10}\,\mathrm{cm}^{-3}$). In general, doping leads to unequal number densities of the charge carriers, called majority and minority charge carriers, respectively. In n-doped semiconductors electrons are the majority charge carriers and in p-doped semiconductors these are the holes. The majority charge carriers are more abundant by many orders of magnitude than minority charge carriers, in silicon about 10^{10} times.

The charge carrier motion in semiconductors follows the Boltzmann transport equation (4.4) which describes the diffusion and drift motions, as sketched in fig. 4.1. Although to solve the Boltzmann transport equation in semiconductor crystals, lattice phenomena have to be taken into account, which in general requires the use of quantum mechanical methods (see e.g. [881, 402]), the charge movements can largely be described using similar models and parameters as for the movement in gases discussed in sections 4.5 and 4.6. The reason mainly is that electrons and holes in a semiconductor can be treated as if they could move freely. This becomes possible by employing the concept of an 'effective mass' for the charge carriers which accounts for the binding inside the lattice (see section 8.2.3).

Solving the transport equation for semiconductors, an essential result is, as discussed before for gases, that time-independent electric fields cause stationary drift motions of the charge carriers. The drift motion results from an equilibrium between electric acceleration and friction by collision processes; in the case of semiconductors these are collisions with the lattice.

The unordered thermal motion causes diffusion of the charge carriers into regions of lower concentrations. For instance, concentration differences occur in semiconductors—as in gases or other ionisable media—along the ionisation trails of a charged particle or around the local ionisation by an absorbed photon. In semiconductors charge carriers also diffuse across boundary surfaces, for example between p- and n-conducting substrates, which is in fact the basic functional principle of semiconductor electronics (see section 8.3).

4.7.1 Drift of electrons and holes

In a semiconductor the acceleration of electrons and holes by an electric field $\vec{E}$ is compensated by scattering off lattice phonons and crystal defects. The motion is generally described by the Boltzmann transport equation (4.4) as in the case of gases. We also mention here an often used ansatz using the Drude model [366] developed already around 1900. According to this model the equation of motion for the average electron movement, that is, the drift velocity $\vec{v}_D$, reads

$$m^* \left(\dot{\vec{v}}_D + \frac{\vec{v}_D}{\tau} \right) = q\vec{E}\,, \tag{4.120}$$

where $q = \pm e$ and m^* are the charge and the effective mass, respectively, of the charge carriers in the semiconductor (see eq. (8.5) on page 266). The various effects of scattering by the lattice are combined into one parameter, the relaxation time τ. The time τ, under the assumptions given in section 4.3.1 also called collision time, is the mean time to the next momentum change, that is, until the next scattering

process that changes the direction of the particle. For silicon the relaxation time is in the range of picoseconds and is strongly temperature dependent. The Drude ansatz corresponds to the relaxation approximation of the Boltzmann transport equation in section 4.2.3.2 applied to average motions.

In the stationary case of constant drift velocity ($\dot{\vec{v}}_D = 0$) the Drude ansatz leads to

$$\vec{v}_D = \frac{q\,\tau}{m^*}\vec{E} = \mu\,\vec{E}\,, \tag{4.121}$$

where the mobilities of the electrons and holes are given by

$$\mu_{e,h} = \frac{q\,\tau}{m^*_{e,h}}\,. \tag{4.122}$$

Compared to the solutions of the Boltzmann transport equation, as discussed in section 4.3, the quantity τ in (4.121) and (4.122) corresponds to an energy weighted average, as defined in (4.52).

At fixed temperature and at small electric fields the mobility μ is roughly constant, similar to the behaviour of ions in gases (section 4.5). But as a function of temperature the mobility changes strongly because of the temperature dependence of the various collision processes (mostly elastic) which determine the collision time τ. In semiconductors we particularly distinguish between the scattering off lattice phonons and off lattice defects. Since the wavelengths of the thermal charge carriers are larger than the lattice spacing they can only interact with long wavelength phonons, called acoustic phonons, corresponding to coherent lattice oscillations (sound waves in solids) [929]. The lattice oscillations impede the electron motion increasingly with higher lattice temperature. For example, the dependence of μ for *acoustic phonon scattering* (APS) on the temperature for electrons in the conduction band is [152, 929]:

$$\mu_{\mathrm{APS}} \propto \frac{1}{(m^*)^{5/2}\,T^{3/2}}\,, \tag{4.123}$$

where m^* here is the effective electron mass at the conduction band edge (for the definition of m^* see (8.3) in section 8.2.3). An analogous relation holds for holes in the valence band. For (Coulomb) scattering off ionised impurities, called *ionised impurity scattering* (IIS), the temperature dependence is [319, 929]:

$$\mu_{\mathrm{IIS}} \propto \frac{T^{3/2}}{n_{\mathrm{II}}(T)\,(m^*)^{1/2}}\,, \tag{4.124}$$

where n_{II} is the impurity density. The mobility increases with temperature since at higher thermal velocities the charge carriers are less deflected from the original direction due to the $1/p^2$-dependence of the Rutherford scattering formula (eq. (3.98)). Note that also the density of ionised impurities increases with the temperature.

In addition scattering off neutral, non-ionised impurities, called *neutral impurity scattering* (NIS), is possible. Although no Coulomb scattering from those impurities takes place they change the lattice locally and thus affect the carrier mobility. With higher temperature they can become ionised, that is, their density n_{NI} decreases at the expense of an increase in n_{II}. A (weak) temperature dependence thus only enters indirectly through the density change [379, 880]:

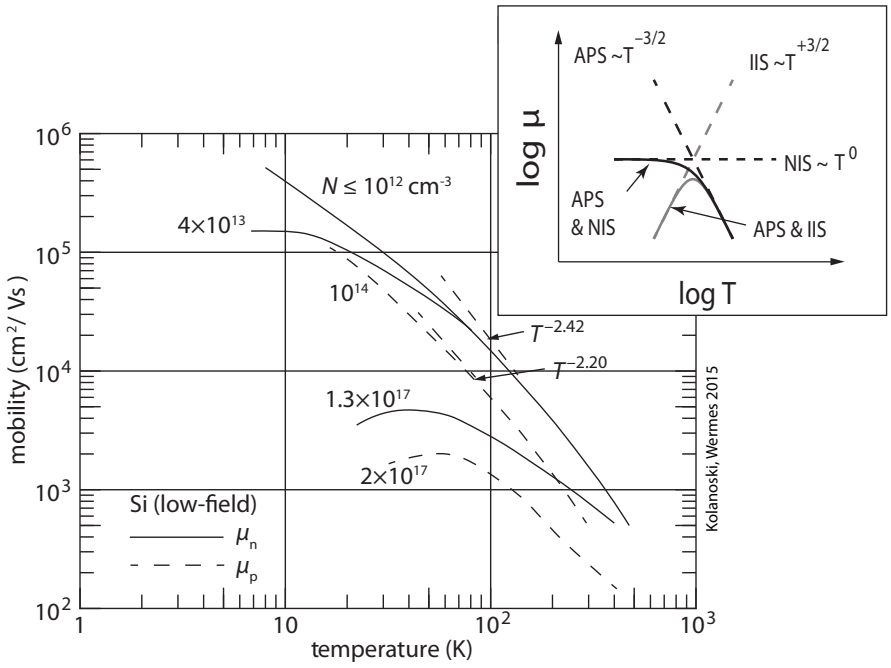


Fig. 4.14 Mobilities of electrons in silicon as a function of temperature for different doping densities N. The insert shows (on a double logarithmic scale) schematically the temperature dependence which is expected from the scattering of the charge carriers off the lattice itself and impurities therein as described in the text (see also [567]). Data are taken from [565], see also [929]. The abbreviations are APS = 'acoustic phonon scattering', IIS = 'ionisation impurity scattering', NIS = 'neutral impurity scattering', and their respective sums APS & NIS and APS & IIS.

$$\mu_{\mathrm{NIS}} \propto \frac{(m^*)^{1/2}\, T^0}{n_{\mathrm{NI}}(T)}\,. \tag{4.125}$$

Other scattering processes as, for example, the scattering of charge carriers from one band minimum to another through interaction with more energetic (optical) phonons, is less relevant, in particular in silicon and germanium.

Therefore, the overall temperature dependence of the mobility in semiconductors follows qualitatively the curve in the insert of fig. 4.14. In the case of different contributing scattering processes i the total mobility is obtained by adding the individual inverse mobilities ('Matthiessen's rule'):

$$1/\mu = \sum_i 1/\mu_i\,. \tag{4.126}$$

The reasoning for this rule is similar to that of the 'Blanc rule' for ions in gases in (4.93). Mobilities in silicon [565], measured for different doping densities, are displayed in fig. 4.14. The strong dependence on the doping arises because the doping atoms act as scattering centres.

At a given temperature and for small field strengths (in Si for $E \ll 10$ kV/cm), the drift velocity v_D is proportional to the electric field (see (4.121)) which means

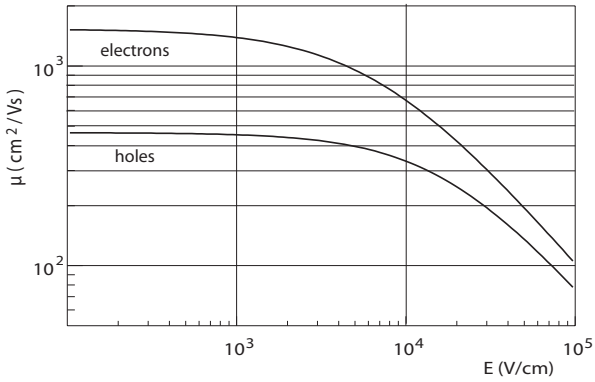


Fig. 4.15 Mobilities of electrons and holes in silicon at 300 K as a function of the electric field strength according to (4.127) with parameters from [837].

according to (4.121) that the mobility is constant. At high field strengths this is no longer true since the mobility becomes field dependent as shown in fig. 4.15 (see for example [929]).

At zero and approximately at low electric fields the system of charge carriers and the lattice (phonons) is in thermal equilibrium; the emission and absorption of phonons by charge carriers leads to (approximately) zero net energy exchange between charge carriers and lattice [929]. The charge carrier distributions at thermal equilibrium follow a Fermi–Dirac or approximately a Maxwell–Boltzmann distribution. With increasing field strength, a stationary equilibrium is reached between the energy spent for the acceleration of the charges and the net energy transferred to the lattice by phonon scattering (phonon absorption and emission), which damps the (unordered) charge velocities and eventually leads to a constant drift velocity, see the corresponding equation (4.106) for electrons in gases. Up to moderate electric fields the scattering processes dominantly involve the emission of *acoustic phonons* (see also page 120). At higher E-fields higher modes of lattice oscillations are increasingly excited, a scattering domain governed by *optical phonons*, i.e. with oscillation frequencies in the optical range. In this regime damping becomes more effective and the mobility drops with increasing field strength, such that the drift velocity v_D becomes more and more field independent approaching a saturation value. At still higher fields the charge carriers can ionise the atoms, generating avalanches of secondary charge carriers.

A simple empirical ansatz [281] yields a good description of the drift velocity over a wide range of field strengths, see fig. 4.15:

$$v_D = \mu(E)\, E = \frac{\mu_0 E}{\left[1 + \left(\frac{\mu_0 E}{v_{\mathrm{sat}}}\right)^{\beta}\right]^{1/\beta}} \,. \tag{4.127}$$

In this equation μ_0 is the *low field mobility* and v_{sat} is the saturation drift velocity, which is reached at high fields in silicon and germanium (fig. 4.15). The empirically determined exponent β, which is not necessarily equal for electrons and holes, typically lies between 1 and 2 (see the corresponding explanations in [941]).

Figure 4.16 shows for silicon and germanium the proportionality of the drift velocity to the electric field over a wide field range as well as the saturation at high fields. At room temperature the saturation velocity is in the range of about 10^7 cm/s. Typical values for the low field mobilities of the charge carriers in silicon detectors at room

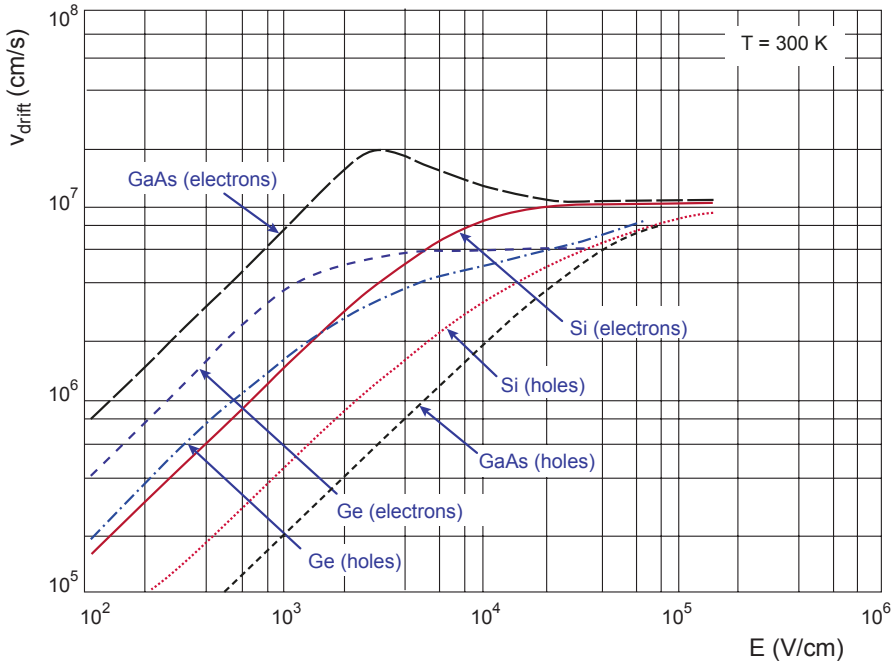


Fig. 4.16 Drift velocities of electrons and holes in silicon, germanium and gallium arsenide at 300 K as a function of the field strength (data taken from [565, 900, 838, 336, 929]). The saturation at high field strengths is a consequence of the mobility decrease as shown in fig. 4.15.

temperature are (see also table 8.2):

$$\mu_n(\mathrm{Si}) = 1450\,\frac{\mathrm{cm}^2}{\mathrm{Vs}}\,, \qquad \mu_\mathrm{p}(\mathrm{Si}) = 500\,\frac{\mathrm{cm}^2}{\mathrm{Vs}} \approx \frac{1}{3}\mu_\mathrm{n}\,.$$

For gallium arsenide (GaAs) the electron and hole mobilities are approximately (see table 8.2):

$$\mu_n(\mathrm{GaAs}) = 8800\,\frac{\mathrm{cm}^2}{\mathrm{Vs}}\,, \qquad \mu_\mathrm{p}(\mathrm{GaAs}) = 320\,\frac{\mathrm{cm}^2}{\mathrm{Vs}} \approx \frac{1}{30}\mu_\mathrm{n}\,.$$

For comparison, the mobilities of ions in gases are typically 1000 times smaller than those of electrons in silicon (table 4.2).

The drift motion of the charge carriers leads to a drift current

$$\vec{j}_\mathrm{drift} = e\,(p\,\mu_p - n\,\mu_n)\,\vec{E} = \sigma\,\vec{E}\,, \tag{4.128}$$

where n, p are the charge carrier densities and σ is the electric conductivity. In solids the mobility μ is usually measured exploiting the μ dependence of the electric conductivity:

$$\sigma = n\,q\,\mu\,. \tag{4.129}$$

Table 4.3 Transport parameters in some semiconductors: low field mobility μ_0, saturation drift velocity v_{sat} (at 300 K) at high fields, diffusion coefficient D (for intrinsic semiconductors), thermal velocities at 300 K (see table 8.1 and [929, 560]).

Semiconductor	μ_n $\left(\frac{\mathrm{cm}^2}{\mathrm{Vs}}\right)$	μ_p $\left(\frac{\mathrm{cm}^2}{\mathrm{Vs}}\right)$	v_{sat} $\left(\frac{\mathrm{cm}}{\mathrm{s}}\right)$	D_n $\left(\frac{\mathrm{cm}^2}{\mathrm{s}}\right)$	D_p $\left(\frac{\mathrm{cm}^2}{\mathrm{s}}\right)$	v_{th}^e $\left(\frac{\mathrm{cm}}{\mathrm{s}}\right)$	v_{th}^h $\left(\frac{\mathrm{cm}}{\mathrm{s}}\right)$
Si	1450	500	1×10^7	36	12	2.3×10^7	1.65×10^7
Ge	3900	1800	7×10^6	100	50	3.1×10^7	1.9×10^7
GaAs	8500	400	1.2×10^7	200	10	4.4×10^7	1.8×10^7
InP	5400	200	1×10^7	130	5	3.9×10^7	1.5×10^7
InAs	40 000	500	8×10^7	1000	13	7.7×10^7	2×10^7
CdTe	1050	90		26	2	3.5×10^7	1.4×10^7
diamond$^{(*)}$	≈1900	≈2300	7×10^6	57	46	$\approx 10^7$	$\approx 10^7$

$^{(*)}$ Values quoted in the literature for the mobility in diamond vary strongly [776, 409, 731, 563]. An approximate value is given in table 4.2. For plasma-deposited monocrystalline single crystal diamond, values as large as 4500 $\mathrm{cm}^2/\mathrm{Vs}$ for electrons and 3800 $\mathrm{cm}^2/\mathrm{Vs}$ for holes are reported in [562].

4.7.2 Diffusion in semiconductors

Owing to the thermal motion, spatial concentration differences of the charge carriers are equalised by diffusion of the charge carriers from regions of higher concentration to those of lower concentration. According to Fick's first law (4.68) this leads to a diffusion current which for the densities n and p of the electrons and holes, respectively, takes the form

$$\vec{j}_{n,\mathrm{diff}} = -eD_n\vec{\nabla}n\,, \qquad \vec{j}_{p,\mathrm{diff}} = -eD_p\vec{\nabla}p\,, \tag{4.130}$$

where $D_{n,p}$ are the diffusion coefficients for electrons and holes, respectively, which are specific for a certain semiconductor (table 4.3).

Together with the drift current (4.128) the total current is obtained as

$$\begin{aligned}\vec{j}_n &= \vec{j}_{n,\mathrm{drift}} + \vec{j}_{n,\mathrm{diff}} = -e\mu_n\, n\vec{E} - e\, D_n\,\vec{\nabla}n\,,\\ \vec{j}_p &= \vec{j}_{p,\mathrm{drift}} + \vec{j}_{p,\mathrm{diff}} = e\mu_p\, p\vec{E} - e\, D_p\,\vec{\nabla}p\,.\end{aligned} \tag{4.131}$$

For a system in thermal equilibrium, the Fermi level $\mathcal{E}_F$ (for definition and use of $\mathcal{E}_F$ see chapter 8, pages 267 ff)[9] is constant over a boundary with a concentration gradient, whereas the conduction and valence band edges ($\mathcal{E}_C$ and $\mathcal{E}_V$) bend up or down (see fig. 8.16 on page 275). For such systems the Einstein relation can be derived from (4.131) as follows (see also (4.85) and (4.92) as well as the footnote on page 106).

For electrons in equilibrium, that is, for $\vec{j}_n = 0$, (4.131) yields in the one-dimensional case (in the x direction):

$$\mu_n n\, E_x = -D_n \frac{\partial n}{\partial x}\,. \tag{4.132}$$

The electric field E_x determines the position of the conduction band edge $\mathcal{E}_C = -e\phi + const$ (see chapter 8), therefore yielding

[9] Note that for reasons of consistency within a chapter the use and meaning of the symbols E for the electric field and $\mathcal{E}$ for energy levels is reversed in chapter 8.

$$E_x = -\frac{\partial\phi(x)}{\partial x} = \frac{1}{e}\frac{\partial\mathcal{E}_C(x)}{\partial x} \tag{4.133}$$

with $\phi = \mathcal{E}/e$ being the electrical potential. The electron density at the conduction band edge is exponentially dependent on the distance of the Fermi energy to the conduction band $\mathcal{E}_C - \mathcal{E}_F$:

$$n = C\,\exp\left(-\frac{\mathcal{E}_C - \mathcal{E}_F}{kT}\right), \tag{4.134}$$

where C is a constant which is not needed any further. The derivative is

$$\frac{\partial n}{\partial x} = -\frac{1}{kT}\left\{C\,\exp\left(-\frac{\mathcal{E}_C - \mathcal{E}_F}{kT}\right)\right\}\frac{\partial\mathcal{E}_C}{\partial x} = -\frac{n}{kT}\frac{\partial\mathcal{E}_C}{\partial x}\,. \tag{4.135}$$

With (4.132) we have

$$\mu n E_x = \mu n\left(\frac{1}{e}\frac{\partial\mathcal{E}_C(x)}{\partial x}\right) = -D_n\left(-\frac{n}{kT}\frac{\partial\mathcal{E}_C}{\partial x}\right). \tag{4.136}$$

From this derivation for electrons and the corresponding one for holes, we obtain the *Einstein relation* which relates the diffusion coefficient and the mobility through fundamental physical constants (e, k):

$$\frac{D_n}{\mu_n} = \frac{kT}{e}, \qquad \frac{D_p}{\mu_p} = \frac{kT}{e}\,. \tag{4.137}$$

The corresponding relation holds for ions in gases (at moderate fields), as shown in section 4.6.1, see (4.92).

4.7.3 Motion of electrons and holes in magnetic fields

Semiconductor detectors are often employed as tracking detectors in magnetic fields. Under the influence of the Lorentz force $\vec{F} = q\,(\vec{E} + \vec{v}\times\vec{B})$ the charge carriers move between two collisions on curved trajectories. The resulting drift motion no longer follows the direction of the E-field, except if $\vec{E}$ and $\vec{B}$ are parallel. For the general case of arbitrary E- and B-fields the equations for the drift components are given by (4.54) to (4.56) in section 4.3.2.

Also in section 4.3.2 the deflection of the drift motion from the original direction by the Lorentz angle α_L was described for the case of perpendicular E- and B-fields. A simplified illustration of this effect is shown in fig. 4.4 on page 101. The strength of the deflection is given by the ratio of the drift velocity components perpendicular and parallel to the electric field, corresponding to (4.59) in section 4.3.2:

$$\tan\alpha_L = \frac{|v^B_{D,1}|}{v^B_{D,3}} = \omega\tau_H\,. \tag{4.138}$$

Here $\omega = q/m^*\,B$ is the cyclotron frequency and τ_H is the relaxation or collision time accounting for the effects of the magnetic field. The time constant τ_H enters in the Hall mobility μ_H (hence the subscript 'H') which is measured via the Hall effect [893, 402]. The Hall mobility is related to the mobility $\mu_{B=0}$ without magnetic field by the Hall factor r_H (see also section 4.3.2) defined by

$$r_H = \frac{\mu_H}{\mu_{B=0}} \,. \tag{4.139}$$

The mobility $\mu_{B=0}$ is measured via the conductivity employing (4.129). Equation (4.138) can now be rewritten as

$$\tan \alpha_L = \omega\tau_H = \frac{e}{m^*}\tau_H \, B = \mu_H B = \mu_{B=0} \, r_H \, B \,. \tag{4.140}$$

The Hall factor was already introduced in section 4.3.2 by (4.62) in the approximation of small fields:

$$r_H = \frac{\langle \tau^2 \rangle_\epsilon}{\langle \tau \rangle_\epsilon^2} \,. \tag{4.141}$$

The subscript ϵ on the averaging symbol indicates the energy weighted average as defined by (4.53).

While for semiconductors the Hall factor is defined by the measurement method, namely by the Hall effect, a corresponding factor occurs also for gases in the evaluation of the Lorentz angle. Without the assumed approximation of small magnetic fields the more complex averages in (4.54) to (4.56) have to be used rather than $\langle \tau \rangle_\epsilon$ and $\langle \tau^2 \rangle_\epsilon$ as in (4.141). Other approximations may have to be scrutinised as well.

Examples.

(1) In the simple—but not realistic—case that τ is constant, $\tau = \tau_H = \tau_\mu$ follows and thus $r_H = 1$.

(2) If one considers only acoustic phonon scattering (thermal lattice vibrations) the mean free path λ is constant and $\tau = \lambda/v$. The distribution function of the charge carriers is in good approximation a Maxwell–Boltzmann distribution which can be employed to determine r_H as follows:

$$r_H = \frac{\langle \tau^2 \rangle_\epsilon}{\langle \tau \rangle_\epsilon^2} = \frac{\int_0^\infty \tau^2 \, \epsilon^{3/2} \, \mathrm{e}^{-\frac{\epsilon}{kT}} \, d\epsilon}{\left(\int_0^\infty \tau \, \epsilon^{3/2} \, \mathrm{e}^{-\frac{\epsilon}{kT}} \, d\epsilon \right)^2} = \frac{\frac{3kT}{m^*}}{\frac{8kT}{\pi m^*}} = \frac{3\pi}{8} = 1.18 \,. \tag{4.142}$$

For scattering processes in which λ is not constant and/or the assumed approximations are not valid, r_H takes other values. As a rule, $r_H \geq 1$ is valid for non-degenerate semiconductors with symmetric band structure. However, $r_H < 1$ also occurs, in particular for p-silicon with non-symmetric curved band structure [658, 560]. For silicon at 300 K and small magnetic fields ($\omega\tau_H \ll 1$) one finds $r_H \approx 1.15$ for electrons and $r_H \approx 0.8$ for holes [659] in a wide range of doping concentrations up to about $10^{15}\,\mathrm{cm}^{-3}$. For large magnetic fields, as typical for particle detectors, $r_H \approx 1$ applies.

In semiconductor detectors with structured electrodes, like microstrip or pixel detectors, the Lorentz angle α_L plays an important role for the optimisation of the electrode dimensions and pitches in order to obtain the best possible spatial resolution. Because the Hall mobility for electrons in silicon is a factor of 4.5 larger than for holes the Lorentz angle for electrons is also larger. At a magnetic field of 1 T and for moderate electric fields, where the mobility is still field independent, α_L amounts to 9.5° for electrons and to 2.1° for holes. The Lorentz angle possibly changes with the operation time of the detector due to radiation damage and because higher electric fields may be required, leading to shorter collision times and hence smaller mobilities and (usually) smaller α_L (see e.g. [469]).

5
Signal formation by moving charges

5.1 Introduction

Particles can only be detected if they deposit energy in a detector medium. Often at least part of that energy is used to ionise the surrounding atoms, thereby creating charge carriers (electron/ion or electron/hole pairs). Detectors sensing those charge carriers are for instance gas-filled ionisation or proportional chambers, semiconductor detectors, photodiodes and others. By separating the ionisation charges in an electric field, electrical signals can be produced on external electrodes which can be further electronically processed. The possibility of processing measured data electronically is one of the most important construction criteria for modern detectors and will only be overruled if there are other, more important criteria—for example photoemulsion may be chosen if particularly precise tracking is required. Even in detectors where the primary signal is not due to ionisation usually the detection chain nonetheless comes down to the detection of ionisation. For example, Cherenkov or scintillation photons are converted to ionisation charge when hitting the photosensitive layers of photodetectors like photomultipliers or photodiodes (chapter 10 specifically addresses the processing of light signals).

In ionisation detectors the formation of signals that can be electronically processed proceeds along the following steps:

- a particle deposits energy in the detector volume;
- charge pairs (electron–ion or electron–hole) are created;
- the charges are separated by an electric field generated by an electrode system moving them towards the electrodes;
- the moving charges induce an accumulation of charges on the electrode surfaces;
- the charge induction will be recorded as a charge, current or voltage signal.

For further discussion of signal formation in electrode systems it is fundamentally important to understand that signals on electrodes form by the motion of charges relative to the electrodes. Therefore it is not important for the signal generation that the charge actually arrives at the electrode. A charge can get lost in the ionisation volume, for example by recombination or attachment, but the signal from a preceding

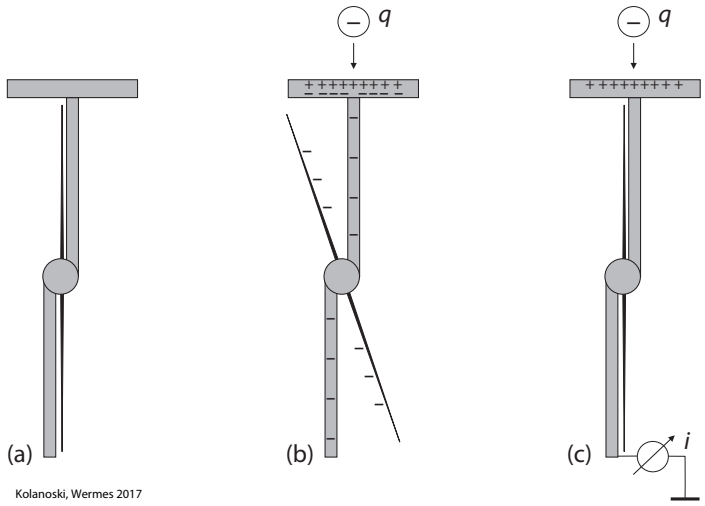


Fig. 5.1 Effect of a charge approaching an electrometer. (a) An uncharged electrometer which is insulated against ground. (b) A charge q, approaching from infinity, induces a counter-charge on the metal surface of the electrometer which increases as the distance becomes smaller. Within the free-floating electrometer the charge is conserved and hence can only be separated. The charge with sign opposite of that of q (here $q < 0$) accumulates on the metal surface close to q and the same sign charge accumulates on the surfaces further away. This generates a deflection of the electrometer's needle. (c) Grounding the electrometer pedestal allows the (negative) charge to drain off so that a current signal can be measured on the path to ground.

movement of that charge can still be detected. Hence 'charge collection' on the electrodes is not the essential phenomenon, but it is rather the movement of the charge relative to the electrodes.

The conditions for signal formation can be illustrated by the simple example shown in fig. 5.1, which is familiar from lectures on electrostatics. When a charge q approaches the electrode of an insulated electrometer an opposite sign charge will be induced, as depicted in fig. 5.1(b). The closer q comes, the larger the deflection of the electrometer's needle becomes. When q stops moving the needle deflection does not change. After grounding the electrometer (fig. 5.1(c)) a measurable current flows which will be a short discharging current pulse if the external charge does not move further. Else, if the external charge moves continuously, a continuous current will flow which will be the stronger the faster q moves and will disappear when q arrives at the electrode or, more generally, when q does not move anymore. This current can be used for forming an electronic signal. Note that, as mentioned before, for a signal to form it is not necessary that the charge is 'collected' by the electrode.

Since ionisation detectors are usually treated as current sources in equivalent circuits the current can be considered as the basic observable. By measuring the current over a resistor or by integrating it on a capacitor the signal can be converted into a voltage or a charge signal, respectively.

5.2 Weighting field and Shockley–Ramo theorem

5.2.1 Weighting potential and weighting field

To obtain general statements about the formation of signals by moving charges we initially consider a simple system of two electrodes, as sketched in fig. 5.2 (left). A point charge q moves in a closed volume bordered by outer and inner conductors, each of them laid on a fixed potential. In this example the voltage between the two conductors is V and the inner conductor is grounded. The voltage V leads to induction charges Q and $-Q$, respectively, on the electrodes. For a capacitance C of the arrangement the induction charge is given by $Q = C\,V$. The charge q at location $\vec{r}_q$ induces an additional charge $\Delta Q(\vec{r}_q)$. While moving the charge q from $\vec{r}_q$ to $\vec{r}_q + d\vec{r}_q$ the field of the electrodes, $\vec{E}_0$, does the work:[1]

$$dW_q = q\,\vec{E}_0\,d\vec{r}\,. \tag{5.1}$$

This work has to be delivered by the power supply (W_V) and/or the field energy (W_E):

$$dW_q + dW_V + dW_E = 0\,. \tag{5.2}$$

If the voltage is kept constant ($dV = 0$) the work done by the power supply is only determined by the charge dQ extracted from the power supply:

$$dW_V = dQ\,V + Q\,dV = dQ\,V\,. \tag{5.3}$$

Here we adopt the convention that dQ is negative if negative charge is moved into the power supply (or, equivalently, if positive charge is taken from the power supply) and that dQ has the opposite sign as the charge induced on the electrode surface.

The total field energy in the volume τ bounded by the electrodes is

$$W_E = \frac{1}{2}\epsilon\epsilon_0 \int_\tau \vec{E}^2 d\tau\,, \tag{5.4}$$

where ϵ_0 is the electric field constant, also called vacuum permittivity, and ϵ the relative permittivity. In the following it will be shown that this field energy does not change when the charge q moves in the field.

As illustrated in fig. 5.2, the field $\vec{E}$ can be expressed as a linear superposition of the field $\vec{E}_0$ generated by the electrodes (without charge in the volume) and the field $\vec{E}_q$ generated by the additional charge:

$$\vec{E} = \vec{E}_0 + \vec{E}_q\,. \tag{5.5}$$

According to (5.1) the work on the charge q will be done by the field $\vec{E}_0$ at the position of the charge, while the field generated by the charge q itself, $\vec{E}_q$, does not contribute to the work, as follows from the laws of electrodynamics.

Accordingly, the potential $\phi(\vec{r})$ in the volume bounded by the electrodes can be represented by the sum of a potential ϕ_0 without charge q and the potential resulting from the additional charge q:

[1] The particle carrying the charge q usually gains kinetic energy by the work of the field. This energy is mostly dissipated in interactions with the detector medium. However, for the considered energy balance of the electrode system only the delivered work of the field is relevant, no matter how this energy is dissipated.

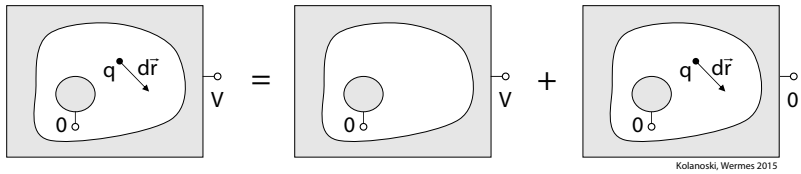


Fig. 5.2 Sketch illustrating the movement of a charge in an arrangement of potentials. The arrangement on the left hand side with the inner conductor at ground and the outer conductor at the potential V can be decomposed into the two arrangements on the right-hand side. See also the discussion in the text.

$$\phi(\vec{r}) = \phi_0(\vec{r}) + \phi_q(\vec{r})\,. \tag{5.6}$$

The boundary conditions for the potentials on the outer and inner electrode surfaces, S_a and S_i, respectively, are

$$\begin{aligned} \phi(\vec{r})|_{S_a} &= \phi_0(\vec{r})|_{S_a} = V, \\ \phi(\vec{r})|_{S_i} &= \phi_0(\vec{r})|_{S_i} = 0\,, \\ \phi_q(\vec{r})|_{S_a} &= \phi_q(\vec{r})|_{S_i} = 0\,. \end{aligned} \tag{5.7}$$

Hence the potential ϕ_0 fulfils these conditions given by the fixed voltages whereas the potential ϕ_q due to the charge q is to be determined with all electrodes grounded. The potentials result from the solutions of the Laplace or Poisson equation,

$$\Delta\phi_0 = 0\,, \tag{5.8}$$

$$\Delta\phi_q = -\frac{q}{\epsilon\epsilon_0}\delta(\vec{r} - \vec{r}_q)\,, \tag{5.9}$$

with the boundary conditions (5.7). The corresponding fields are

$$\vec{E}_0 = -\vec{\nabla}\phi_0\,, \tag{5.10}$$

$$\vec{E}_q = -\vec{\nabla}\phi_q\,. \tag{5.11}$$

Using Green's theorem one can show (see e.g. [512]) that the field energy of a static electrode field (approximately also of a low-frequency field) is also separable into the two components from the electrode field and from the point charge field:

$$W_E = W_{E_0} + W_{E_q}\,. \tag{5.12}$$

Under a movement of the charge q both these field energy contributions do not change,

$$dW_E = dW_{E_0} + dW_{E_q} = 0\,, \tag{5.13}$$

since according to the assumptions, the field $\vec{E}_0$ is static because of the constant voltage and the field $\vec{E}_q$ neither does work on the charge q nor does it exchange energy with the power supply as the electrodes are kept on fixed potential (here on ground). Therefore it follows from (5.1), (5.3) and (5.13):

$$dW_q + dW_V = q\,\vec{E}_0\,d\vec{r} + dQ\,V = 0 \qquad \Rightarrow \qquad dQ\,V = -q\,\vec{E}_0\,d\vec{r}\,. \tag{5.14}$$

This means that the work on the charge q is exclusively provided by the voltage source yielding

$$dQ = -q\,\frac{\vec{E}_0}{V}\,d\vec{r}\,. \tag{5.15}$$

The field $\vec{E}_0$ is given by the geometry of the electrode arrangement and its absolute value is proportional to V. This means that in (5.15) the quantity $\vec{E}_0/V$ and thus also the induced charge dQ are independent of the applied voltage V. Hence we can simply set the voltage to $V = 1$ (in any units) to determine the so-called *weighting potential* and the corresponding *weighting field*:

$$\phi_w = \frac{\phi_0}{V}\,, \qquad \vec{E}_w = -\vec{\nabla}\phi_w\,. \tag{5.16}$$

Thus the weighting potential ϕ_w is the specific potential ϕ_0 that solves the potential equation (5.8) with the boundary condition $V = 1$. According to its definition ϕ_w is dimensionless and $\vec{E}_w$ has the unit of an inverse length. In the following we will show that the weighting field concept goes beyond a simple division by V, particularly when dealing with multi-electrode systems.

The charge signal in (5.15) can now be expressed using the weighting field:

$$dQ = -q\,\vec{E}_w\,d\vec{r}\,. \tag{5.17}$$

The time evolution of the induced charge is given by the charge motion $\vec{v}_D$ and corresponds to the signal current,[2] which is drawn from the voltage source while keeping the voltage constant:[3]

$$i_S = -\frac{dQ}{dt} = q\,\vec{E}_w\,\vec{v}_D \qquad \text{with} \quad \vec{v}_D = \frac{d\vec{r}}{dt}\,. \tag{5.18}$$

In general, the direction of the charge velocity $\vec{v}_D$ can be different from the direction of the weighting field $\vec{E}_w$, for example in the case of a superimposed magnetic field or in multi-electrode systems (see next section for the generalisation to systems with an arbitrary number of electrodes).

Presence of polarisation charges. We have already pointed out that the weighting field depends only on the geometrical arrangement of the electrodes. For the derivation of the weighting field we did not explicitly exclude the case that dielectric polarisation charges exist in the detector volume. The relative permittivity ϵ enters into the field energy W_E term in (5.4) and in the potential equation for the point charge (5.9). Therefore it might be surprising that the result for the weighting field is independent thereof. The reason is that the field of the point charge is inversely proportional to ϵ because the divergence of the field is

$$\vec{\nabla}\cdot\vec{E}_q = \frac{q}{\epsilon\,\epsilon_0}\,\delta(\vec{r}-\vec{r}_q)\,. \tag{5.19}$$

For the induced charge this permittivity dependence is cancelled because, according to Gauss's theorem, the charge induced on the outer conductor is given by

$$\Delta Q = \epsilon\epsilon_0 \int_{S_a} \vec{E}_q\,d\vec{S}\,. \tag{5.20}$$

[2] See the remarks on the sign convention for the induced current in section 5.2.4.

[3] For current and voltage signals, $i_S(t)$ and $v_S(t)$, respectively, lower case letters are used.

The electrode signals are independent of the polarisation charges in the volume between the electrodes to the extent that the dielectric properties of the medium can be treated linearly.

Presence of space charge. In practice it is important that the weighting field is also independent of possible space charge in the detector medium, as long as the space charge remains locally fixed. An important example is the space charge in a depleted semiconductor layer. Though the space charge changes the electric field that causes the movement of the charge, and thus the drift velocity $\vec{v}_D$, it does not change the weighting field. This can be explained as follows. In fig. 5.2 one can add a space charge contribution with the density ρ in the inner volume and with vanishing potentials on the electrodes. This corresponds to an extension of the potential in (5.6) by a potential $\phi_\rho(\vec{r})$,

$$\phi(\vec{r}) = \phi_0(\vec{r}) + \phi_q(\vec{r}) + \phi_\rho(\vec{r})\,, \tag{5.21}$$

which fulfils the Poisson equation with boundary conditions:

$$\Delta\phi_\rho = -\frac{\rho}{\epsilon\epsilon_0} \qquad \text{with} \qquad \phi_\rho(\vec{r})|_{S_a} = \phi_\rho(\vec{r})|_{S_i} = 0\,. \tag{5.22}$$

Considering the energy balance for this field shows that it does not contribute to the signal. Since the electrodes are grounded no energy can be exchanged with the power supply, and since the space charge is stationary the field energy does not change either. Both together lead to the conclusion that this field cannot do work on the point charge and thus neither contributes to the energy balance (5.14) nor to the weighting field.

Signal computation using weighting fields. An important advantage of signal computation by means of weighting fields according to (5.17) and (5.18) is the decoupling of the charge motion (caused by the electric field) from the field which determines the charge induction. The direction of motion does not have to coincide with the direction of the weighting field. Such a situation arises, for example, if the charge motion is influenced by a magnetic field, or if the field is formed by many electrodes (see next section). Another advantage is that the algorithms can simply be extended to systems with an arbitrary number of electrodes. The task to compute the signals of a moving charge on all electrodes, which appears in the first instance confusingly complex, becomes quite manageable by employing the Shockley–Ramo theorem as will be discussed in the next section.

5.2.2 Shockley–Ramo theorem

We consider an arrangement of k electrodes to which voltages $V_1,\ V_2, \ldots,\ V_k$ are applied. Thanks to the superposition principle the potential generated by the electrodes $\phi_0(\vec{r})$ in the space between the electrodes can be represented by a sum over different configurations of potentials ϕ_i $(i = 1, \ldots, k)$. For each configuration i all electrodes remain at a potential 0 except for the potential of the electrode i which lies on its actual voltage V_i (fig. 5.3):

$$\phi_0(\vec{r}) = \sum_{i=1}^{k} \phi_i(\vec{r})\,, \qquad \text{with} \qquad \begin{array}{l} \phi_i|_{S_i} = V_i\,, \\ \phi_i|_{S_j} = 0\,, \quad j \neq i\,. \end{array} \tag{5.23}$$

The potential ϕ_0 is defined analogously to the decomposition (5.6) or (5.21). That means ϕ_0 is the potential between the electrodes calculated without including the point

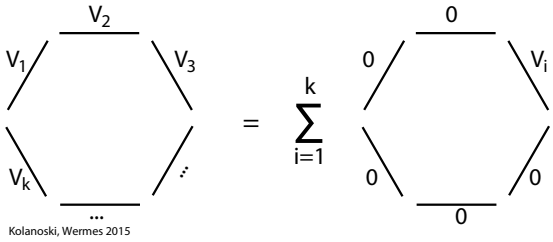


Fig. 5.3 Schematic illustration of a system with k electrodes lying on respective potentials V_i $(i = 1, \ldots, k)$. This system can be represented by a sum of k arrangements, each one with another electrode on its actual potential and all others grounded.

charge and possibly existing polarisation and space charges. The weighting potential and weighting field, respectively, for each configuration i are

$$\phi_{w,i}(\vec{r}) = \frac{\phi_i(\vec{r})}{V_i}\,, \qquad \vec{E}_{w,i} = -\vec{\nabla}\phi_{wi}\,. \tag{5.24}$$

Each weighting potential $\phi_{w,i}$ individually fulfils the Laplace equation with the above defined boundary conditions $V_i = 1$ and all other $V_{j\neq i} = 0$:

$$\Delta\phi_{w,i}(\vec{r}) = 0\,, \qquad \text{with} \quad \phi_{w,i}|_{S_i} = 1, \quad \phi_{w,i}|_{S_{j\neq i}} = 0\,. \tag{5.25}$$

Generalising (5.17) and (5.18), the Shockley–Ramo theorem [891, 806] states that the signal induced by a moving charge q on an electrode i is given by the weighting potential or the weighting field of the potential configuration $\phi_i(\vec{r})$:

$$\begin{aligned} dQ_i &= -q\,\vec{E}_{w,i}\,d\vec{r}\,, \\ i_{S,i} &= q\,\vec{E}_{w,i}\,\vec{v}_D\,. \end{aligned} \tag{5.26}$$

Here and in the following the charge signal dQ_i has the sign as adopted for (5.3): 'dQ_i is negative if positive charge is taken from the power supply and has the opposite sign to the charge induced on the electrode surface'. In the following we denote dQ_i and $i_{S,i}$ as defined in (5.26) as 'signal charge' and 'signal current', respectively. The Shockley–Ramo theorem implies that the signals dQ_i or $i_{S,i}$ on an electrode do not directly depend on the strength of the electric field between the electrodes or on the applied high voltage. However, the electrical field usually determines the direction and the velocity of the charge motion and thus according to the Shockley–Ramo theorem (5.26) the signal shape. The weighting field $\vec{E}_{w,i}$, which depends only on the geometry of the arrangement, determines the signal on the electrode i generated by the movement of the charge with the size and sign of the signal current being proportional to the velocity component in the field direction. Figure 5.4 shows an example with four electrodes. The weighting field and the actual electric field, and thus the direction of motion of the charge, have different orientations at the position of the charge. If in addition space charges are present, as in semiconductor detectors, then even in two-electrode systems the weighting and electric fields are no longer proportional. For instance, in a parallel plate capacitor with homogeneous space charge the electric field increases linearly with the distance x from an electrode as $E \propto x$. In contrast the weighting field remains constant, given as $E_w = 1/d$ where d is the electrode distance.

The Shockley–Ramo theorem was proven in the previous section for a two-electrode system using energy balance arguments. The generalisation to multi-electrode systems only exploits the superposition in (5.23) in addition and the fact that in the configura-

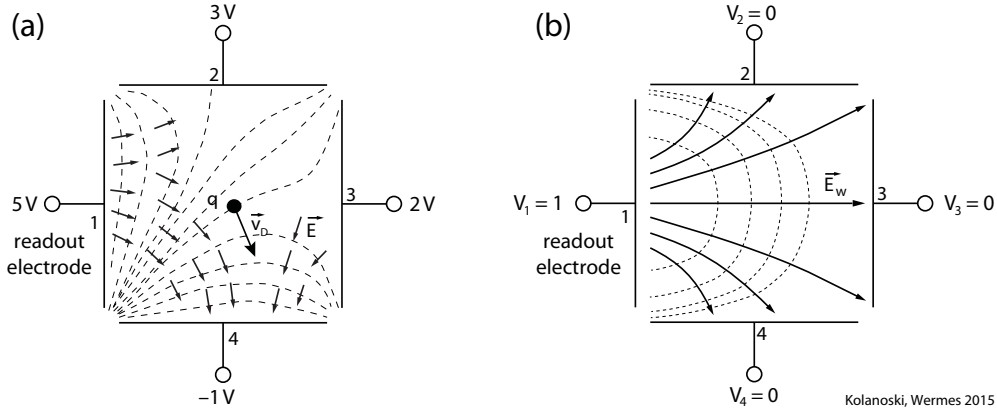


Fig. 5.4 Example of an arrangement of four electrodes. (a) Actual field and potential configuration and the resulting direction of motion of a charge; (b) Configuration for the computation of the weighting field corresponding to electrode 1. Field profiles are depicted by arrows and potential contours by dashed lines.

tion where all but one electrode are grounded (as in fig. 5.2) the grounded electrodes do not perform work.

In several publications it was clarified that the weighting potential should be computed without including stationary charges (space charges) and polarisation charges [572,282,447,512], as long as the dependence of the polarisation on the applied voltage remains linear.[4] The proof has been sketched at the end of the previous section (see also eqs. (5.19) and (5.20) on page 131).

Extensions of the weighting field method as shown here and in the following can be found in the literature. Time dependent weighting fields are treated in [449]. For silicon it was shown that time-dependent weighting fields are required to describe the response of substrates with un-depleted or partially depleted regions [822] which are (partially) conductive. To what extent this affects silicon sensors damaged by radiation (see section 8.12.1) has been studied in [875] concluding that the weighting field of a highly irradiated silicon detector is about the same as that of a fully depleted, undamaged detector due to a compensating effect, since by irradiation the resistivity of the substrate is changed (see also section 8.12.1).

5.2.3 Recipe for the computation of signals on electrodes

For practical applications of the Shockley–Ramo theorem a recipe for the step-by-step calculation of signals is given here:

- Determine the weighting field $\vec{E}_{w,i}(\vec{r})$ of an electrode i by setting the potentials to $V_i = 1$ and $V_{j\neq i} = 0$.
- Determine the real field $\vec{E}(\vec{r})$ between the electrodes.
- Determine the drift velocity $\vec{v}_D(\vec{r}(t))$ of the charge q moving along the trajectory $\vec{r}(t)$; without a magnetic field the velocity is $\vec{v}_D(\vec{r}(t)) = \mu\vec{E}(\vec{r}(t))$, where μ is the mobility (see chapter 4).

[4] The general case which includes nonlinear behaviour of the medium and space charges is discussed in [500].

– Determine the signal current according to $i_{S,i}(t) = q\,\vec{E}_{w,i}\,\vec{v}_D$.

Knowing the space-time relation $\vec{r}(t)$ one obtains the time evolution of the signal current $i_{S,i}(t)$ at the electrode i. Integrating the current over the time interval t_0 to t provides the detected charge:

$$\begin{aligned} Q_{S,i}(t) &= -\int_{t_0}^{t} i_{S,i}(t')\,dt' = -q\int_{t_0}^{t} \vec{E}_{w,i}\,\vec{v}_D\,dt' \\ &= -q\int_{\vec{r}(t_0)}^{\vec{r}(t)} -\vec{\nabla}\phi_{w,i}\,d\vec{r} = q\left[\phi_{w,i}(\vec{r}(t)) - \phi_{w,i}(\vec{r}(t_0))\right]\,. \end{aligned} \tag{5.27}$$

Here the relative sign is fixed by the 'conventional current direction' (see explanations in the next section).

5.2.4 On the sign of the induction current

A current I is defined as the integral of the current density $\vec{j}$ over the area $\vec{A}$ that $\vec{j}$ passes:

$$I = \int_A \vec{j}d\vec{A}\,. \tag{5.28}$$

The direction of the current density is unambiguously given by the velocity vector $\vec{v}_D$ of charge carriers:

$$\vec{j} = n\,q\,\vec{v}_D\,, \tag{5.29}$$

where n is the density of the charge carriers and q is their charge. For positive charges the current density vector has the direction of the charge movement and for negative charges the inverse direction. In the definition of (5.28) the size of the current is determined but not its sign since the orientation of the area $\vec{A}$ can be arbitrarily chosen. We define the 'conventional current direction' as the direction of the movement of positive charges or the opposite direction of the movement of negative charges, implying that the area in (5.28) is oriented in the direction of the current density $\vec{j}$. Consequently, with this definition, the motion of free positive or negative charges in an electric field always leads to a positive current.

The integral in (5.28) over a closed surface S with the surface's normal pointing outwards yields

$$\oint_S \vec{j}d\vec{A} = -\frac{d}{dt}Q_{inner} = -\frac{d}{dt}\int_V \rho dV\,, \tag{5.30}$$

where the current flow corresponds to a decrease of charge Q_{inner} inside the surface area. The differential form of (5.30) yields the continuity equation,

$$\vec{\nabla}\cdot\vec{j} = -\frac{\partial\rho}{\partial t}\,, \tag{5.31}$$

which is valid independent of any sign convention. When applied to a current node that is surrounded by a closed surface, (5.30) yields Kirchhoff's nodal rule. According to this rule the sum of all currents at a node is zero, provided there is no source or sink in the node, corresponding to $dQ_{inner} = 0$. Thus, the nodal rule controls the relative signs of incoming and outgoing currents at a node.

We consider now a charged, free-floating capacitor which stores a charge $Q_C > 0$ and $-Q_C$ on its positive and negative electrode, respectively. A discharge, either by

shorting the electrodes or by generating ionisation charges in the volume between the electrodes, leads to a decrease of the charge on the positive electrode ($dQ_C < 0$) and to an increase on the negative electrode. Adopting the 'conventional current direction' the discharge of a capacitor always leads to a positive current:

$$I = -\frac{dQ_C}{dt} \quad (dQ_C < 0) \quad \text{(conventional current direction)}\,. \tag{5.32}$$

That a positive current leads to the decrease of charge according to (5.32) might not seem very intuitive and therefore the sign of the current is often reversed [800]. Adopting this convention the discharge of a capacitor leads to a negative current accompanied by a decrease of charge on the electrodes:

$$I = \frac{dQ_C}{dt} \quad (dQ_C < 0) \quad \text{(inverted current direction)}\,. \tag{5.33}$$

The discharge of a capacitor with capacitance C which is decoupled from a voltage source leads to a voltage jump across the capacitor,

$$dv = \frac{1}{C}\, dQ_C\,, \tag{5.34}$$

which is always negative since in both conventions dQ_C is negative.

Both sign conventions for currents are used in the literature. For the derivation of the Shockley–Ramo theorem usually the conventional current direction is chosen (for example in [512]), while for the description of signal processing the inverted current direction is preferred (e.g. in [799]). Often the conventions are swapped even within an article or talk when turning from the theoretical description of the Shockley–Ramo theorem to applications. Therefore, when the sign of the currents is important, one should pay attention to the convention used.

Naturally the signs of the physically observable quantities, like charges and voltages, have to be independent of the convention. In fact, the charge induced by a positive charge on a conductor surface is always negative. When the positive charge moves away from the electrode the induced charge will become more positive. At fixed electrode potential this positive charge will be delivered by the voltage source. The change of the induced charge on the electrode is positive but according to the convention we used in (5.3) the charge taken from the voltage source has the reverse sign.

5.3 Signal formation in two-electrode systems without space charge

To begin the discussion of signal formation we restrict ourselves to systems with only two electrodes, without space charge and without magnetic field ($\rho = 0$, $B = 0$). In such a situation the drift velocity of a charge usually has the same or opposite direction of the electric field. Polarisation charges in a linear isotropic medium with a relative permittivity $\epsilon \neq 1$, should not be excluded (according to the Shockley–Ramo theorem current and charge signals do not depend on polarisation). Furthermore we assume that the charges drift in the electric field along the field lines (without magnetic field) with a velocity that is constant for a given field strength. We further assume that no charge is lost along the drift path. Examples for such systems are ionisation chambers filled with a gas or a liquid as well as gas-filled proportional and drift chambers. For

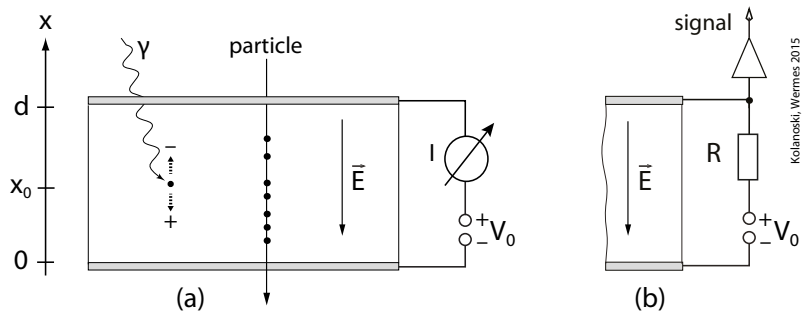


Fig. 5.5 Parallel plate detector (plate capacitor) as ionisation chamber. The figure shows examples for the generation of a point-like primary charge distribution by a photon and of a linear charge distribution by a through-going charged particle. The ionisation charges will be sucked to the electrodes by the electric field generated by the applied voltage V_0. In (a) the current of the ionisation charges is measured; in (b) the voltage drop, caused by the current flow over the resistor, can be recorded (here the signal readout is assumed to be high-ohmic).

signal formation it is an important property of gases and liquids that the drift velocity of electrons for usual drift fields typically is about three orders of magnitude larger than that of ions.

5.3.1 Signal formation in homogeneous electric fields

We consider an ionisation chamber consisting of a parallel plate capacitor with an approximately homogeneous electric field (fig. 5.5). When charged particles or photons deposit energy in the chamber medium, free charges are created which then drift to the electrodes. With the plate distance d, the plate area A and an applied voltage V_0 the electric field and the capacitance are

$$\vec{E} = -\frac{V_0}{d}\vec{e}_x \ , \qquad C = \frac{\epsilon\epsilon_0 A}{d} \, . \tag{5.35}$$

The weighting field for the readout of the positive electrode is obtained with $V_0 = 1$ and $V = 0$ on the opposing electrode:

$$\vec{E}_w = -\frac{1}{d}\,\vec{e}_x \, . \tag{5.36}$$

Electron–ion pairs which are generated in the capacitor volume are separated by the electric field and move to the opposite electrodes with constant drift velocity. Drift with acceleration will be treated in section 5.4.1 for configurations with space charge.

Local charge deposition. We will assume that an electron–ion pair with charges $q^- = -e$ and $q^+ = +e$ is generated at x_0, for example by absorption of a photon, and that the charges drift to the corresponding electrodes. According to the Shockley–Ramo theorem (5.26) the signal current is

$$i_S^{\pm} = q^{\pm}\vec{E}_w\,\vec{v}_D^{\,\pm} = -\frac{q^{\pm}}{d}\,\vec{e}_x\,\vec{v}_D^{\,\pm} = \frac{e}{d}v_D^{\pm} \, , \tag{5.37}$$

where $v_D^{\pm} = |\vec{v}_D^{\,\pm}|$ is the magnitude of each of the vectors and $e = |q^{\pm}|$ is the elementary charge. Since charges with opposite sign move in opposite directions the signal current

of both charges is positive (employing the conventional direction), as required for a discharge current (see section 5.2.4). Integrating over the time until the charges reach the electrodes,

$$T^- = \frac{d - x_0}{v_D^-}\,, \qquad T^+ = \frac{x_0}{v_D^+}\,, \tag{5.38}$$

yields the total signal charge:

$$\begin{aligned} Q_S^{tot} &= Q_S^- + Q_S^+ = -\frac{e}{d}\left(\int_0^{T^-} v_D^- dt + \int_0^{T^+} v_D^+ dt\right) \\ &= -\frac{e}{d} v_D^- \left(\frac{d - x_0}{v_D^-}\right) - \frac{e}{d} v_D^+ \left(\frac{x_0}{v_D^+}\right) = -e\,. \end{aligned} \tag{5.39}$$

The sign in front of the integral corresponds to the convention in (5.32). According to (5.39) the contributions of the positive and of the negative drifting charges to the total induced charge depend on the generation point x_0. At the opposite electrode the sign of the signal pulse is opposite since the weighting field in (5.36) for this electrode is reversed while the velocities maintain their directions.

In typical ionisation chambers the contributions of both charge components to the signal currents are very different due to the different drift velocities for electrons and ions in gases and liquids (see above):

$$i_S^- = \frac{e}{d}\, v_D^- \;\gg\; i_S^+ = \frac{e}{d}\, v_D^+\,. \tag{5.40}$$

This is depicted in fig. 5.6. The instantaneously starting current is different in magnitude for electrons and ions but for one charge type it has the same magnitude on both electrodes, independent of x_0. This changes if the electrodes are segmented (see section 5.5).

The signal charge (5.26),

$$dQ_S^\pm = -i_S^\pm dt = -\frac{e}{d} v_D^\pm dt\,, \tag{5.41}$$

leads to a measurable voltage change on a capacitor if the electrodes are isolated from the voltage source or decoupled from it by a high resistance R (see fig. 5.5(b)):

$$dv_S^\pm = \frac{1}{C} dQ_S^\pm = -\frac{q}{C} \vec{E}_w d\vec{r} = -\frac{e}{Cd}\; v_D^\pm dt\,. \tag{5.42}$$

The voltage change of a discharging capacitor is always negative, as was already pointed out in connection with (5.34). In the equation above C is the detector capacitance including possible parasitic capacitances of the wiring. Note that the relative permittivity ϵ enters through the capacitance C into the computation of the voltage signal while the induced currents and charges are independent of the polarisation charges, as argued in section 5.2.1. To be able to integrate the charge the resistor R in fig. 5.5(b) has to be chosen such that the time constant for charging the capacitor, $\tau = RC$, is large compared to the drift times.[5] Then the time evolution of the voltage signal $v_S(t)$ is

[5]The integration of the induced charge accompanied by a corresponding voltage change over the electrodes does not satisfy the conditions under which the Shockley–Ramo theorem was derived,

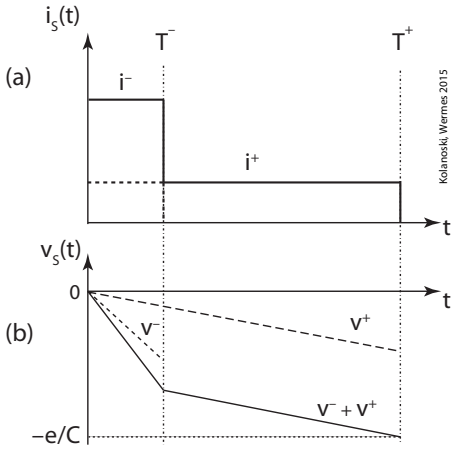


Fig. 5.6 Schematic illustration of (a) current and (b) voltage signals of a charge pair in an ionisation chamber in parallel plate configuration as in fig. 5.5. Currents and voltages are given by the equations (5.37) and (5.43), respectively. For clarity of the illustration, the drift velocity of the ions is set to only 1/3 of the electron velocity instead of the typical 1/1000.

$$v_S(t) = v_S^-(t) + v_S^+(t) = \begin{cases} -\dfrac{e}{Cd}(v_D^- + v_D^+)t & 0 < t < \min(T^-, T^+)\,, \\ -\dfrac{e}{Cd}(d - x_0 + v_D^+ t) & T^- < t < T^+\,, \\ -\dfrac{e}{Cd}(x_0 + v_D^- t) & T^+ < t < T^-\,, \\ -\dfrac{e}{Cd} & t > \max(T^-, T^+)\,. \end{cases} \quad \text{for} \tag{5.43}$$

In fig. 5.6 we restrict ourselves to the case $T^- < T^+$ since T^- is almost always smaller than T^+ due to the very different drift velocities.

Hence both the voltage and charge signals exhibit a linear increase with time, with different slopes for the electron and the ion component, respectively (fig. 5.6). If predominantly only the electron signal is observed, for example because the longer drifting ions recombine, the magnitude of the signals become dependent on the charge generation point, a feature which one usually tries to avoid.

For gas-filled ionisation chambers, as used for dosimetry, usually only the current generated by many ionisation processes can be measured with a circuit, as in fig. 5.5(a). Without gas amplification (see section 7.4.1) a single ionisation with a charge of about 10^{-19} C is not measurable. One should also realise that a measurable current of, let's say, 1 pA requires an ionisation rate of about 10^7 Hz. In denser media, solid or liquid, ionising tracks also generate sufficient charge without amplification so that single pulses can be observed.

Charge deposition along a particle track. Another important case, besides the local charge deposition, is the generation of charge as a uniform line charge along a track with a charge density $dq^{\pm}/dx = \pm Ne/d$ (fig. 5.5), which is for each charge sign normalised to the total charge, $q_{tot}^{\pm} = \pm Ne$.

specifically at fixed voltage. We assume here that the voltage change is sufficiently small so that the drift velocities do not change. Since the weighting fields are independent of the voltage the theorem should also hold in this case. This argument has been confirmed for the examples discussed in this chapter by comparing our calculations to results obtained employing the concept of time-dependent weighting fields (see e.g. [449, 820]).

The current signal induced on the readout electrode by the movement of the charges results from the contributions of all drifting individual charges. The charges continuously reach the electrode leading to a linearly decreasing current. To determine the current contributions at a time t (5.37) has to be integrated over that part of the line charge for which the charges did not yet arrive at the electrodes:

$$\begin{aligned} i_S^-(t) &= \frac{v_D^-}{d}\int_{v_D^- t}^{d}\frac{Ne}{d}dx = \frac{Ne}{T^-}\left(1-\frac{t}{T^-}\right) && \text{for } 0<t<T^- \\ i_S^-(t) &= 0 && \text{else} \\ i_S^+(t) &= \frac{v_D^+}{d}\int_{0}^{d-v_D^+ t}\frac{Ne}{d}dx = \frac{Ne}{T^+}\left(1-\frac{t}{T^+}\right) && \text{for } 0<t<T^+ \\ i_S^+(t) &= 0 && \text{else} \end{aligned} \tag{5.44}$$

where now $T^- = d/v_D^-$ and $T^+ = d/v_D^+$ are for both charge carriers the longest possible times to drift over the full detector thickness. After arrival of the last charge of each type the corresponding current signal vanishes.

The charge and voltage signals (for high impedance signal readout) are obtained by integrating the currents (5.44):

$$v_S^{\pm}(t) = \frac{Q_S^{\pm}(t)}{C} = -\frac{1}{C}\int_0^t i_S^{\pm}(t')dt' = -\frac{Ne}{C}\left(\frac{t}{T^{\pm}} - \frac{1}{2}\left(\frac{t}{T^{\pm}}\right)^2\right). \tag{5.45}$$

The total voltage signal is then (assuming the usual case $T^- < T^+$):

$$\begin{aligned} v_S(t) &= v_S^-(t) + v_S^+(t) \\ &= \begin{cases} -\dfrac{Ne}{C}\left[\dfrac{t}{T^-} + \dfrac{t}{T^+} - \dfrac{1}{2}\left(\left(\dfrac{t}{T^-}\right)^2 + \left(\dfrac{t}{T^+}\right)^2\right)\right] & \text{for } 0<t<T^-, \\ -\dfrac{Ne}{C}\left[\dfrac{1}{2} + \dfrac{t}{T^+} - \dfrac{1}{2}\left(\dfrac{t}{T^+}\right)^2\right] & \text{for } T^-<t<T^+, \\ -\dfrac{Ne}{C} & \text{for } t>T^+. \end{cases} \end{aligned} \tag{5.46}$$

Initially the increase of the pulse with time is completely determined by the electrons because of the about 1000 times slower ion movement. After the electron collection time $T^- = d/v_D^-$ the pulse develops with a much slower increase. In time-critical applications often only the fast part of the pulse is utilised. From (5.44) one finds that the total charge signal of the electrons on the readout electrode is

$$Q_{tot}^- = -\frac{Ne}{2}, \tag{5.47}$$

thus about half of the total electron charge; the other half is due to the ion signal. This holds independently of the angle under which a particle track crosses the detector. An inclined track would however increase the number N of electron–ion pairs because of the longer path length.

Example. A *sampling calorimeter* (see chapter 15) is composed of alternating layers of metal plates in which particles generate showers and of 'active' layers for the de-

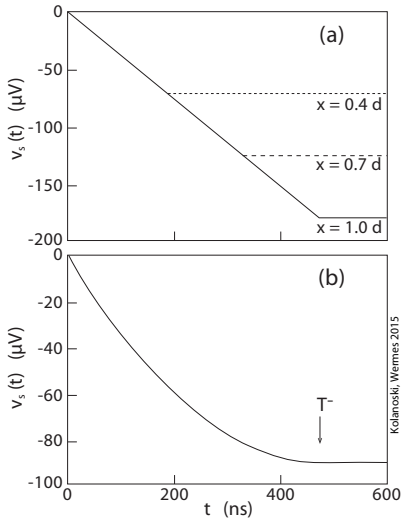


Fig. 5.7 Signal formation on the electrodes of an ionisation chamber configured as a parallel plate capacitor (plate separation d) for (a) a point charge generated at the distance x from the anode and for (b) a line charge along a line perpendicular to the plates. The slope of the signal is in both cases dominated by the electron movement; on this time-scale ($< 1\,\mu$s) the ion signal shows a hardly noticeable contribution. The maximal signal voltage (180 μ V for the assumed 10 000 electron–ion pairs and a capacitance of 9 pF) will in (b) only be reached after a very long time (ms), far outside the drawn time range.

tection of the shower particles. A common choice is lead plates alternating with gaps filled with liquid argon.[6] A high voltage across the gaps sucks the ionisation charges generated by the shower particles to the plates. Hence such a calorimeter is operated as an ionisation chamber. The active medium is a liquid rather than a gas because of its higher density and thus higher signal yield. However, this does not change the principle features of signal formation as discussed here. To give an example, we consider typical parameters of a detector cell of such a calorimeter:

$$\begin{aligned} V_0 &= 1.5\,\mathrm{kV}, \quad v_D^+ \approx 15\,\mathrm{cm/s}, \quad v_D^- \approx 0.5\,\mathrm{cm/\mu s}\,, \\ d &= 2.35\,\mathrm{mm}, \quad A = 16\,\mathrm{cm}^2, \quad \epsilon = 1.5 \quad \Rightarrow C \approx 9\,\mathrm{pF}\,. \end{aligned}$$

The resulting electron and ion collection times are

$$T^- = \frac{d}{v_D^-} = 470\,\mathrm{ns}\,, \qquad T^+ = \frac{d}{v_D^+} = 15.7\,\mathrm{ms}\,. \tag{5.48}$$

Thus the two collection times are different by more than four orders of magnitude.

Figure 5.7 shows the signal evolution for different primary charge distributions. The distributions are (a) point-like at different distances from the electrodes and (b) along a line perpendicular to the plates, each with 10 000 electron–ion pairs. Because of the distinctly different drift times the signal shape is mostly determined by the electron movement. In both cases the signal rises over the full ion collection time to the same value ($\approx$ 180 μV).

As already mentioned, for time-critical applications often only the electron signal is used. For use in a fast trigger, particularly at high event rates as at the LHC, the signal can be differentiated by a high-pass filter with a suitable time constant. In this way pulse widths of some 10 ns at still acceptable charge resolutions can be reached.

[6] This calorimeter type was realised, for example in the ATLAS detector [4] at LHC and in the H1 detector [34] at the HERA collider.

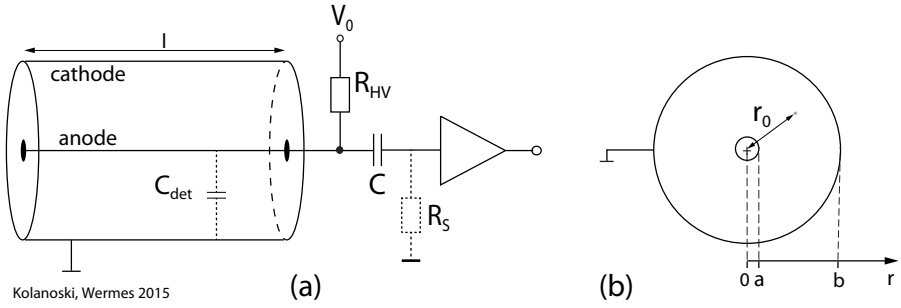


Fig. 5.8 Principle of a proportional counter tube with an anode wire on the axis of a cylindrical cathode: (a) side view, (b) cross-sectional view. The charge is generated at the radial distance r_0 from the axis. The detector has the capacitance C_{det} including all parasitic capacitances like those due to the wiring. R_{HV} is a high-ohmic resistor in series with the high-voltage power supply (HV). In order to decouple the signal electrode from the high voltage the signal is AC-coupled to an amplifier by a capacitance C which is large compared to C_{det}. The signal amplifier is here assumed to be an ideal operational amplifier with infinitely high input impedance (see also section 17.2). Hence the actual input impedance is given by the resistor R_S. This circuit corresponds to a high-pass filter differentiating the detector signal with the time constant $\tau = R_S\, C_{det}$.

5.3.2 Signal formation in cylinder-symmetric electrical fields

The signal formation in a wire chamber with gas amplification (chapter 7) is significantly different from that in a parallel plate assembly. The movement of the primary charge carriers to the wire (anode) or to the cathode, respectively, can be neglected as a contribution to the generation of the signal compared to the movement of charges created in the amplification process. Due to the strongly rising electric field near the wire the amplification only starts very close to the wire. Since the electrons therefore have a very short drift path to the wire they contribute little to the signal. In this case the signal is predominantly generated by the ions because in the high field near the wire they reach high drift velocities comparable to those of electrons (for ions the drift velocity is proportional to the field, see chapter 4).

As a typical representative of wire chamber geometries we consider a proportional counter tube (section 7.2.2) to demonstrate important aspects of signal formation in a cylindrical geometry. The counter tube has an anode wire with radius a, surrounded by a cylindrical cathode with radius b, and a length l which should be large compared to the radii (fig. 5.8).

Weighting field. In section 7.2.2, field strength and potential for an applied voltage V_0 as well as the capacitance per length are given by (7.5) and (7.6) for this configuration:

$$\vec{E}(r) = \frac{1}{r}\frac{V_0}{\ln b/a}\frac{\vec{r}}{r}\,, \qquad \phi(r) = -V_0\frac{\ln r/b}{\ln b/a}\,, \qquad C_l = \frac{2\pi\epsilon_0}{\ln b/a}\,. \tag{5.49}$$

The radius vector $\vec{r}$ is perpendicular to the cylinder axis with $r = |\vec{r}|$ being the radial distance to the cylinder axis in this direction. This yields the weighting field and potential for the computation of signals on the wire by setting $V_0 = 1$ while keeping the cathode on ground:

$$\vec{E}_w(r) = \frac{1}{r}\frac{1}{\ln b/a}\frac{\vec{r}}{r}\,, \qquad \phi_w(r) = -\frac{\ln r/b}{\ln b/a}\,. \tag{5.50}$$

The potential fulfils the boundary conditions

$$\phi_w(a) = 1\,, \quad \phi_w(b) = 0\,. \tag{5.51}$$

Employing the Shockley–Ramo theorem (5.26),

$$dQ_S = -q\vec{E}_w d\vec{r} \qquad \text{or} \qquad i_S = -\frac{dQ_S}{dt} = q\vec{E}_w\frac{d\vec{r}}{dt}\,, \tag{5.52}$$

we can compute the induced signals on the readout electrode (the wire).

Induced charges. We assume that at a distance r_0 from the wire N electron–ion pairs are generated which are separated in the field of the chamber. The whole charge which is induced on the wire contains contributions from the movement of the electrons from r_0 to a and of the ions from r_0 to b, respectively:

$$Q_S^- = -(-Ne)\frac{1}{\ln b/a}\int_{r_0}^{a}\frac{1}{r}dr = -Ne\frac{\ln r_0/a}{\ln b/a}\,, \tag{5.53}$$

$$Q_S^+ = -(+Ne)\frac{1}{\ln b/a}\int_{r_0}^{b}\frac{1}{r}dr = -Ne\frac{\ln b/r_0}{\ln b/a}\,. \tag{5.54}$$

The sum of both charge contributions,

$$Q_S^{tot} = Q_S^- + Q_S^+ = -Ne\,, \tag{5.55}$$

is the charge which flows from the anode to the voltage source (the same charge, but with reversed sign, flows from the cathode to ground). The total charge is independent of the position r_0 where the primary charge pair was created. However, in contrast to an arrangement with a homogeneous field as discussed in the previous section, the contributions of the electron and ion movements are in general very different and dependent on r_0 because of the $1/r$ behaviour of the field. The r_0 dependence of the ratio of the contributions is

$$\left(\frac{Q_S^-}{Q_S^+}\right)_{r_0} = \frac{\ln r_0/a}{\ln b/r_0}\,. \tag{5.56}$$

It is instructive to consider two special cases: (a) charge generation at a large distance from the wire ($r_0 \gg a$) and (b) very close to the wire where the amplification sets in. For typical parameters of a wire chamber with wire radius $a = 10\,\mu\text{m}$, cathode distance $b = 10\,\text{mm}$ and for example $r_0 = b/2$ the ratio is

$$\left(\frac{Q_S^-}{Q_S^+}\right)_{r_0=b/2} \approx 9\,. \tag{5.57}$$

Hence in this case the electron component determines the signal because the electrons travel through a region with much larger weighting potential difference (see eq. (5.27)) than the ions.

In detectors with gas amplification, however, the charge movement in the drift space practically does not play a role for the signal. Here it is essential that the overwhelming

part of the charge is generated in an avalanche very close to the wire. In fact, about half of the total charge is generated over the last interaction length, λ_{ion}. Therefore we approximate the distance of the charge production from the wire by $\epsilon \approx \lambda_{ion} \approx 1\,\mu\text{m}$ ($r_0 = a + \epsilon$) yielding an estimate of the ratio (5.56):

$$\left(\frac{Q_S^-}{Q_S^+}\right)_{r_0=a+\epsilon} \approx 0.01 - 0.02\,. \tag{5.58}$$

Hence, in detectors with gas amplification, the integrated charge signal is dominated by the contribution from the ions (generated in the avalanche).

Time evolution. Because of the dominance of the ion signal we first consider the contribution of the ion movement to the time evolution of the current signal. The ions start near the anode at $r_0 \approx a$ where the field strength and hence the initial velocity of the ions is very high. According to the Shockley–Ramo theorem (5.26) the time dependence of the induced current is

$$i_S^+(t) = Ne\,\frac{1}{\ln(b/a)}\frac{1}{r}v_D^+\,. \tag{5.59}$$

The radial dependence $r(t)$ is obtained from the drift motion of the ions:

$$\frac{dr}{dt} = v_D^+ = \mu^+ E(r) = \frac{\mu^+ V_0}{\ln(b/a)}\frac{1}{r}\,, \tag{5.60}$$

where μ^+ is the mobility of the ions in the gas (see section 4.3, eq. (4.45)). The integration of this differential equation yields (with $r_0 = r(t=0)$):

$$\begin{aligned} &\int_{r_0}^{r(t)} r\,dr = \frac{1}{2}\left(r^2(t) - r_0^2\right) = \frac{\mu^+ V_0}{\ln b/a}\,t \\ &\Rightarrow \quad r(t) = \sqrt{r_0^2 + \frac{2\mu^+ V_0}{\ln b/a}\,t} = r_0\sqrt{1 + \frac{t}{t_0^+}}\,, \end{aligned} \tag{5.61}$$

with a characteristic time t_0^+ defined as

$$t_0^+ = \frac{r_0^2 \ln b/a}{2\mu^+ V_0}\,. \tag{5.62}$$

This parameter roughly corresponds to the rise time of the ion signal at small t. The time an ion needs to arrive at the cathode is

$$T^+ = t(r=b) = t_0^+\,\frac{b^2 - r_0^2}{r_0^2}\,. \tag{5.63}$$

The result $r(t) \propto \sqrt{t_0^+ + t}$ in (5.61) shows that the ion movement is fast at the beginning ($t \lesssim t_0^+$) and slows down with increasing distance from the wire because of the strongly falling electric field. Inserting (5.60) and (5.61) into (5.59) yields the explicit time dependence of the ion signal:

$$i_S^+(t) = \frac{Ne}{2\ln b/a}\,\frac{1}{t + t_0^+}\,. \tag{5.64}$$

As opposed to the ions the electrons travel only a very short distance (few micrometres) after the onset of gas amplification and thus generate only an extremely short signal. If one assumes for the electrons in the gas amplification region an average free path of $\lambda_{ion} \approx 1\,\mu\text{m}$ and a constant velocity of about $v_D^- \approx 5\,\text{m}/\mu\text{s}$ (see section 7.4.1) the typical time until the arrival of the electrons at the wire is

$$T^- = (r_0 - a)/v_D^- \approx \frac{1\,\mu\text{m}}{5\,\text{m}}\,\mu\text{s} = 0.2 \times 10^{-12}\,\text{s}\,. \tag{5.65}$$

This intrinsically high time resolution can hardly be exploited in drift chambers because the primary ionisation is spatially distributed, yielding an arrival time smearing in the order of nanoseconds. Since the electrons contribute only about 1% of the ion signal, as estimated above, we will focus in the following on the ion signal.

The time dependence of the charge signal is obtained by integration of (5.64):

$$Q_S(t) \approx Q_S^+(t) = -\frac{N\,e}{2\ln b/a}\int_0^t \frac{dt'}{t' + t_0^+} = -\frac{Ne}{2\ln b/a}\,\ln\left(1 + \frac{t}{t_0^+}\right). \tag{5.66}$$

The induced charge will be accumulated on the electrode if the voltage source is decoupled by a high-ohmic resistor yielding a voltage signal with a rise time determined by the ion drift (see footnote on page 138) and a height depending on the capacitance of the detector $C_{det} = C_l\,l$ (see (5.49)):

$$v_s(t) = \frac{Q_S(t)}{C_{det}} = -v_0 \ln\left(1 + \frac{t}{t_0^+}\right) \quad \text{with} \quad v_0 = \frac{Ne}{2\ln b/a\,C_{det}} = \frac{Ne}{4\pi\epsilon_0 l}. \tag{5.67}$$

This means that the peak voltage only depends on the total charge and on the length of the counting tube, but not on radial dimensions (provided that additional stray capacitances, caused e.g. by the connection to the signal readout, can be neglected).

As a quantitative example we discuss the voltage signal of a proportional counter tube with typical parameters:

$$\begin{aligned} &a = 10\,\mu\text{m} \quad b = 10\,\text{mm}, \;\; l = 0.2\,\text{m} \quad \Rightarrow \quad C_{det} = 1.61\,\text{pF}, \\ &V_0 = 2\,\text{kV}, \quad \mu^+(\text{argon}) = 1.7\,\text{cm}^2\,\text{V}^{-1}\text{s}^{-1}, \\ &G = 10^4, \quad q = 100\,e^- \times G = 1.6 \times 10^{-13}\,\text{C}. \end{aligned}$$

In addition to the already introduced parameters the gas amplification G is also listed. In fig. 5.9(a) the curve with the label '$\tau = \infty$' presents the voltage signal according to (5.67) for these parameters. In this example the characteristic times are

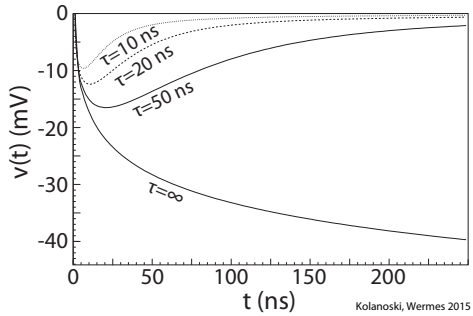
$$T^+ = 1\,\text{ms} \quad \text{and} \quad t_0^+ = 1\,\text{ns}\,.$$

Hence the total charge collection time is typically about 10^6 times larger than the characteristic signal rise time. The maximal voltage signal is reached after the time T^+ yielding

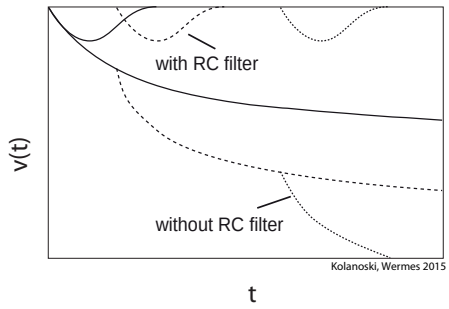
$$v_S(T^+) = -\frac{Ne}{2\ln b/a\,C_{det}}\ln\left(1 + \frac{T^+}{t_0^+}\right) \approx -100\,\text{mV}\,.$$

However, because of the steep voltage rise at small times, due to the high ion velocities near the anode, nearly 20% of the voltage is already reached after a time of $10\,t_0^+ \approx 10\,\text{ns}$:

$$\frac{v_S(10\,t_0^+)}{v_S(T^+)} \approx \frac{\ln 11}{\ln 10^6} \approx 0.17\,.$$



(a) Signal evolution in a counting tube.



(b) Suppression of 'pile-up'.

Fig. 5.9 (a) Signal formation in a cylindrical drift chamber with gas amplification. The parameters are chosen as described in the text. The relevant time constants are $t_0^+ = 1\,\text{ns}$ (fast signal rise) and $T^+ = 1\,\text{ms}$ (collection time for the ion signal). The different pulse shapes show the influence of the differentiation of the signal with different times $\tau = R_S C_{det}$. (b) 'Pile-up' signals with and without RC filter. In the drawing it is assumed that the long ion tails have been cancelled as well ('ion tail cancellation') so that the pulses quickly reach the baseline (see section 17.3.5).

Differentiation of signals. To avoid signal pile-up in high rate experiments the signal is often cut off by differentiation employing a suitable high-pass filter, as depicted in fig. 5.8(a): The anode wire is connected to high voltage via a high-ohmic resistor R_{HV} (some MΩ). The voltage signal generated by the drifting charges rises at the input of the amplifier, which ideally has an infinitely high input impedance and is decoupled from the high voltage by the coupling capacitor C. The actual input impedance is given by R_S. A voltage change at the amplifier input leads to a current $i(t) = v(t)/R_S$ across R_S which damps the voltage signal with a time constant $\tau = R_S\, C_{det}$ as long as the coupling capacitance C is large compared to C_{det}:

$$v(t) = v_0\, \mathrm{e}^{-t/\tau}\,. \tag{5.68}$$

For a given detector capacitance $C_{det} = C_l\, l$ (here 1.61 pF) the time constant is determined by R_S. The further pulse shaping, including cancellation of the 'ion tail' by means of dedicated filter circuits, is described in section 17.3.

Each charge contribution at time t' contributes to the voltage signal with an exponential factor as in (5.68). Altogether the result is that the integrand (5.66) has to be weighted with an exponential function:[7]

$$v_S(t) = -v_0 \int_0^t \frac{\mathrm{e}^{-\frac{t-t'}{\tau}}}{t_0 + t'}\, dt' = -v_0 \mathrm{e}^{-\frac{t+t_0}{\tau}} \left[\mathrm{Ei}\left(\frac{t+t_0}{\tau}\right) - \mathrm{Ei}\left(\frac{t_0}{\tau}\right) \right], \tag{5.69}$$

[7] This result fully agrees with the more general approach for the calculation of signals on electrodes embedded in an electronic network using the concept of 'time-dependent weighting fields' as for example described in [449, 820], see also the footnote on page 138. This approach yields the voltage signal in Laplace space (with the Laplace variable s, see appendix H):

$$v(s) = -v_0 \frac{\tau}{t_0} \exp(s\, t_0)\, \frac{\mathrm{E}_1(s\, t_0)}{1 + s\, \tau}\,.$$

Here v_0 is defined as in (5.67) and E_1 is the exponential integral function for complex values of the argument. The inverse Laplace transform yields (5.69) as expected.

with v_0 defined as in (5.67). The function Ei(x) is the 'exponential integral' (see e.g. [255, 976]). Figure 5.9(a) shows the signal at small times for different values of the time constant $\tau = R_s C_{det}$. The curve for $R_s = \infty$, corresponding to $\tau = \infty$, depicts the signal development without differentiation. The exponential decay of the differentiated signal is flattened at larger times due to the Ei terms in (5.69), caused by the persistent ion current.

Figure 5.9(a) demonstrates that by differentiation and possible further pulse shaping on a time-scale of about 10 ns, only a fraction of the total charge will be measured, typically about 20% ('ballistic deficit', see also section 17.3.4). The loss can be accepted since in turn a better signal separation is achieved at high rates (fig. 5.9(b)) improving the resolution of closely succeeding hits from nearby particle tracks (double-track resolution). At high rates differentiation alone does not prevent a shift of the voltage baseline due to the very long ion signals. This can make the signal thresholds rate dependent (fig. 5.9(b)). With smart electronic circuits, called 'ion tail cancellation' and 'baseline restoration'), a baseline shift can be mostly suppressed (see section 17.3.5 or, for example, [799, 905]).

With such a differentiation of the signal the very fast electron signal fully contributes so that the signal fraction due to electrons can increase to about 5–10% as compared to the 1–2% fraction of the full charge, as calculated in (5.58).

5.4 Signal formation in detectors with space charge

5.4.1 Charge and current signal of an electron–hole pair in silicon

A semiconductor detector resembles a parallel plate detector, filled with a dielectric (see chapter 8). The difference, however, is that the electric field is not constant because of space charge in the depletion zone of the detector. The field decreases linearly from the pn junction side to the opposite side of the detector (fig. 5.10). In the following we refer to a silicon detector with a weakly n-doped substrate material on which a more strongly doped p-layer is applied. Starting from the pn junction the depletion layer grows with increasing inverse bias voltage into the detector. The n^+ layer on the opposite side serves as contact. Other arrangements and more details are described in chapter 8.

At full depletion the field reaches zero at the $n^- n^+$ boundary at the side opposite to the readout electrode of the detector. At incomplete depletion (underdepletion), the field ends at the conductive, (still) undepleted layer (fig. 5.10(a)). If the applied voltage is larger than required for full depletion a constant field component is superimposed on the linear component (overdepletion). The field is given by

$$\begin{aligned} \vec{E}(x) &= -\left[\frac{2V_{dep}}{d^2}(d-x) + \frac{V - V_{dep}}{d}\right]\vec{e}_x \\ &= -\left[\frac{V + V_{dep}}{d} - \frac{2V_{dep}}{d^2}x\right]\vec{e}_x\,, \end{aligned} \tag{5.70}$$

where V is the applied voltage and V_{dep} is the voltage necessary for full depletion. The voltage V_{dep} follows directly from (8.59) in section 8.3.1 and depends mainly on the doping density N_D and the thickness of the detector d:

$$V_{dep} \approx \frac{eN_D}{\epsilon\epsilon_0}\frac{d^2}{2}\,. \tag{5.71}$$

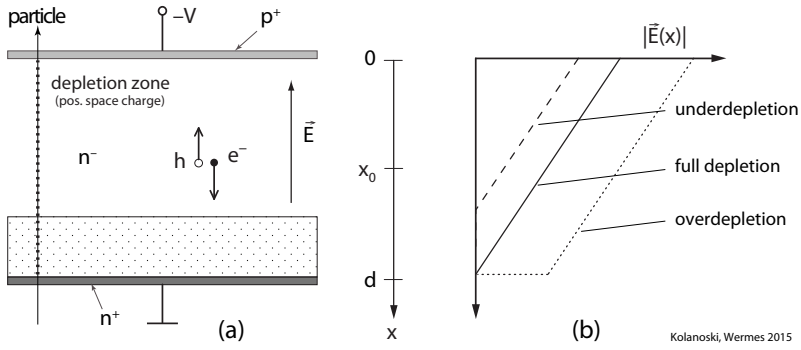


Fig. 5.10 Layout and electric field of a silicon detector. (a) Sketch of a silicon detector with incomplete depletion (underdepletion) and e/h pair generation at x_0; (b) linear behaviour of the electric field for full depletion as well as for over- and underdepletion.

At sufficiently large overvoltage the conditions approach more and more those of a parallel plate detector with constant field (see section 5.3.1). With the abbreviations

$$a = \frac{V + V_{dep}}{d}\,, \qquad b = \frac{2V_{dep}}{d^2}\,, \tag{5.72}$$

the electrical field reads

$$\vec{E}(x) = -\,(a - bx)\,\vec{e}_x\,. \tag{5.73}$$

We now consider an electron–hole pair (e/h) generated at $x = x_0$ (fig. 5.10(a)). The electrons drift in the direction opposite to the field and the holes in the direction of the field:

$$\begin{aligned} v_e &= -\mu_e E_x(x) = +\mu_e\,(a - bx) = +\frac{1}{\tau_e}\left(\frac{a}{b} - x\right) = \dot{x}_e\,, \\ v_h &= +\mu_h E_x(x) = -\mu_h\,(a - bx) = -\frac{1}{\tau_h}\left(\frac{a}{b} - x\right) = \dot{x}_h\,. \end{aligned} \tag{5.74}$$

The equations use the characteristic times for the charge carrier movements,

$$\tau_{e,h} = \frac{1}{\mu_{e,h}\,b} = \frac{d^2}{2\,\mu_{e,h}\,V_{dep}}\,. \tag{5.75}$$

The solutions of the differential equations (5.74) are

$$\begin{aligned} x_e(t) &= \frac{a}{b} - \left(\frac{a}{b} - x_0\right)\,\mathrm{e}^{-t/\tau_e}\,, \quad v_e = \dot{x}_e = \left(\frac{a}{b} - x_0\right)\,\frac{1}{\tau_e}\,\mathrm{e}^{-t/\tau_e}\,, \\ x_h(t) &= \frac{a}{b} - \left(\frac{a}{b} - x_0\right)\,\mathrm{e}^{+t/\tau_h}\,, \quad v_h = \dot{x}_h = -\left(\frac{a}{b} - x_0\right)\,\frac{1}{\tau_h}\,\mathrm{e}^{t/\tau_h}\,. \end{aligned} \tag{5.76}$$

Thus the result is an accelerated movement of the charge carriers. At high field strength, however, the drift velocity can saturate leading to conditions as in the case of parallel plates with constant drift velocity, as described in section 5.3.1.

At full depletion the electrons reach the electrode at $x_e = d$ at the time $t = T^-$. The holes reach the electrode at $x_h = 0$ at the time $t = T^+$. With (5.76) we obtain for these times:

$$T^- = \tau_e \ln \frac{a - bx_0}{a - b\,d}\,, \qquad T^+ = \tau_h \ln \frac{a}{a - bx_0}\,. \tag{5.77}$$

The collection times are finite only if the denominators are positive. With a, b from (5.72) this means for electrons:

$$a - b\,d \;>\; 0 \quad \Rightarrow \quad V > V_{dep}\,. \tag{5.78}$$

For smaller voltages V than the depletion voltage V_{dep} the electrons no longer reach the electrode. However, as long as the charges are created in a region of non-vanishing field, there are still signals induced at the readout electrode (top electrode in fig. 5.10(a)) due to the movement of both carriers, albeit smaller than at full depletion. Thus, in order to obtain full signal efficiency, the applied voltage should always be larger than the voltage needed for full depletion. In the following we use the upper electrode in fig. 5.10(a) as readout electrode (hole collection) for which we will compute the current signal. According to the Shockley–Ramo theorem (5.26) the current signal is (for the sign convention see section 5.2.4):

$$i_S^{e,h}(t) = q\vec{E}_w\vec{v}_{e,h}\,. \tag{5.79}$$

As argued in section 5.2.2 the weighting field is independent of stationary space charges and thus the same as in the parallel plate configuration in section 5.3.1. Hence for the upper electrode the weighting field is given by

$$\vec{E}_w = -\frac{1}{d}\,\vec{e}_x\,. \tag{5.80}$$

With that and the velocities in (5.76) we obtain the time evolution of the charge signal:

$$\begin{aligned} i_S^h(t) &= \frac{e}{d}\frac{1}{\tau_h}\left(\frac{a}{b} - x_0\right)\mathrm{e}^{+t/\tau_h} \qquad \text{for} \qquad t < T^+\,, \\ i_S^e(t) &= \frac{e}{d}\frac{1}{\tau_e}\left(\frac{a}{b} - x_0\right)\mathrm{e}^{-t/\tau_e} \qquad \text{for} \qquad t < T^-\,. \end{aligned} \tag{5.81}$$

The current signal ends when the charge stops drifting. Typical numerical values for electron and hole mobilities in semiconductors are found in table 8.2 in section 8.2.1. In order to obtain an estimate for the drift times we examine the characteristic times in (5.75):

$$\tau_{e,h} = \frac{d^2}{2\mu_{e,h}V_{dep}} \approx \frac{\epsilon\,\epsilon_0}{\mu_{e,h}eN_D}\,. \tag{5.82}$$

For silicon detectors with $\epsilon_{Si} = 11.9$ and typical doping concentrations $N_D = 10^{12}\,\mathrm{cm}^{-3}$ (see section 8.3.1), $\tau_e \approx 5\,\mathrm{ns}$ and $\tau_h \approx 15\,\mathrm{ns}$ are obtained. Applying a voltage $V = 1.5\,V_{dep}$ (overdepletion, to get finite drift times for the electrons) and assuming charge generation either in the centre of the detector or far away from the readout electrode one finds for the drift times:

$$\begin{aligned} x_0 = 0.5\,d \quad &\Rightarrow \quad T^+ = 0.5\,\tau_h = 7.7\,\mathrm{ns}\,, \\ x_0 = 0.95\,d \quad &\Rightarrow \quad T^+ = 1.4\,\tau_h = 21\,\mathrm{ns}\,, \\ x_0 = 0.5\,d \quad &\Rightarrow \quad T^- = 1.1\,\tau_e = 5.5\,\mathrm{ns}\,, \\ x_0 = 0.95\,d \quad &\Rightarrow \quad T^- = 0.2\,\tau_e = 1.0\,\mathrm{ns}\,. \end{aligned} \tag{5.83}$$

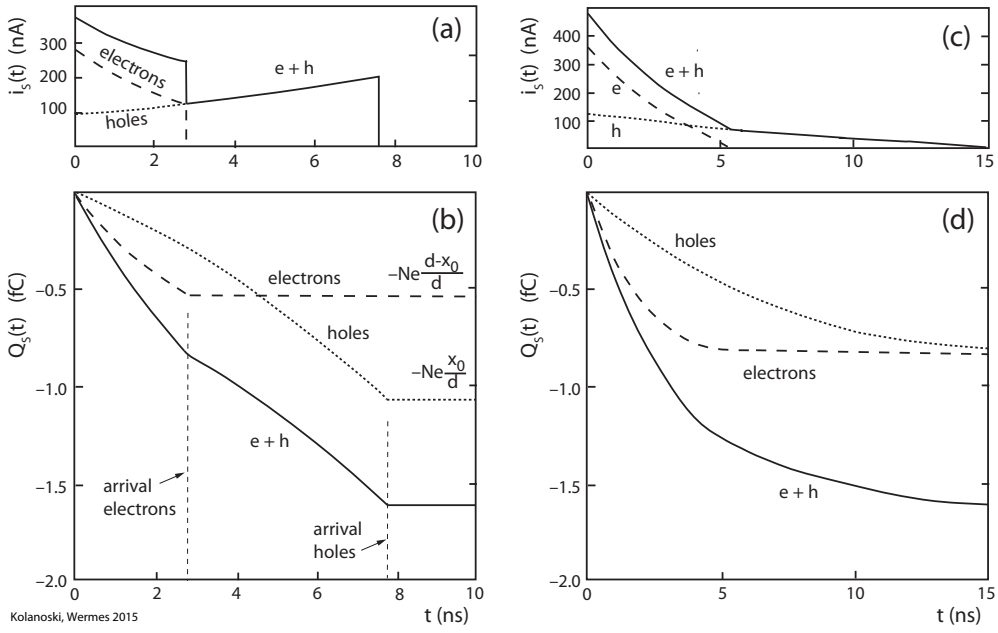


Fig. 5.11 Typical signal evolution for current $i_S(t)$ and charge $Q_S(t)$ for electrons and holes at the negative electrode of a silicon detector as depicted in fig. 5.10. The detector parameters are: $V = 150\,\mathrm{V}$, $V_{dep} = 100\,\mathrm{V}$, $d = 300\ \mu\mathrm{m}$. (a, b) Point charge with starting point $x_0 = \frac{2}{3}d$ for $N = 10\,000$ elementary charges; (c, d) charged particle with $N = 10\,000$ elementary charges, equally distributed along the track. The finite arrival time is ensured by 50% 'overdepletion'.

Thus silicon detectors have a relatively short collection time for electrons and holes and can be deployed as 'fast detectors'.

The sum of the electron and hole components yields the total current (fig. 5.11(a)):

$$\begin{aligned} i_S(t) &= i_S^e(t) + i_S^h(t) \\ &= \frac{e}{d}\left(\frac{a}{b} - x_0\right)\left(\frac{1}{\tau_e}\mathrm{e}^{-t/\tau_e}\,\Theta(T^- - t) + \frac{1}{\tau_h}\mathrm{e}^{t/\tau_h}\,\Theta(T^+ - t)\right), \end{aligned} \tag{5.84}$$

where Θ is the Heaviside step function.

By integrating the current signal one obtains the induced charge signal $Q_S(t)$ (fig. 5.11 (b)):

$$\begin{aligned} Q_S(t) = -e\,\frac{a - b\,x_0}{b\,d}\Big[&\left(1 - \mathrm{e}^{-t/\tau_e}\right)\Theta(T^- - t) + \left(\mathrm{e}^{\,t/\tau_h} - 1\right)\Theta(T^+ - t) \\ &+ \left(1 - \mathrm{e}^{-T^-/\tau_e}\right)\Theta(t - T^-) + \left(\mathrm{e}^{\,T^+/\tau_h} - 1\right)\Theta(t - T^+)\Big]\,. \end{aligned} \tag{5.85}$$

The equation can be written in a more compact form as long as none of the charges has reached the respective electrode:

$$Q_S(t) = -e\,\frac{a - b\,x_0}{b\,d}\left(\mathrm{e}^{\,t/\tau_h} - \mathrm{e}^{-t/\tau_e}\right) \quad \text{for} \quad t < T^-,\, T^+\,. \tag{5.86}$$

The first line in (5.85) shows the growth of the induced charge as long as the

charges move. The second line contains the totally induced charges after the respective collection times. Using the expressions for $T^{\pm}$ in (5.77) the total charges can be written as follows:

$$Q_S^{tot-} = -e\,\frac{a - b\,x_0}{b\,d}\left(1 - \mathrm{e}^{-T^-/\tau_e}\right) = -e\,\frac{d - x_0}{d}\,, \tag{5.87}$$

$$Q_S^{tot+} = -e\,\frac{a - b\,x_0}{b\,d}\left(\mathrm{e}^{T^+/\tau_h} - 1\right) = -e\,\frac{x_0}{d}\,. \tag{5.88}$$

This yields the total signal charge originating from the movement of both charge carriers:

$$Q_S^{tot} = Q_S^{tot-} + Q_S^{tot+} = -e\,. \tag{5.89}$$

With the detector capacitance C the charge signal can be converted into a voltage signal,

$$v_S(t) = Q_S(t)/C\,, \tag{5.90}$$

which may be utilised for further electronic processing.

5.4.2 Signal shape for the passage of a particle

Computing the signal of an ionising particle with N e/h pairs equally distributed along its track (fig. 5.10(a)) the contributions of the individual starting points x_0 have to be superposed. Neglecting diffusion and applying (5.81) for the contribution of an individual charge, the current signal of a particle track can be calculated. For a given time t the currents (5.81) of the individual charges have to be integrated over all those generation points x_0 from which the charge did not yet arrive at the electrode at the time t:

$$\begin{aligned}
i_S^-(t) &= -\tfrac{Ne}{d^2}\tfrac{1}{\tau_e}\,\mathrm{e}^{-\frac{t}{\tau_e}} \underbrace{\int_0^{x^e_{max}} \left(\frac{a}{b} - x\right) dx}_{\frac{a}{b}x^e_{max} - \frac{1}{2}x^{e\,2}_{max}} && \text{for}\quad 0 < t < T^-_{max}\,,\\
i_S^-(t) &= 0 && \text{else}\,,\\
i_S^+(t) &= -\tfrac{Ne}{d^2}\tfrac{1}{\tau_h}\,\mathrm{e}^{+\frac{t}{\tau_h}} \underbrace{\int_{x^h_{min}}^{d} \left(\frac{a}{b} - x\right) dx}_{\frac{a}{b}\left(d - x^h_{min}\right) - \frac{1}{2}\left(d^2 - x^{h\,2}_{min}\right)} && \text{for}\quad 0 < t < T^+_{max}\,,\\
i_S^+(t) &= 0 && \text{else}\,.
\end{aligned} \tag{5.91}$$

Each current flows for up to the maximal drift time that is needed to pass the detector thickness according to (5.77):

$$\begin{aligned}
T^-_{max} &= T^-(x_0 = 0) = \tau_e \ln\frac{a}{a - b\,d}\,,\\
T^+_{max} &= T^+(x_0 = d) = \tau_h \ln\frac{a}{a - b\,d}\,.
\end{aligned} \tag{5.92}$$

Charges starting at the integral boundaries x^e_{max} and x^h_{min} in (5.91) reach their respective electrode at the time t at which the coordinates are $x_e(t) = d$ for $x_0 = x^e_{max}$ and $x_h(t) = 0$ for $x_0 = x^h_{min}$ in (5.76). Solving the equations for x_0 yields

$$x^e_{max} = \frac{a}{b} + \left(d - \frac{a}{b}\right)\mathrm{e}^{+t/\tau_e}, \qquad x^h_{min} = \frac{a}{b}\left(1 - \mathrm{e}^{-t/\tau_h}\right). \tag{5.93}$$

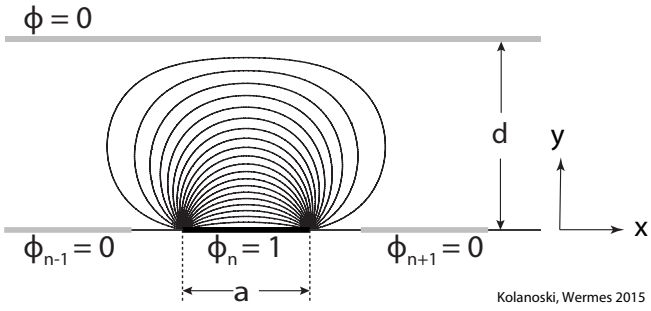


Fig. 5.12 Schematic representation of a cut through a strip detector perpendicular to the strips. The weighting potential $\phi_w(x, y)$ for the nth strip is plotted as computed according to (5.94).

Inserting (5.93) into (5.91) yields the current signal of a track passing perpendicularly through the detector and, after integration over time, the charge or voltage signal. In fig. 5.11(c, d) the result is presented for a particle with $N = 10\,000$ e/h pairs along its track.

5.5 Signal formation in detectors with segmented electrodes

The previous discussions apply for two-electrode configurations with unsegmented electrodes. However, often a position sensitive readout is desired which can be achieved by subdividing the electrodes into smaller elements. Common examples are strip and pixel detectors. We condider here a multi-electrode system with one-dimensional subdivision into strips. The conclusions, however, to a large extent also hold for pixel detectors with a two-dimensional subdivision (see e.g. [823]).

We assume that the strip width a is large compared to the distance between strips. The strip length should be large relative to their width and the thickness of the detector so that the problem can be treated as depicted in fig. 5.12. The figure shows the weighting potential, delineated by lines of constant potential, for one of the strips, taken as the readout electrode. With the above assumptions the weighting potential can be computed by solving the homogeneous potential equation, $\Delta\phi_w(x, y) = 0$, with the respective boundary conditions on the electrodes ($\phi_w = 1$ on the readout strip, $\phi_w = 0$ else):

$$\phi_w(x, y) = \frac{1}{\pi} \arctan \frac{\sin(\pi y)\, \sinh(\pi \frac{a}{2})}{\cosh(\pi x) - \cos(\pi y) \cosh(\pi \frac{a}{2})} \,. \tag{5.94}$$

Here x, y are the coordinates in the sectional plane, x lies in the plane with $x = 0$ in the middle of the considered strip and y perpendicular to it such that the electrode planes lie at $y = 0$ and $y = 1$ (thus $d = 1$). This solution of the potential equation can be found, for example, by means of conformal mapping,[8] as shown in appendix B. Evaluating the potential numerically one has to ensure that the arctan function is mapped onto the domain $[0, \pi]$ by accounting for the signs of numerator and denominator separately.

In fig. 5.13(a) the weighting potential is plotted for different strip widths as a function of the distance y from the electrode. For a very large strip width ($a \to \infty$) the configuration approaches the geometry of a parallel plate detector with the weighting potential depending almost linearly on the distance from the electrode. For

[8] Usually the computation of complex field configurations is performed numerically with the aid of special programs (e.g. Garfield [959] or FlexPDE [415]).

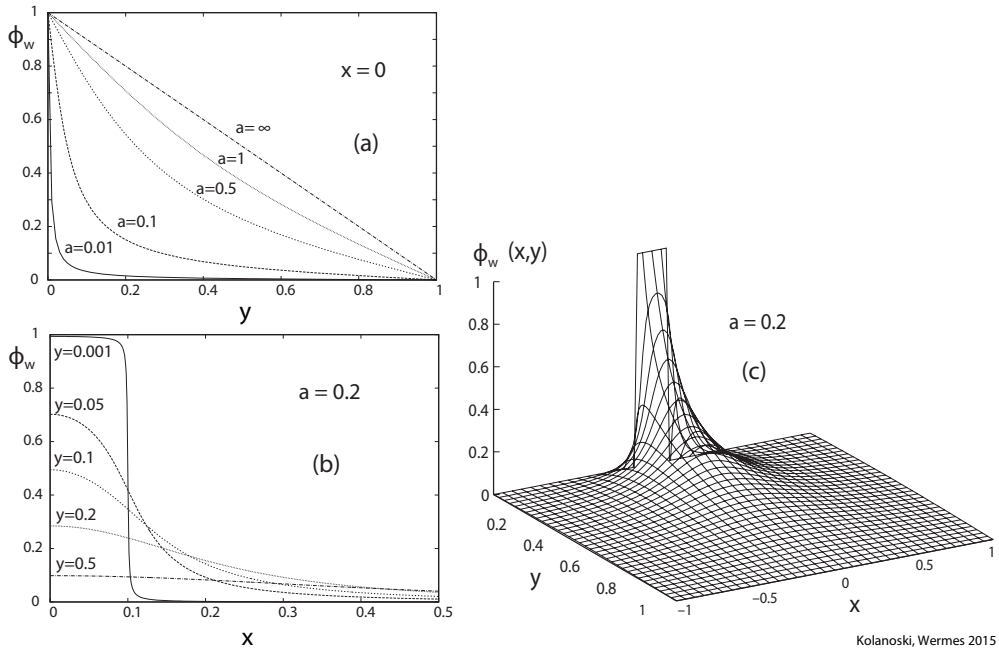


Fig. 5.13 Plots of the weighting potential ϕ_w according to (5.94): (a) ϕ_w as a function of the distance y from the strip plane at $x = 0$ (strip centre) for different strip widths a; (b) ϕ_w as a function of the distance x from the strip centre for different distances y from the strip plane at fixed strip width $a = 0.2$; (c) ϕ_w in two-dimensional representation at fixed strip width $a = 0.2$. All lengths values are given in units of the detector thickness d.

small strips ($a \ll 1$), in contrast, the weighting potential has a steep slope near the electrode. This means, according to the Shockley–Ramo theorem, that the strongest signal contributions result from the charge movement near the readout strips or pixels. Therefore this effect is called the 'small pixel effect'.

Figure 5.13(b) displays the potential for $a = 0.2$ (e.g. 60 µm for a 300 µm thick detector) as a function of the horizontal distance x from the strip centre at different depths y; in fig. 5.13(c) the same is displayed two-dimensionally as a function of x and y. With sufficient distance to the strip plane, the weighting potentials for the 'signal electrode' where the drifting charge is collected and its neighbour electrodes are very similar. In the middle of the detector, at $y = 0.5$, the potential is nearly independent of the x position, as can be seen in fig. 5.13(b). A charge drifting towards the segmented electrode (strips or pixels) initially induces a broad surface charge distribution, causing similar signals on several neighbouring strips. Upon further movement of the charge towards the strip electrode the induced signal further increases on the signal electrode while it decreases on the neighbouring electrodes. For electrodes whose dimensions are small compared to the distance of the drifting charge from the electrode, the signal development is the same on all strips (or pixels) for most of the time the charge moves. It differs for the signal electrode and its neighbours only shortly before the drift ends.

The implications of the 'small pixel effect' have to be taken into account in the design of a detector. This is particularly true if electron and hole mobilities are very different ($\mu_h < \mu_e$) as for example for the semiconductors CdTe or CdZnTe, or if the

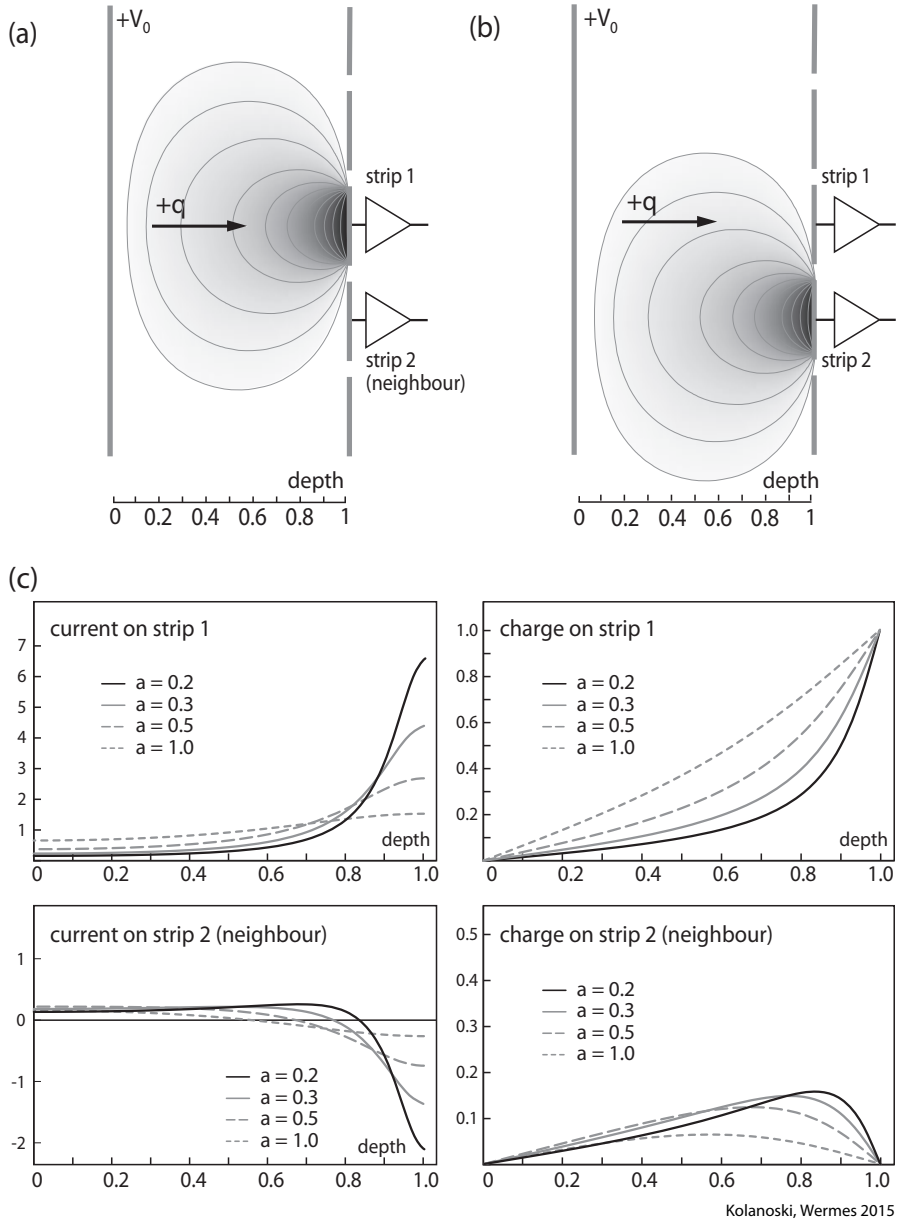


Fig. 5.14 Weighting potentials (a, b) and resulting current and charge signals (c) for a strip or pixel detector at different values of the strip width a (strip width and depth coordinate are given in units of the detector thickness d). Strip 1 is the signal electrode where the charge is collected, strip 2 is a neighbouring electrode. In (a) the weighting potential is shown for the signal electrode (strip 1) and in (b) for the neighbouring electrode. A charge moving towards strip 1 generates a monotonically increasing current signal (c). The charge signal increase as well until it reaches its maximum at arrival. For the neighbouring electrode the current signal crosses the zero line and the charge signal vanishes at arrival of the drifting charge on the signal electrode.

charge carriers can be trapped (see section 8.12). Let us consider an example where the holes are very immobile and the electrons are trapped after a short distance. In such a case the local absorption of a photon, at a distance from the readout electrode which is large compared to the strip width, generates only a small signal which is evenly distributed over several electrodes and thus is not very sensitive to the position.

Figure 5.14 presents the two situations where the charge drifts towards the measuring electrode or towards a neighbouring electrode, respectively. According to the Shockley–Ramo theorem (5.26) the current signal always has the same sign if the charge moves towards the measuring electrode since the product of the velocity and the weighting field, $\vec{E}_w \vec{v}_D$, does not change the sign. In this case the integral over the current yields the value of the moving charge, $\pm q$. In contrast, in the region of the neighbour electrode the product $\vec{E}_w \vec{v}_D$ changes sign, and thus also the current, yielding a vanishing integral over the current corresponding to the fact that no charge arrives at that electrode.

At the beginning of the charge movement the neighbouring electrodes see nearly as much signal as the electrode on which the charge arrives. If the electronic readout cuts the signal off after short drift times, for example to cope with high rates, several neighbouring strips (or pixels) could have a similar signal—an effect which has the same impact on a measurement as cross talk between channels.

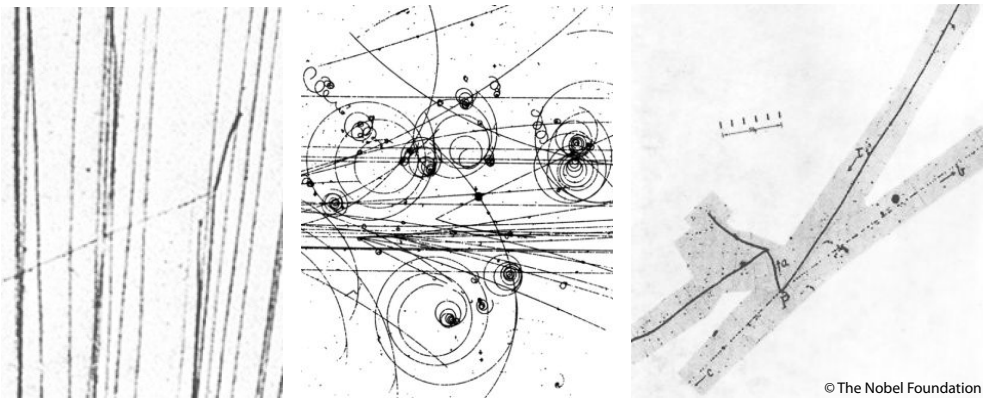
6
Non-electronic detectors

The first detectors which made trajectories of ionising particles visible were cloud chambers, photoemulsions and bubble chambers. All of these 'classical' detectors can not be read electronically. Today, cloud and bubble chambers have no or little relevance anymore since the relatively cumbersome data acquisition and data processing for these detectors is no longer competitive given the fast progressing developments in electronics and its application to detectors. Only photoemulsions, despite their laborious evaluation procedures, are still employed in modern experiments if spatial resolutions in the micrometre range are requested where emulsions are still unrivalled. Today cloud chambers are often used for demonstration purposes, because they instructively visualise particle tracks from radioactivity and cosmic rays. They are also relatively easy to construct and operate. Bubble chamber pictures are also often employed for demonstrations of whole reaction sequences and typical event topologies. A guidance for the understanding of bubble chamber pictures is offered by CERN, see [284]. For the visualisation of events in this book we will also resort to pictures from cloud chambers, bubble chambers and photoemulsions, as for example in fig. 6.1 (see also figs. 2.3, 2.4, 14.1, 15.9 and 15.26).

Historically these detectors played a decisive role for the development of nuclear and particle physics, see section 2.1. Until the beginning of the 1950s, particle physics was explored employing cloud chambers and photographic plates; afterwards, until the 1980s, bubble chambers played a prominent role for research at accelerators. For a detailed description of these classical detectors one has to go back to older literature, for example the contributions in the Encyclopedia of Physics 'Nuclear Instrumentation II' from 1958 [330]. The newer developments on the subject of photoemulsions are described in the volume on 'Detectors for Particles and Radiation' of the Landolt–Börnstein series [345].

6.1 Cloud chamber

The cloud chamber, introduced 1912 by C.T.R. Wilson (Nobel Prize 1926), was the first instrument with which particle tracks could be made visible and thus the particle kinematics could be analysed. The principle relies on making the microscopic ionisation charges, which a charged particle generates in a gas along its trajectory (section 3.2), macroscopically visible. When a gas, saturated with water vapour, is in a supercritical state, the generated ions form condensation nuclei such that the particle track becomes visible via the condensing droplets (see fig. 6.1(a) and also figs. 2.3 and 2.7).



(a) Claud chamber. (b) Bubble chamber. (c) Nuclear emulsion.

Fig. 6.1 Visualisation of particle reactions by means of the detectors which are described in this chapter. (a) Detail of a cloud chamber picture of a nuclear reaction [208]. An α particle reacts with a nitrogen nucleus (in the filling gas of the chamber), the intermediate state decays into a proton (thin line going to the left) and the oxygen isotope ^{17}O (thick track going to the right). (b) Bubble chamber reaction. Interaction of a pion beam (from left) in a bubble chamber (BEBC, see section 6.2) filled with hydrogen. One sees the straight-lined beam particles, primary vertices at which many particles are generated and secondary vertices of decaying particles. Source: CERN. (c) Interaction of a secondary particle of the cosmic radiation recorded by nuclear emulsion [793]. The picture shows a charged kaon (then still called τ) that enters from the top right, comes to rest at the point P and then decays into three charged pions. One of the pions makes a further interaction in the emulsion. Source: The Nobel Foundation.

6.1.1 Expansion cloud chamber

In Wilson's cloud chamber the supercritical state is reached by cooling of the gas volume through adiabatic expansion. The principle of the apparatus is shown in fig. 6.2(a). A usually cylindrical gas volume is rapidly expanded by means of a piston movement. The droplets which form along ionising tracks are illuminated from the side and photographed through a pressure-resistant glass pane which closes the cylinder on the side opposite to the piston. In order to study nuclear radiation, radioactive sources can be directly inserted into the gas volume (e.g. an α radiator as in fig. 3.21 on page 51). Usually a magnetic field is superimposed perpendicular to the glass pane, so that the track curvature lies in the plane of the photograph, as for example in fig. 2.3 on page 6.

At the beginning, cloud chambers were expanded in a random time sequence, with the result that only a few per cent of the pictures contained interesting events. Blackett and Occhialini reached an essential improvement of the data acquisition efficiency by triggering the camera upon a coincidence of pulses from Geiger counters [209, 210]. Figure 6.2(b) shows a schematic sketch of an assembly used for studying cosmic radiation [210]. The diffusion of the ions during the time between the particle passage and the recording of the photograph after the termination of the expansion determines the spatial resolution. A minimum of 10–20 ms is quoted in [210] for this time interval which, however, was also selectively varied to make individual ion clusters visible. The diffusion becomes smaller with either increasing pressure or larger mass of the ions,

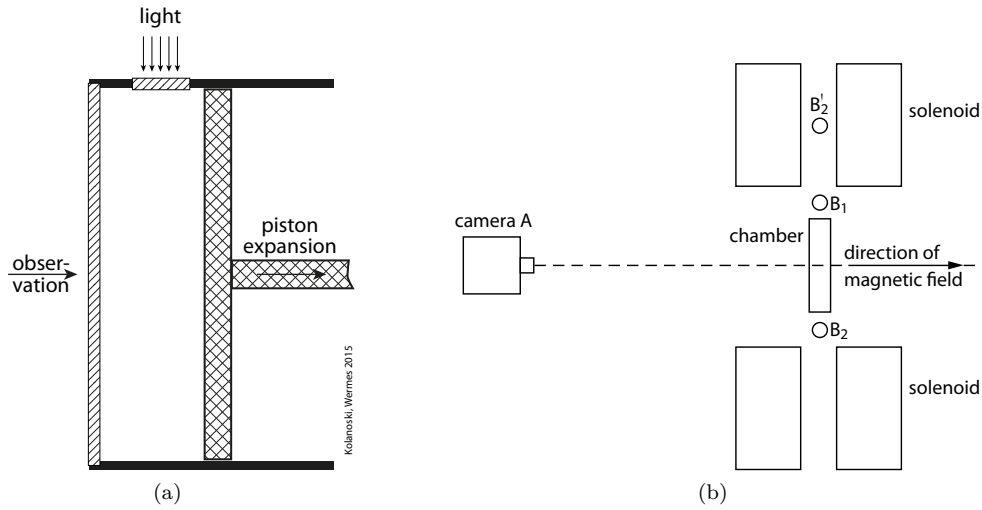


Fig. 6.2 (a) Functional principle of a cloud chamber; explanations are given in the text. (b) Arrangement for the observation of cosmic radiation with a cloud chamber (adapted from [210]). The cloud chamber has the shape of a disk with a sensitive volume of 13 cm diameter and 3 cm depth. The chamber is set up vertically in order to allow the observer, here a camera, to look perpendicular to the preferential direction of the cosmic radiation. Three Geiger–Müller tubes (B_1, B_2, B_2'; diameter 2 cm) are positioned above and below the chamber. A coincident signal of the tubes triggers the expansion and the camera. A magnetic solenoid field deflects the particles in the plane perpendicular to the camera view.

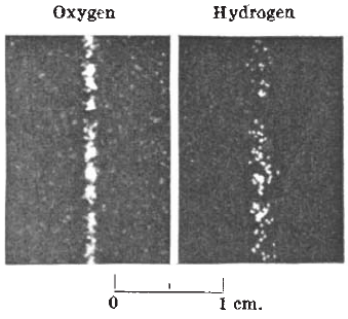


Fig. 6.3 Cloud chamber tracks in oxygen (left) and in hydrogen (right) [209]. The lighter hydrogen ions diffuse more strongly than the oxygen ions. Reprinted with kind permission of the Nature Publishing Group.

as can be inferred from eq. (4.90). In fig. 6.3 tracks in oxygen and in hydrogen are compared under the same conditions [209].

In principle, a cloud chamber picture can contain all information necessary for kinematic reconstruction and particle identification. As a rule the pictures are taken with two cameras and the kinematics is stereoscopically reconstructed,[1] yielding momentum and direction of the particles. By counting the number of ionisation clusters (and possibly of δ electrons) the velocity can be determined and therefore in principle

[1]It is interesting how tracks were stereoscopically reconstructed without the help of a computer [996, 210]. For this purpose the developed photographic plates were inserted again into the cameras, illuminated from the back and thus back-projected through the camera optics. A mechanical track model could then be aligned with the intersection points of the projection rays. In principle, this stereoscopic back-projection is a deconvolution algorithm as implemented today in computer programs which generate tomographic pictures.

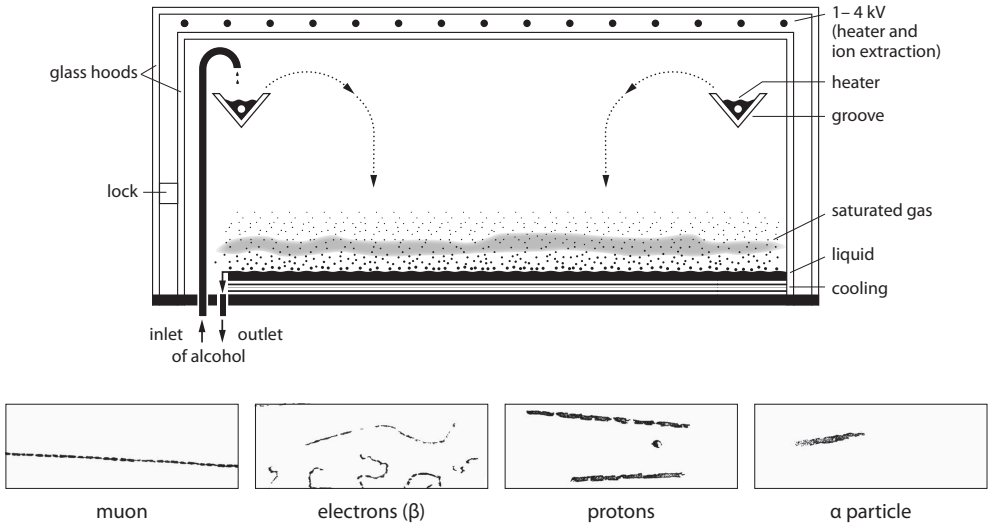


Fig. 6.4 Diffusion cloud chamber. Top: Schematic sketch, see explanations in the text. Bottom: Samples of tracks from different particle species. Source: DESY.

also the mass if the momentum is measured (see section 14.2.2). In these early times electrons, protons, nuclei and 'mesons', which are particles with masses between those of electrons/positrons and protons, could thus be distinguished.

The energy of relatively slow particles can be determined by measuring the range if they stop in the chamber, as demonstrated using the example of α rays in fig. 3.21 in section 3.2.4. By inserting absorber plates into a cloud chamber (figs. 2.3 and 2.7) characteristic signatures of particle species can in addition be exploited, like the ability to penetrate through the plates or to develop a shower.

6.1.2 Diffusion cloud chamber

While the just described expansion cloud chamber hardly plays a role anymore, a variant, the diffusion cloud chamber asserted itself for demonstration purposes. In particular, it is eminently suited for the observation of the natural background radiation.

The principle is depicted in fig. 6.4. A glass case encloses an atmosphere that is saturated with alcohol. From top to bottom a strong temperature gradient develops between a channel, in which the alcohol is evaporated, and a cooled base plate, in which the alcohol condenses. In between is a region, typically a few centimetres thick, where the alcohol is supercritical and where droplets condense out after passage of an ionising particle. Under stationary conditions, that is, under constant temperature gradient and regulated alcohol supply, a continuous operation is possible, in contrast to the properties of expansion cloud chambers.

The shape of the cloud tracks is an indicator for the particle species and its kinetic energy. Natural background radiation generates various track types (fig. 6.4 bottom): heavy, slow particles from nuclear decays (protons and α particles) that leave short, broad tracks; low-energetic β particles that generate thin tracks which often change the direction due to scattering; tracks from high-energetic muons that are thin as well as being, however, quite straight.

Diffusion cloud chambers are offered by producers of teaching materials. In the internet one finds also instructions for do-it-yourself chamber construction with very simple means; see for example [764].

6.2 Bubble chamber

The bubble chamber, invented 1952 by D. Glaser, employs a similar principle to the cloud chamber [462, 463]. Here a supercritical state of a medium is also used to transform a microscopic perturbation into a macroscopic signal. In this case the ionisation tracks become visible by bubbles in a superheated liquid. Liquids are typically three orders of magnitude denser than gases so that a bubble chamber is usually both the target for particle beams and the detection instrument for the reaction products. The employed liquids are, amongst others, hydrogen, propane, neon, freon or mixtures thereof.

6.2.1 Functional principle

The functional principle of bubble chambers is depicted in fig. 6.5. By adiabatic decompression of the liquid the chambers come into a state of boiling retardation. In this state bubbles are seeded by the ionisation clusters of particles which passed the volume shortly before. After the photographic recording the liquid will be compressed again (5–20 bar). The cycle can be repeated about every second, the rate being mainly limited by the propagation of the pressure change in the container. Since the chamber is usually situated in a magnetic field the momenta of the ionising particles can be determined through the track curvature.

In contrast to cloud chambers, bubble chambers have the essential disadvantage that they cannot be triggered. In liquids the seeds for bubble formation are local heat accumulations whose energy is delivered by the recombination of the electrons [884]. While in cloud chambers the ions, which are in this case the condensation nuclei, have lifetimes of many milliseconds, the heat accumulations are dissipated within not more than about 10^{-10} s. Since within this short relaxation time the liquid must already be in the critical state, the triggering of the expansion, for example by using external counters, is in practice not possible. This limits the possible applications appreciably. Therefore bubble chambers have mainly been operated in external accelerator beams

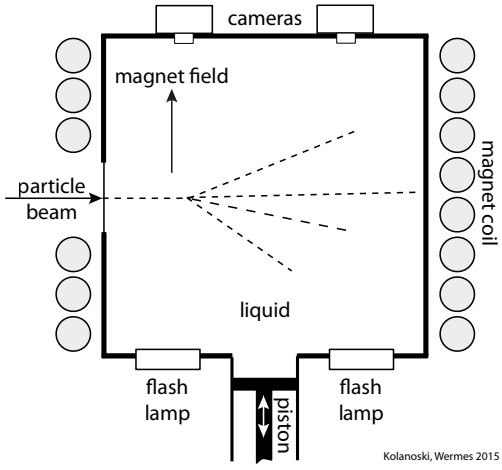
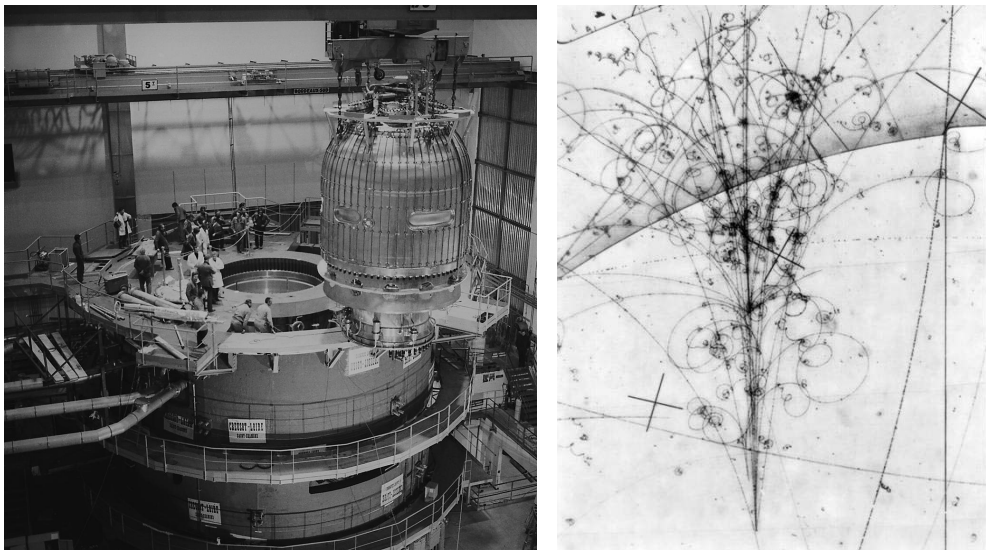


Fig. 6.5 Functional principle of a bubble chamber. In a container a liquid is compressed and decompressed by the movement of a piston. By decompressing the liquid it goes into a superheated state where bubbles form along the tracks of ionising particles. After formation of the bubbles the tracks are illuminated and stereoscopically photographed by several cameras. Normally the chamber is placed in a magnetic field (here a solenoid field), which bends the tracks perpendicular to the main observation direction.



(a) (b)

Fig. 6.6 The Big European Bubble Chamber (BEBC). (a) The vessel containing the liquid with a diameter of 3.7 m, a height of 4 m and a capacity of 35 m^3 is mounted inside the coil of a superconducting magnet. The piston with a weight of 2 t compresses and decompresses the liquid, for example hydrogen with a pressure of 4 bar at a temperature of 26 K at the boiling point. From 1973 onwards hadron and neutrino beams have been directed to the bubble chamber, which was operated with hydrogen, deuterium or a neon–hydrogen mixture (maximal pressures between 5 and 20 bar). By the end of its lifetime in 1984 BEBC had produced 3000 km of film. Source: CERN. (b) Example of a high-energetic neutrino interaction in a neon–hydrogen mixture (72% neon). The short radiation length of the mixture of 44 cm, which is short compared to pure hydrogen with 890 cm, permits the observation of converted photons and bremsstrahlung of electrons. Source: [750].

where the expansion can be triggered synchronously to the periodically ejected particle bunches. Since every bubble grows after being initialised, the photographs can be triggered with a delay after the bubbles reach the desired size. On the other hand, the shortness of the time in which a bubble chamber is sensitive has the advantage that the bubbles can be very precisely localised (if mechanical and optical perturbations can be controlled). With holographic recording techniques bubble resolutions in the order of 10 µm have been reached.

6.2.2 Bubble chamber systems

Until the early 1980s the bubble chamber was one of the most important apparatuses for the detection of high-energy particle interactions.[2] For instance, the first evidence for the neutral current of weak interactions, a direct confirmation of the electro-weak theory, was found with the bubble chamber Gargamelle at CERN in 1973 [492]. The Big European Bubble Chamber (BEBC) (fig. 6.6), which obtained data from 1973 to

[2]For a historical review see [504].

1984, was the largest instrument of its type with a volume of 35 m^3 and a magnetic field of 3.5 T [503].

Since bubble chambers are both targets for particle interaction and detectors for the reaction products, they provide particularly complete and detailed observations of particle reactions. The charged particles can be kinematically reconstructed with high momentum and angular resolution. In addition, the determination of the specific energy loss dE/dx is possible by counting the bubbles along the tracks yielding the Lorentz variables β or γ, where the measurement is sensitive to β at non-relativistic energies and to γ at relativistic energies. The dE/dx measurements can be substantially improved by excluding δ electrons ('restricted energy loss', section 3.2). Together with the momentum measurement the mass and thus the particle identity can be determined, as discussed already for cloud chambers. In bubble chambers, however, the kinematic measurements are generally much more precise because the ionisation is denser and the observable track length is usually larger. In 'heavy' liquids like neon, freon or xenon, photons, electrons and positrons can be identified by their electromagnetic interactions, see fig. 6.6(b) and fig. 3.35.

In their ability of precisely measuring particle reactions, bubble chambers can definitely compete with modern detectors or may even be superior. Given the availability of digital cameras, even the complex photographic recording cannot necessarily be regarded as an exclusion criterion. Nevertheless, because of a combination of various disadvantages bubble chambers have largely been replaced by detectors with electronic readout. For collider experiments there is no need for a target and hence the advantage of bubble chambers to offer target and detector as an entity becomes irrelevant. But in general the decisive disadvantage is primarily the missing possibility for triggering and a comparatively low tolerance for high rates. However, the favourable properties of bubble chambers are still exploited for specific applications, as for example for the detection of dark matter (see section 16.7.3.3) or for the use as neutron dosimeters [557].

6.3 Photoemulsions

6.3.1 Introduction

The radioactivity of atomic nuclei was discovered in 1896 by the blackening of a photographic plate [166] (see also section 2.1). Until the 1950s photoemulsions, when used as detectors also called nuclear emulsions, played a prominent role in the discovery of elementary particles in cosmic radiation. In order to maximise the probabilities to record events, the photoemulsions were brought onto high mountains or to high altitudes with balloons. In this way C.F. Powell and co-workers, for example, discovered the pion (see fig. 16.2 and also section 2.1).

When passing a photosensitive layer an ionising particle leaves a blackened trace after development of the emulsion which can be microscopically measured with resolutions of better than 1 µm. Despite the rapid development of micro-structured, electronically readable detectors (chapter 8) this is still the best spatial resolution of particle detectors that has been reached to date.. Therefore, in the case of particularly high demands on the spatial and angular resolutions, as for measurements of decay vertices of short-lived charm and bottom hadrons or τ leptons, photoemulsions are still today employed [345], if the rate conditions do not require electronic processing.

Typical application areas, where the reaction rates are low enough that the accumulation of tracks in an emulsion is possible during the lifetime of an experiment, are

balloon experiments for measurements of cosmic radiation (section 16.2) and experiments using neutrino beams (see examples in this section). Outside of particle and astroparticle physics, nuclear emulsions are also employed in medicine, biology and geology [345].

6.3.2 Properties of nuclear emulsions

Nuclear emulsions are photographic plates with a transparent gelatine layer which is particularly thick (up to some mm) compared to the emulsions used for light exposures (up to some 100 μm). A large thickness yields larger track segments providing direction information already in a single layer. In the gelatine layer tiny crystals of silver halides (dominantly AgBr, less AgI or AgCl) are uniformly distributed. For nuclear emulsions the AgBr part is generally larger than in normal photoemulsions for light, and the crystals are finer with diameters of typically about 0.2 μm.

The AgBr crystals are semiconductors with a band gap of 2.6 eV, in which electron–hole pairs are created when traversed by ionising particles. The electrons diffuse through the crystal and are trapped preferentially by lattice defects at the crystal surface. Positive silver ions, which sit movably on interstitial sites, can be attracted by the negatively charged defects where they are neutralised and become stationary. By reiteration of this process at the same lattice defect, clusters of metallic silver are formed consisting typically of three to four atoms (see e.g. [345, 934]). Depending on the specific properties of the emulsion these clusters can be stable for periods between several days and several years, but they are not visible because of the small number of atoms. They constitute the so-called latent image which only becomes visible by the development process, augmenting the number of silver atoms of a cluster by a factor of 10^8–10^{10}.

In the development process a metallic silver atom acts as seed for a growing silver grain, which is two to three times larger than the original crystal and becomes visible in the microscope as a black point. The silver halides that remain outside the latent centres are flushed out by the fixation process. The developed silver grains mark the track of an ionising particle with a high spatial resolution of better than 1 μm. The number of Ag grains per 100 μm generated by a minimum-ionising particle is called the 'sensitivity' of the emulsion. It depends on the number and size of the silver halide grains and also on ambient conditions, like temperature and humidity. Typically the sensitivity amounts to about 20–40 Ag grains per 100 μm with grain sizes of about 0.1–1 μm (fig. 6.7) [345].

The density of photoemulsions varies between 2.5 and 4 g/cm^3 depending on the AgBr content and also fluctuates with the ambient relative humidity. In [762] a density of 3.8 g/cm^3, a radiation length of 3 cm and an hadronic interaction length of 35 cm is quoted for a photoemulsion[3] (see also table 3.4). The reason for the relatively short radiation length is the high content of silver and bromine in a nuclear emulsion (mass fractions of 47% and 35%, respectively [762]). For the OPERA experiment [39] (see next section) emulsions with a density of 2.7 g/cm^3 and a radiation length of 5 cm have been employed.

[3]The quoted properties correspond to those of the emulsion Ilford G.5 at a relative air humidity of 58% in [889]; see table 4 there.

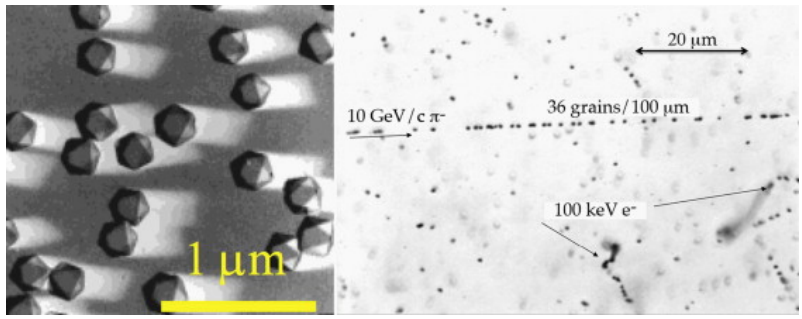


Fig. 6.7 Left: Picture of AgBr crystals in an emulsion layer taken with a scanning electron microscope (SEM). The crystal diameter is about 0.25 µm. Right: Track of a minimum-ionising pion with a momentum of 10 GeV/c. Source: [605], with kind permission of Elsevier.

6.3.3 Emulsions as detectors

6.3.3.1 Emulsion stacks

Nuclear emulsions are normally produced as photographic plates consisting of glass or plastic carriers on which the emulsion is applied with thicknesses between 10 µm and 1000 µm. As already mentioned, thick layers offer the advantage that a track can already be coarsely measured in one layer. At about perpendicular incidence of the particles, for example, the coarse direction of a particle can be measured by focusing the measurement microscope at different depths. Such a direction estimate helps a lot for associating tracks in different layers. On the other hand, thin layers are more stable against deformations during the processing, in particular during the drying after the development.

Usually emulsions are layered in thick packages which are separated into single sheets for microscopic analysis. The relative alignment of the package layers can be made with the help of tracks that pass through several layers or by using dedicated X-ray beams.

In the early 1950s 'emulsion cloud chambers' (ECC) were introduced that are big piles of emulsions, sometimes with absorber plates between the emulsion layers, in which complete reaction processes could be recorded [587]. As for bubble chambers the ECCs serve both as target and as detection instrument. A typical example is shown in fig. 6.8.

6.3.3.2 Emulsion experiments

Example: DONUT experiment. The production of τ leptons by τ neutrinos was first observed in an emulsion target of the DONUT experiment. Figure 6.9 shows the tracks reconstructed in an emulsion target for such an event. Essential for the identification of the event is that the τ lepton tracks are seen before they decay (in collider experiments only the decay vertex can be reconstructed) and that the reconstructed origin agrees with the primary vertex. The average decay length is

$$\lambda = \gamma\beta c\tau = \frac{\gamma^2}{\sqrt{\gamma^2 - 1}} c\tau \,, \tag{6.1}$$

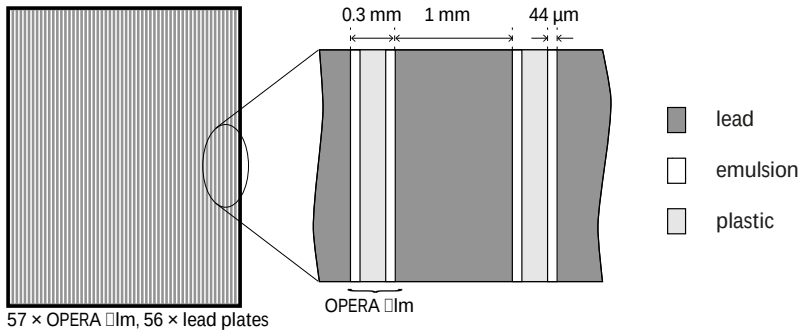


Fig. 6.8 Example for an emulsion target. Shown is the layer structure of a building block (brick) of the emulsion target of the OPERA experiment (from [39]). The photographic films, consisting of emulsion layers on plastic carriers, alternate with lead absorber layers. The front face of a brick has an area of 10.2 cm × 12.8 cm, the depth is 7.9 cm. A target wall is composed of 2912 bricks, in total the experiment has 62 walls which are split into two identical 'supermodules'.

where γ, β are the Lorentz variables and τ is the average lifetime. The relevant parameters of a τ lepton are $c\tau = 87\,\mu\text{m}$ and $m_\tau = 1.777\,\text{GeV}$; thereby a τ lepton with an energy of, for example, 10 GeV has an average decay length of about 500 µm.

In particle physics experiments emulsion detectors are in most cases combined with electronic detectors forming hybrid systems. DONUT is such a hybrid detector, as shown in fig. 6.10 [621]. While the emulsions provide the precise vertex information the electronic detectors provide trigger and time stamps, the association with tracks in the emulsion, energy and momentum determination (magnetic field, calorimeter), and particle identification (muon spectrometer, calorimeter, ...).

Other emulsion experiments. We presented the DONUT experiment as a more recent example for the successful deployment of emulsion targets. Experiments with similar concepts are for example CHORUS [380] and OPERA [39] (see also the review on the use of emulsions in neutrino experiments [377]). The experiment CHORUS at CERN, a predecessor of DONUT, searched for oscillations of muon neutrinos to τ neutrinos. In this case the search was in vain because the relevant ratio of neutrino energy to the detector distance did not cover the right range, as we know today.

The discovery was later made by the OPERA experiment at a neutrino beam which was sent from CERN to the Gran Sasso Laboratory at a distance of 730 km (the beam is called CNGS = *CERN Neutrinos to Gran Sasso*). The OPERA experiment introduced a completely new scale for emulsion experiments. Given the large distance from the beam source, the beam divergence has to be accounted for. Accordingly the target cross section is 6.7 m × 6.7 m, thus nearly 200 times larger than that of the DONUT experiment which was operated only 36 m behind the beam source (fig. 6.10). The OPERA emulsion films have a total area of about 110 000 m^2, which certainly is a challenge for the analysis of the films; see next section.

Common to the three experiments is the fact that they were operated at neutrino beams where the event rates are relatively low such that the emulsion packages could stay for quite long times (months to years) in the experiments.

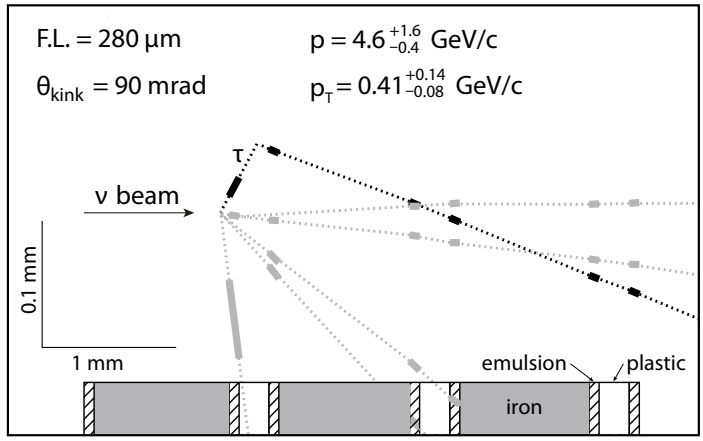


Fig. 6.9 Production of a τ lepton at the primary vertex of a neutrino reaction, recorded in the emulsion target of the DONUT experiment [620] (with kind permission of Elsevier). The reaction is interpreted as $\nu_\tau + X \to \tau^- + Y$ followed by decay of the τ into a charged particle, two neutrinos and possibly neutral hadrons. The τ is identified by its decay causing the kink at 280 μm distance from the primary vertex (F.L. = flight length). The relative position in the emulsion target, which is similar to the emulsion stack shown in fig. 6.8 but with iron as absorber, can be read off from the structure at the bottom of the picture. The different scales for the longitudinal and the transverse dimensions are given at the bottom left. The primary vertex lies in an iron layer and the decay in the following plastic layer. The measurements of the tracks in the emulsion layers are also depicted.

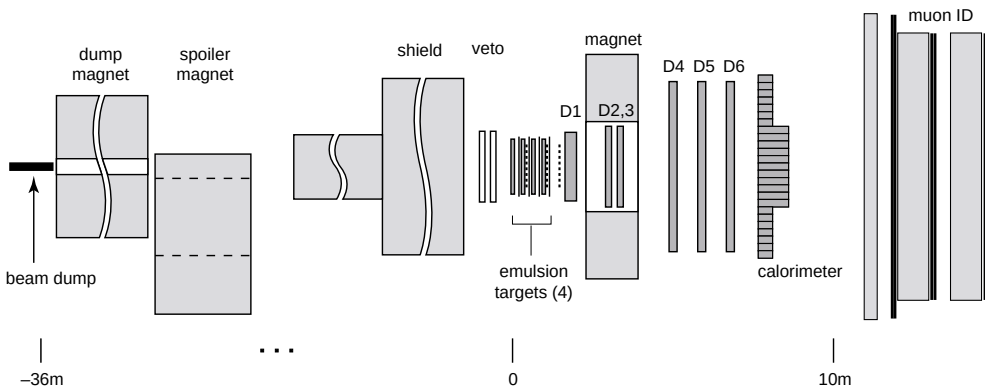


Fig. 6.10 The DONUT experiment [620] (with kind permission of Elsevier). The experiment has taken data at a neutrino beam which was generated by an 800-GeV proton beam in a 1 m long tungsten block ('beam dump'). Such a beam consists mainly of electron and muon neutrinos, however, in charm decays also τ neutrinos are produced. The τ leptons generated by τ neutrinos can be detected in the emulsion target (see fig. 6.9). Subsequent detector components provide additional measurements: a magnet and tracking chambers (D_1–D_6) for the momentum analysis of the charged tracks, a calorimeter for the energy determination of electrons and photons and finally a muon detector for the muon identification.

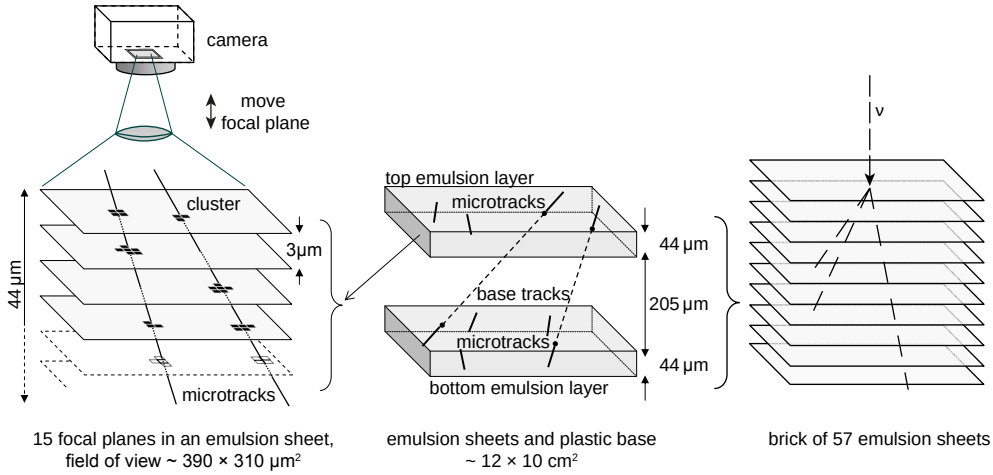


Fig. 6.11 Analysis of the emulsion tracks in a block of the OPERA emulsion target (according to [118, 898], with kind permission of Elsevier). The structure of the block is shown in fig. 6.8. The left pane shows the scanning principle for a single 44 µm thick emulsion layer which is scanned in 15 focal planes at equal distances of 3 µm. Correlated points in different focal planes are combined to a 'micro-track'. The middle pane shows how the micro-tracks in the two emulsion films of a layer are connected across the about 200 µm thick plastic carrier to form 'base-tracks'. This reconstruction yields spatial resolutions of 1 µm and direction resolutions of a few mrad [118] for tracks which perpendicularly hit the emulsion layers. The right pane shows the combination of the tracks in a block of 57 double emulsion layers.

6.3.4 Scanning of emulsion films

The emulsion layers are individually scanned for tracks using microscopes with a field of view of some 100 µm. By variation of the focal plane in steps of few micrometres a track can be measured three-dimensionally (tomographic imaging). As an example the scanning of the OPERA emulsion target is depicted in fig. 6.11.

Today's operation of emulsions with areas of several 1000 m^2, in the case of OPERA more than 100 000 m^2, would not be possible without automated, computer-assisted scanning and reconstruction procedures. The first fully automated system is described in [109]. Meanwhile, for the automated scanning of the OPERA films a rate of 72 cm^2/h is quoted [717]. But even then the demands can only be satisfied using many scanning microscopes in parallel.

In order to make the scanning as efficient as possible, in many hybrid experiments the extrapolations of external tracks into the emulsion target are employed to restrict the scanning to regions in which a vertex is expected with a certain probability. It also became customary to insert quickly exchangeable sheets (CS = changeable sheets) between and behind emulsion targets which are first analysed in order to localise tracks in the compact target material more accurately than with external tracks [377].

6.3.5 Other applications of nuclear emulsions

Besides the described applications of emulsions in particle and astroparticle physics, emulsions are also employed in other fields like medicine, biology or geology. Besides the well-known use of emulsions as X-ray films, emulsion films are also used, for exam-

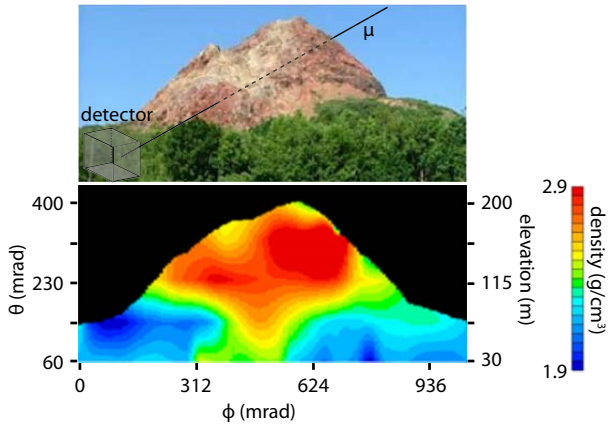


Fig. 6.12 Muon radiography of the volcano Usu in Japan: shown is the average density in the various lines of sight of the emulsion detector. The volcanic lava has a higher density than the surrounding rock. Source: [345], with kind permission of Springer Science+Business Media.

ple, in auto-radiography for the detection of radiation from radioisotopes which mark chemical compounds in biological and other substances (see the references in [556]). In the following we give examples for imaging techniques using radiation of higher energy where the capability of ECCs for the determination of a particle's direction and possibly also its range is employed.

Muon radiography. In 1965 Luis Alvarez proposed investigation of the Chephren pyramid for unknown cavities using cosmic muons [86]. The idea is to use muons as diagnostic radiation like X-rays for the examination of a body. Muons, the most penetrating component of cosmic rays, allow us to explore particularly massive objects like pyramids, volcanos or the rocks above a mine gallery.

Muons can be measured using any detector which provides charged particle tracking with sufficient direction accuracy [745]. Detectors used for this purpose include wire chambers (chapter 7), scintillation horoscopes (chapter 13) and also emulsions. Emulsions, for this application as ECCs (page 165) with good direction accuracy, offer the advantage of not requiring an electric power supply and can thereby be deployed as compact and relatively light units for field applications. The density profile of the volcano Usu in fig. 6.12 was determined using an ECC [345].

Beam characterisation in tumour therapy. In section 3.2.4 the employment of protons and heavy ions in tumour therapy was discussed. As an alternative for existing methods of proton therapy verification (see e.g. [617]) it was proposed to use ECCs that are equipped with tissue-equivalent absorber plates. Studies on such so-called phantoms showed that such a combination of absorber and detector can deliver very accurate three-dimensional dose distributions [243]. In this case the phantom was constructed following the principle of an ECC with OPERA films (fig. 6.8) and polystyrene (plastic).

Radiating with carbon ions (see fig. 3.22 on page 53) the determination of the deposed energy distribution becomes more difficult because carbon undergoes more nuclear reactions than protons of comparable energy and thus more frequently generates nuclear fragments with ranges beyond the Bragg peak of carbon. In [344] a measurement of the properties of a therapeutic carbon beam with an energy of 400 MeV per nucleon is described using an ECC composed of polycarbonate absorber plates and emulsions of variable sensitivity. The results for energy loss, multiple scattering, absorption and secondary particle production demonstrate the potential of ECCs to become important auxiliary means for irradiation planning.

7
Gas-filled detectors

7.1 Overview

In many particle physics experiments detectors are employed which measure charged particles through their ionisation of gases. Such gas-filled detectors allow for the determination of particle trajectories in large volumes, often in a magnetic field. For these applications they have largely replaced cloud and bubble chambers (see chapter 6), mostly because the charges generated along a track can directly be converted into electronic signals which can be more efficiently read out and further processed. Compared to semiconductor detectors (see chapter 8) gaseous detectors are in most cases cheaper, in particular for large volumes, and tend to represent less material for the passing particles.

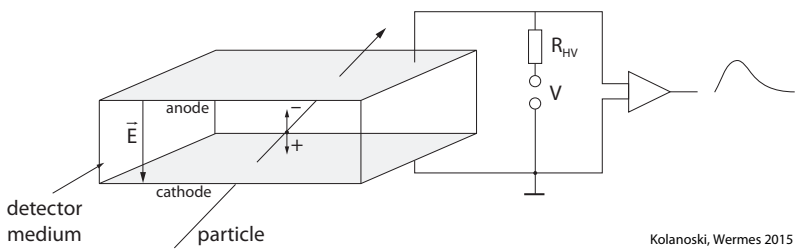


Fig. 7.1 Principle of an ionisation detector. The sensitive detector medium, in this case gas, resides between two electrodes with a voltage applied across. During the passage of a charged particle through the gas charges are liberated which move in the electric field towards the electrodes. The moving charges induce a current signal at the electrodes (see chapter 5). In principle, the detector is a capacitor which discharges upon ionisation of the medium. Electrically it acts as a current source.

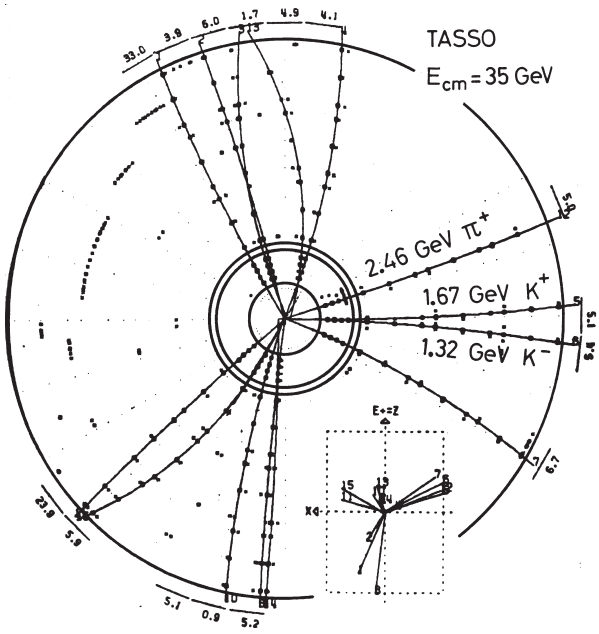


Fig. 7.2 Tracks of charged particles in a three-jet event, interpreted as a quark–antiquark pair with emission of a gluon, in the drift chamber of the TASSO detector at the PETRA storage ring. The picture displays a cross section perpendicular to the colliding electron–positron beams. The insert shows a projection onto a plane containing the beams. Source: DESY.

The principle of a detector which senses the ionisation charges is shown in fig. 7.1: particles ionise a gas in the electric field between capacitor electrodes. The electrons and ions are separated by the field, causing a current between the electrodes which can be registered. This is the principle of an ionisation chamber for the measurement of radiation fluxes containing a large number of particles. For the detection of single particles this current signal is not sufficient for electronic registration because it would not be discernable above the noise level. However, the signal can be intrinsically increased by *gas amplification* in the very high field around the anode. The high field is achieved by a suitable electrode geometry where secondary ionisation can be initiated, leading to charge avalanches.

As low-priced, compact electronics became available by the beginning of the 1970 detectors could be equipped with high granularity readout covering large volumes (Charpak 1968 [293, 291], Nobel Prize 1992 [290]). Figure 7.2 shows as an example the tracks of a three-jet event in the central drift chamber of the TASSO experiment at the PETRA storage ring [222]. This chamber was put into operation in 1978 and was one of the first large drift chambers comprising a sensitive volume of about 16 m^3 and 2340 readout channels. Today such detectors for charged particles may have 10^5 or more readout channels.

7.2 Detector types

For different application areas a variety of different detector types exist using gas as the sensitive medium. On the one hand they can be distinguished according to the way they are used (see also section 7.4 about operation modes): detectors which measure radiation fluxes (radiation monitors such as ionisation chambers and proportional counters) and those which provide information about the locations where the particle passed the detector (position-sensitive detectors such as multiwire proportional chambers, drift chambers and gaseous micro-structure detectors). On the other hand

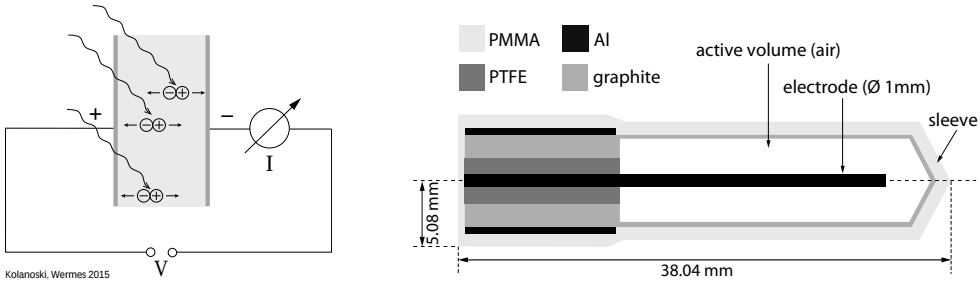


Fig. 7.3 (a) An ionisation chamber is essentially a capacitor, charged up by a voltage V, which discharges due to the ionisation in the capacitor volume. The current is a measure of the radiation flux. (b) A cylindrical ionisation chamber for dosimetry in medical applications (model NE2571) (adapted from [376], with kind permission of Elsevier). The ionisation volume (air) amounts to only $0.6\,\mathrm{cm}^3$ at a diameter of 0.64 cm. The materials of the detector are chosen such that the radiation field is influenced as little as possible (low atomic number Z when measuring in water or air). Here the electrodes are made of graphite and aluminium, insulated by Teflon (PTFE= polytetrafluoroethylene), and the whole detector is covered by a sleeve made of acrylic glass (PMMA = polymethyl methacrylate).

one can distinguish according to whether the primary ionisation charge is measured directly, non-amplified (ionisation chamber) or amplified (e.g. proportional chamber and Geiger counter).

7.2.1 Ionisation chamber

In an ionisation chamber particles ionise a gas volume within a capacitor (fig. 7.3). The electrons and ions are separated by an applied electric field and are measured as current between the electrodes. Ionisation chambers are not sensitive to single particles but they are used as radiation monitors for high particle fluxes. The detectability for charges starts at about a femtocoulomb, that is, about 10^4 elementary charges. A current measurement in the range of nanoamperes is not too difficult, but with some more effort is also possible down to the picoampere regime. A current of 1 nA corresponds to about 10^{10} ionisations per second, while a minimum-ionising particle passing through argon gas generates only about 100 ions per cm (see table 7.1).

In radiation therapy ionisation chambers serve as calibration standard for dosimetric measurements. With an appropriate geometrical format ionisation chambers can also be introduced into body orifices for in-vivo dosimetry (fig. 7.3(b)).

Ionisation chambers are also deployed as smoke detectors or, more generally, for monitoring the purity of gases. In such devices the detector volume is directly connected to the ambient air (or any gas to be monitored) and will be ionised by a weak radioactive source (e.g. the alpha-radiator ^{241}Am, see table A.1 on page 820). Impurities, for example from smoke, boost the neutralisation of the ionisation charges leading to a decrease of the measured ionisation current.

7.2.2 Counting tubes with gas amplification

The principle of a counting tube is shown in fig. 7.4. In a cylindrical tube, which is filled with a special gas, a thin wire is stretched along the cylinder axis. A high voltage is applied between the wire (anode) and the cylinder wall (cathode). In the

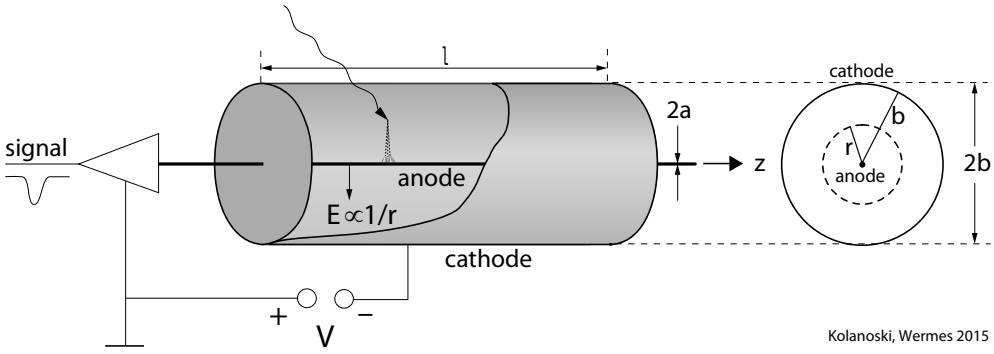


Fig. 7.4 Principle of a counting tube. The plot delineates the relevant geometrical parameters and the cylinder coordinates r and z which are used for the calculation of the field and the detector capacitance (the azimuth is not shown since it is not relevant for the calculations).

$1/r$ field of this cylinder capacitor the electrons created in the ionisation processes can be sufficiently accelerated near the wire that they can initiate secondary ionisation. This leads to the formation of avalanches and thus to the amplification of the ionisation charge with typical amplification factors of 10^4–10^6. This principle of *gas amplification* at a thin wire on the axis of a gas-filled tube was developed by H. Geiger [842]. It eventually lead to the development of the Geiger–Müller counter (section 7.6.2) [454].

Counting tubes are distinguished according to their operation regime: proportional counting tubes work in a regime where the amplified signals are about proportional to the primary ionisation; Geiger–Müller counters are operated with higher amplification in a regime of saturation in which the signals become independent of the primary ionisation (trigger counter). Counting tubes can also be run as ionisation chambers without charge amplification, as depicted in fig. 7.3(b). The various operation modes with different gas amplifications will be discussed in section 7.4.2.

Electric field and capacitance of a counting tube. The electric field in a counting tube corresponds to that of a cylinder capacitor. For the field calculations we use cylinder coordinates: z as longitudinal coordinate along the axis with the origin in the middle of the tube, ϕ as azimuth around the longitudinal axis and r as the perpendicular distance from the longitudinal axis (fig. 7.4). We assume that the length l of the tube is large as compared to the diameter $2b$; the diameter of the wire is $2a$.

The field will be derived from the static Maxwell equations,

$$\vec{\nabla}\vec{E} = \frac{1}{\epsilon_0}\rho \qquad \text{or} \qquad \int_S \vec{E}d\vec{S} = \frac{1}{\epsilon_0}\int_V \rho dV = \frac{1}{\epsilon_0}\int_V dQ\,, \tag{7.1}$$

where S is the surface surrounding the volume V which contains the charge density ρ. In the considered case of a cylinder capacitor the charge is concentrated on the surfaces of the electrodes. Taking into account that here the charge density corresponds to a surface density on an electrode, we obtain

$$E(r)\,2\pi r\Delta z = \frac{1}{\epsilon_0}\Delta Q\,, \tag{7.2}$$

where ΔQ is the charge on the anode surface per length Δz. Assuming a very long wire the charge density per length of the anode wire, dQ/dz, can be taken as constant.

Then the field has only a radial component perpendicular to the anode surface and can be expressed as the radial gradient of a potential Φ:

$$E(r) = \frac{1}{2\pi\epsilon_0}\frac{1}{r}\frac{dQ}{dz} = -\frac{\partial\Phi}{\partial r}\,. \tag{7.3}$$

The field is generated by an applied voltage V for which the following holds:

$$V = \int_a^b E(r)dr = \frac{1}{2\pi\epsilon_0}\ln\frac{b}{a}\frac{dQ}{dz} \overset{(7.3)}{=} E(r)\,r\,\ln\frac{b}{a}\,. \tag{7.4}$$

This finally yields the field and the potential as a function of the applied voltage:

$$E(r) = \frac{V}{r\ln b/a}, \qquad \Phi(r) = -V\,\frac{\ln r/b}{\ln b/a}\,, \tag{7.5}$$

where Φ is fixed such that $\Phi(a) = V$ and $\Phi(b) = 0$. The capacitance per length is

$$C = \frac{dQ/dz}{V} = \frac{2\pi\epsilon_0}{\ln\frac{b}{a}}\,. \tag{7.6}$$

For example, an applied voltage of $V = 1\,\mathrm{kV}$ yields with $a = 10\,\mu\mathrm{m}$, $b = 10\,\mathrm{mm}$ a field strength on the anode surface of

$$E(a) = \frac{1\,\mathrm{kV}}{10\,\mu\mathrm{m}\,\ln\left(\frac{10\,\mathrm{mm}}{10\,\mu\mathrm{m}}\right)} = 145\frac{\mathrm{kV}}{\mathrm{cm}}\,,$$

and the capacitance per length is $C = 8\,\mathrm{pF/m}$.

The $1/r$ behaviour of the electric field and also of the weighting field (see section 5.3.2) leads to a logarithmic increase with time of the charge or voltage signals, respectively (eq. (5.67) in section 5.3.2). In that section it was also shown that typically the signal contribution by the ion movement strongly outweighs the contribution by the electron movement (eq. (5.58)).

7.2.3 Position sensitive chambers

Figure 7.5 shows the principle of a gas-filled chamber with which the coordinate of a traversing particle can be detected. The most important types of position-sensitive gaseous detectors are the multiwire proportional chamber (MWPC, section 7.8, used as example in fig. 7.5), the drift chamber (section 7.10) and relatively novel chambers with micro-structured electrodes (section 7.9). As in proportional tubes, gas amplification takes place at the thin wires or fine microstrips if the anode–cathode voltage is sufficiently high. The wire spacing of about 2 mm determines the spatial resolution in an MWPC. In a drift chamber, an advancement of the MWPC, the drift time of the electrons from the generation by ionisation to their arrival at the wire is measured in order to calculate from this the location of the ionisation. The wire separations in drift chambers can be much larger than in MWPCs, in the range of centimetres, while reaching typical spatial resolutions of about 100 µm.

Predecessors of these chamber types were spark chambers (section 7.7.1) which today are still employed for presentations of cosmic radiation because of their public appeal. At very high gas amplification sparks develop along ionisation trails which can be sensed both optically and acoustically.

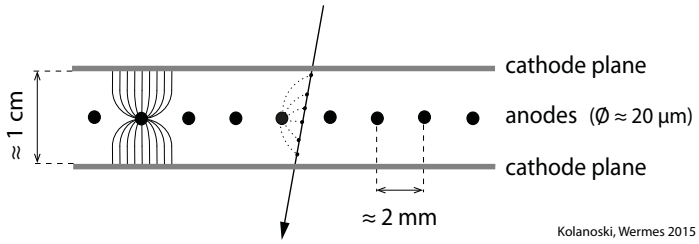


Fig. 7.5 Principle of a position-sensitive multiwire proportional chamber (MWPC). The continuous lines represent the field lines around an anode and the dashed lines the drift paths of electrons generated by ionisation.

7.3 Ionisation and charge loss in gases

The basis for the detection of particles in gas-filled detectors is the creation of charges by ionisation. For the efficiency of signal formation it is also important that the charges do not get lost by recombination or attachment while they are moving towards the electrodes.

7.3.1 Ionisation

Primary ionisation. Particles which pass the detector ionise the gas along their trajectories. The mean energy loss per path length can be determined using the Bethe–Bloch formula (3.25). In table 7.1 the material constants which are important for ionisation are compiled for the most common drift gases. The mean energy loss dE/dx and the number n_p of primary ions per path length are given for minimum-ionising particles. For example, in argon 29.4 primary electron–ion pairs are generated with a total energy expenditure of 2.44 keV, hence on average about 83 eV per primary pair. A part of the energy is lost for excitation of atoms, another part is converted into kinetic energy of the liberated charge carriers, mainly of the electrons. The kinetic energy T of the electrons follows approximately a $1/T^2$ distribution according to (3.46) (see also fig. 3.11(b)).

Secondary ionisation. A part of the electrons has sufficient energy for generating more ions. This secondary ionisation leads to a total number n_{tot} of generated electron–ion pairs per path length which can be multiples of the number n_p of primary electron–ion pairs. For the example of argon n_{tot} is 94/cm and thus about three times more than n_p. The primary ionisations are Poisson-distributed along the particle track. However, the secondary ionisation causes the charges to be produced in ionisation clusters with a gas-specific distribution of cluster sizes. A discussion of cluster distributions obtained from experimental data can be found in [217].

Table 7.1 also contains the energy w_i which has on average to be provided for the production of one electron–ion pair:

$$w_i = \frac{\Delta E}{n_{\text{tot}}} \qquad \text{with} \qquad \Delta E = \left\langle \frac{dE}{dx} \right\rangle \Delta x\,. \tag{7.7}$$

For example, table 7.1 lists for argon $w_i = 26\,\text{eV}$. This energy is distinctly higher than the threshold energy for ionisation, $E_{ion} = 15.8\,\text{eV}$. The difference is due to additional energy loss processes, for example, energy transfer by atomic excitations or energy dissipation by elastic scattering.

Table 7.1 Properties of various gases which are used for detectors. The table was compiled with the help of [849, 890, 762, 310, 856, 511]. For DME I was calculated using (3.37) on page 36; the DME density was calculated employing the van-der-Waals formula for extrapolation to NTP conditions with data from [511]; for $C_2H_2F_4$ (R134a) the parameter set found in the literature is incomplete. In some cases the available data are not quite consistent and a reasonable choice had to be made.

Gas	Z	A (g/mol)	ρ (mg/cm^3)	E_{ex} (eV)	E_{ion} (eV)	I (eV)	w_i (eV)	$dE/(\rho dx)$ (MeV cm^2/g)	dE/dx (keV/cm)	n_p (cm^{-1})	n_{tot} (cm^{-1})	X_0 (m)
H_2	2	2.02	0.084	10.8	15.4	19.2	37	4.10	0.34	5.2	9.2	7522
He	2	4.00	0.166	19.8	24.5	41.8	41	1.94	0.32	5.9	7.8	5682
N_2	7	28.01	1.17	8.1	15.6	82	38	1.83	2.13	(10)	56	325
O_2	8	32.00	1.33	7.9	12.1	95	33	1.80	2.40	22	73	257
Ne	10	20.18	0.839	16.6	21.6	137	37	1.72	1.45	12	39	345
Ar	18	39.95	1.66	11.5	15.8	188	27	1.52	2.53	29	94	118
Kr	36	83.80	3.49	10	14	352	25	1.36	4.73	32	192	33
Xe	54	131.29	5.48	8.4	12.1	482	22	1.26	6.88	44	307	15
CO_2	6,8	44.01	1.84	5.2	13.8	85	37	1.82	3.35	36	91	183
CH_4	6,1	16.04	0.667	9.8	15.2	41.7	30	2.42	1.61	25	53	646
C_2H_6	6,1	30.07	1.26	8.7	11.7	45.4	26	2.30	2.91	41	111	340
i-C_4H_{10}	6,1	58.12	2.49	6.5	10.6	48.3	29	2.28	5.67	84	195	169
CF_4	6,9	88.00	3.78	12.5	15.9	115	64	1.69	6.38	51	100	92
C_2H_6O (DME)	6,1,8	46.07	1.93	6.4	10.0	62.9	24	2.11	4.08	55	160	212
$C_2H_2F_4$ (R134a)	6,1,9	102.03	4.32			95		1.77	7.65	82		81

Z, A, ρ, dE/dx	atomic number, atomic/molar mass, density, mean specific energy loss
E_{ex}	excitation energy (minimal energy for excitation of the ground state to a higher electronic state),
E_{ion}	ionisation threshold,
I	mean excitation energy (defined in the Bethe–Bloch formula (3.25)),
w_i	mean energy for the generation of an electron–ion pair,
n_p, n_{tot}	number of primary ionisations per cm, total number of generated electron–ion pairs per cm,
X_0	radiation length.
ρ, dE/dx, n_p, n_{tot}, X_0	apply for *normal conditions* of temperature and pressure (NTP) as defined in section 2.5.2 on page 20 ($T = 20\,^\circ$C and $p = 1.01325$ hPa = standard atmosphere).

Table 7.2 Different types of electron attachment reactions.

Mode	Reaction	Example
radiative	$e^- + \mathrm{X} \rightarrow \mathrm{X}^- + h\nu$	$e^- + \mathrm{O_2} \rightarrow \mathrm{O_2^-} + h\nu$
dissociative	$e^- + \mathrm{XY} \rightarrow \mathrm{X}^- + \mathrm{Y}$	$e^- + \mathrm{O_2} \rightarrow \mathrm{O}^- + \mathrm{O}$
three-body collision	$e^- + \mathrm{X} + \mathrm{Y} \rightarrow \mathrm{X}^- + \mathrm{Y}$	$e^- + \mathrm{Ne} + \mathrm{O_2} \rightarrow \mathrm{O_2^-} + \mathrm{Ne}$

When choosing a drift gas it is important to reach a high ionisation density but at the same time one wants to minimise multiple scattering (see section 3.4) which increases with the atomic number Z. In table 7.1, helium, for example, has a large radiation length and thus experiences only little multiple scattering, but on the other hand it has a very low ionisation density. A counterexample presents xenon with the opposite behaviour. Favourable in this sense are argon, CO_2 or CH_4. Nevertheless, sometimes also gases with high Z, like xenon, are used for the detection of X-ray photons, for example in transition radiation detectors (see chapter 14).

7.3.2 Recombination and electron attachment

For the operation of proportional and drift chambers it is important that as many electrons as possible reach the anode wire where they are amplified. Recombination of electrons with positive ions and attachment to electronegative components of the gas can lead to losses. For gas amplification it is primarily important to avoid the loss of positive ions near the amplification region because in the high field region ions deliver the largest contribution to the signal (see section 5.3.2).

Recombination. Recombination of electrons with positive ions often proceeds as a radiative process:

$$\mathrm{X}^+ + e^- \rightarrow \mathrm{X} + h\nu\,. \tag{7.8}$$

The excess energy can also be transferred in collisions with other molecules of the gas. In particular, recombination is not avoidable if the detector geometry causes the electrons to drift predominantly along the ionisation trace.

The photons emitted in radiative recombination are problematic, in particular if they are produced in large numbers in an avalanche. The photons can lead to unwanted additional ionisations and, without countermeasures, to runaway discharges. For further discussion of the subject, see section 7.5.

Electron attachment. Atoms and molecules with nearly complete shells can attach electrons under the release of energy. The binding energy liberated by the attachment and the kinetic energy of the electrons can be dissipated by radiation, dissociation of the molecule or in collisions (see table 7.2).

Electronegative additions to drift gases, for example oxygen or halogens, are normally undesired since they reduce the signal. However, sometimes electronegative components are added on purpose to proportional chamber gases in order to quench signals and thus avoid uncontrolled discharges.

The attachment loss of electrons drifting towards the anode can best be parametrised by the electron range in the drift direction. We therefore define an absorption length λ_a or the inverse, the absorption coefficient η_a. For a drift velocity of v_D^e they are given by

$$\lambda_a = v_D^e \langle t_a \rangle = \left\langle \frac{1}{\sigma_a\, n_a} \right\rangle , \qquad \eta_a = \frac{1}{\lambda_a}\,. \tag{7.9}$$

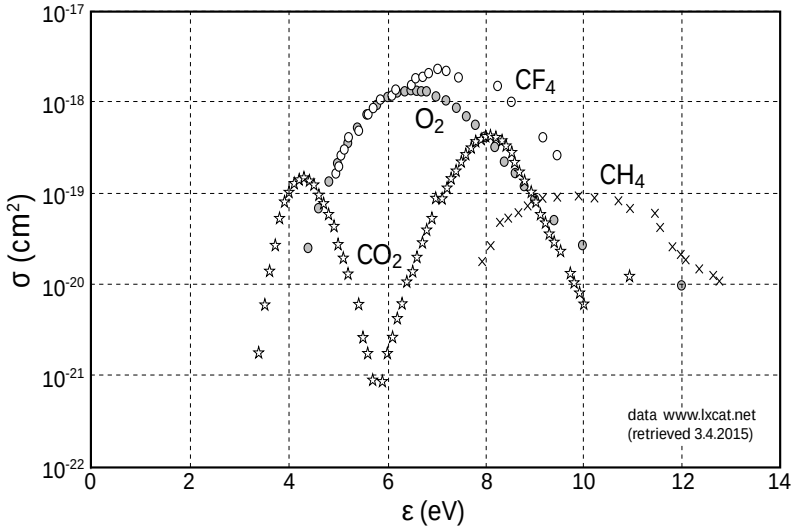


Fig. 7.6 Cross sections as a function of electron energy ϵ for electron attachment of the polyatomic gases CO_2, CH_4 and CF_4, for which the scattering cross sections have already been shown in fig. 4.7. Additionally the dissociation cross section for oxygen is plotted. The data have been retrieved from the website [756]. The used databases are: Hayashi database for CH_4 and CF_4; Itikawa database for CO_2; Biagi-v8.9 for O_2.

The term $1/(\sigma_a n_a)$ is the definition of the mean free path according to (3.7) where in this case σ_a is the—in general energy dependent—attachment cross section and n_a the molecule density. The average absorption time $\langle t_a \rangle$ has to be determined by averaging over the respective energy distribution.

In general, λ_a is dependent on the drift gas, the partial pressure of the electronegative additive and the drift field. Usually a precise determination of λ_a is accomplished by employing a simulation program (e.g. MAGBOLTZ [196]) for averaging the attachment cross sections over the energy distribution. In fig. 7.6 such cross sections are shown for the polyatomic gases CO_2, CH_4 and CF_4 for which scattering cross sections are already contained in fig. 4.7. The attachment proceeds for these gases through various dissociation reactions which exhibit resonance-like behaviour above the respective thresholds. The resonances lie mostly above the usual energy distributions of the electrons in the drift gas and thus do not lead to harmful losses during the drift even over long drift passes. However, the electrons pass these energies in the region of gas amplification (section 7.4.1) which then leads to signal losses. This effect is particularly noticeable for CF_4 featuring high absorption around 7 eV (fig. 7.6).

The dissociation cross section of oxygen is additionally shown in fig. 7.6. The cross section for three-body collisions, which are important for electron loss in oxygen at low energies, is not shown because it is dependent on the density or partial pressure, respectively, of the oxygen in the gas mixture and is thus not easily representable. As an example one finds for an admixture of 1% O_2 in argon at a field of 1 kV/cm an average range for electrons of $\lambda_a \approx 5$ cm.

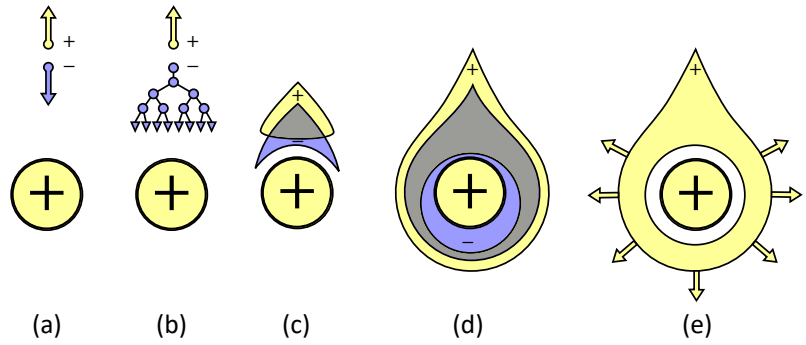


Fig. 7.7 Model of the development of an ionisation avalanche at the anode wire of a proportional tube or chamber (adapted from [849]). (a) In the drift volume an electron and an ion are created which drift to their respective electrode. (b) Near the wire the electron reaches such a high field that it can initiate secondary ionisation which leads to an avalanche. (c) The charges generated in the avalanche are separated by the electric field. (d) Because of the lateral diffusion of the electrons, which is much stronger than that of the ions (see section 4.6), the avalanche spreads out around the whole circumference of the wire and the positive charge cloud forms the shape of a drop. (e) The electrons from the avalanche very quickly reach the anode, within a few nanoseconds, while the ions drift over a longer period (up to milliseconds) to the cathode.

7.4 Gas amplification and operation modes

7.4.1 Gas amplification

If the electric field is strong enough, for example near the anode of a wire chamber, the drifting electrons can be accelerated such that they can initiate secondary ionisations. Then an avalanche develops (fig. 7.7) and thus the ionisation charge is amplified with typical amplification factors of 10^4–10^6 (see section 7.4.2). Beginning at field strengths of about 10–50 kV/cm the energy gain over a free path length between two collisions becomes sufficient for ionisation of the gas. The number α of ions generated per path length, the so-called *first Townsend coefficient*, is given by

$$\alpha = \sigma_{ion}\, n = \frac{1}{\lambda_{ion}}\,, \tag{7.10}$$

where σ_{ion} and λ_{ion} are the cross section for ionisation and the mean free path between ionisations, respectively. Analogously to the field strength in (4.113) this quantity can be normalised to the particle density or the pressure (at fixed temperature):

$$\alpha/n = \sigma_{ion} \qquad \text{or} \qquad \alpha/p \propto \sigma_{ion}\,. \tag{7.11}$$

Figure 7.8(a) shows the first Townsend coefficient as a function of the electron energy for various noble gases. In general the electrons have an energy distribution which is given by the applied electric field (see e.g. fig. 4.8). Hence it is more informative to plot an effective Townsend coefficient as a function of the field strength, as in fig. 7.8(b) where α/p is displayed as a function of the reduced field strength E/p.

The increase dN of the number of electron–ion pairs over a path length ds is

$$dN = \alpha\,(E)\; N\, ds\,. \tag{7.12}$$

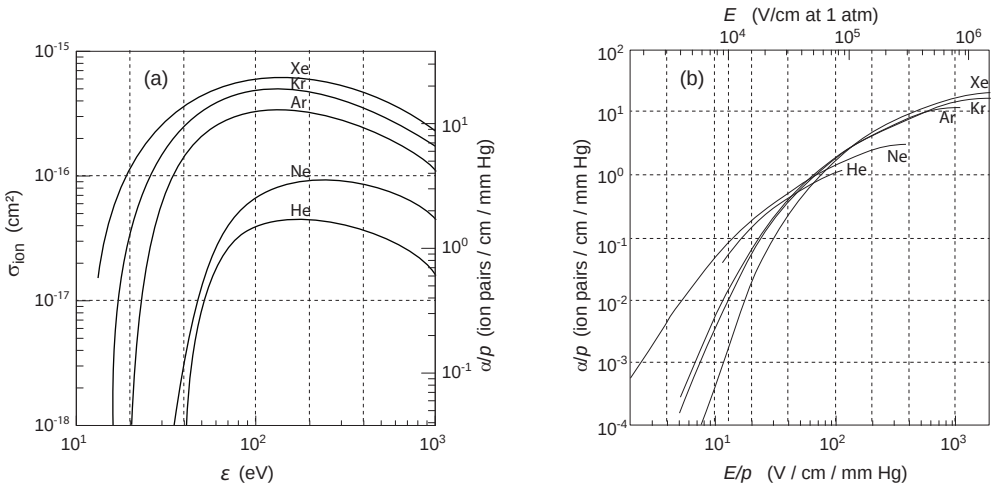


Fig. 7.8 The ionisation cross section and the first Townsend coefficient for electrons in various noble gases (a) as a function of the electron energy and (b) as a function of the electric field (top axis) and reduced electric field (bottom axis), respectively (from [849]).

It follows that

$$N(s_\mathrm{a}) = N_0 \, \exp\left(\int_{s_0}^{s_\mathrm{a}} \alpha\,(E(s))\,ds\right)\,, \tag{7.13}$$

where N_0 is the number of unamplified electrons at $s = s_0$, where the amplification starts, and $N(s_\mathrm{a})$ the number of electrons which reach the anode at $s = s_a$. The ratio of both numbers defines the gas amplification G:

$$G := \frac{N(s_\mathrm{a})}{N_0} = \exp\left(\int_{s_0}^{s_\mathrm{a}} \alpha\,(E(s))\,ds\right)\,. \tag{7.14}$$

If α does not depend on s this yields

$$G = \exp\left(\alpha\,(s_\mathrm{a} - s_0)\right)\,. \tag{7.15}$$

In general, however, α is energy dependent. The energy distribution varies with the electric field which in turn is position dependent. Then the amplification has to be calculated using (7.14) which is in general only possible numerically. For practical purposes parametrisations of the gas amplification as a function of the applied voltage have been developed for which we give an example later in this section.

In the regime where the amplified signal is proportional to the original ionisation charge (see section 7.4.2) G lies between about 10^3 and 10^6. We can estimate the number of collisions necessary to obtain such an amplification by assuming that the number of electrons duplicates at each collision:

$$G = 2^n \Rightarrow n = 13\text{–}20\,. \tag{7.16}$$

In this estimate half of all charges in the avalanche are created on the last mean free path length $\lambda_{ion} \approx 1{-}2\,\mu\mathrm{m}$. Hence most of the electrons have a vanishingly small drift path to the anode while the ions have to traverse a long way to the cathode. This plays an important role in signal formation at the anode (see section 5.3.2).

The gas amplification cannot become arbitrarily large because space charges screen the field near the anode (*Raether limit*) [803]; see also the discussion on streamer and spark formation on page 186. Empirically, the avalanche development reaches a saturation at an amplification of the primary electron number N_0 of about

$$G_{max}\, N_0 \approx 10^8\,. \tag{7.17}$$

Then the current pulse at the electrode becomes independent of the primary ionisation. Counters which just trigger on ionising particles passing the detector, like a Geiger counter, work in this operation mode. Upon further increase of the voltage, it results in discharges; see details in section 7.4.2.

Amplification in cylinder geometry. In the following the gas amplification in a cylindrical device like the counting tube introduced in section 7.2.2 will be discussed. A cylindrical geometry has a more general applicability. If the gas amplification takes place at a cylindrical anode wire of a detector with otherwise quite arbitrary geometry, one can assume a $1/r$ behaviour of the field in the amplification region very near to the wire, as in a counting tube. For the case of a $1/r$ field various parametrisations of the gas amplification as a function of the applied voltage exist which are based on different assumptions. Here we introduce the derivation of the so-called *Diethorn formula* [354].

To begin with, the path integral over the Townsend coefficient in (7.14) can be rewritten as an integral over the field strength. With the expression for the field of a counting tube (7.5) one obtains

$$E(r) = \frac{V}{r \ln \frac{b}{a}} \qquad \Longrightarrow \qquad dE = -\frac{V}{r^2 \ln \frac{b}{a}}\, dr = -E^2 \frac{\ln \frac{b}{a}}{V}\, dr\,, \tag{7.18}$$

where V is the voltage between the electrodes and a, b are the radii of the electrodes. Inserting this into (7.14) and substituting $dr = -ds$ yields the integral

$$G = \exp\left(\frac{V}{\ln \frac{b}{a}} \int_{E_{min}}^{E(a)} \frac{\alpha(E)}{E^2} dE\right)\,. \tag{7.19}$$

The integration limits are the minimal field strength E_{min} at which the multiplication of the electrons starts and the field strength $E(a)$ on the anode surface.

With the knowledge of $\alpha(E)$ the integral can at least be solved numerically. In order to obtain an analytic solution Diethorn [354] made the assumption that the Townsend coefficient is proportional to the field strength,[1] which is in the relevant region of low field strength quite well fulfilled:

$$\alpha(E) = k\, E\,, \tag{7.20}$$

where k is the proportionality constant. Then the integral (7.19) yields:

$$\ln G = \frac{kV}{\ln \frac{b}{a}} \ln\left(\frac{E(a)}{E_{min}}\right)\,. \tag{7.21}$$

[1] An alternative, dating back to Rose and Korff, is the assumption that the Townsend coefficient depends linearly on the electron energy [833, 630], see for example also [849].

The parameter k can be related to the energy w_i (table 7.1) which has on average to be spent for one ionisation. This is the energy which the electron has to acquire by traversing a potential difference ΔV; thus we get $w_i = e\Delta V$. It follows that the number of multiplication steps in a potential difference $\Delta\phi$ is

$$m = \frac{\Delta\phi}{\Delta V}\,, \tag{7.22}$$

where $\Delta\phi$ is the potential difference over the amplification region:

$$\Delta\phi = \int_a^{r(E_{min})} E(r)\,dr = \frac{V}{\ln\frac{b}{a}} \ln\left(\frac{E(a)}{E_{min}}\right)\,. \tag{7.23}$$

With the assumption $G = 2^m$ (duplication at each collision) as in (7.16) and the relation (7.22) one finds

$$\ln G = m\,\ln 2 = \frac{V\,\ln 2}{\ln\frac{b}{a}\,\Delta V} \ln\left(\frac{E(a)}{E_{min}}\right) \overset{(7.21)}{=} \frac{k\,V}{\ln\frac{b}{a}} \ln\left(\frac{E(a)}{E_{min}}\right) \;\Rightarrow\; k = \frac{\ln 2}{\Delta V}\,. \tag{7.24}$$

This yields an expression for the gas amplification G as a function of the voltage V with the parameters $\Delta V (= w_i/e)$ and E_{min} which depend on the gas properties. While $E(a)$ is fixed for a given voltage,

$$E(a) = \frac{V}{a\ln\frac{b}{a}}\,, \tag{7.25}$$

E_{min} is obtained from the condition that an electron is accelerated over a free path just as much that it can ionise. Since the free path is inversely proportional to the pressure (or the particle density of the gas) we can additionally include into the parametrisation the gas pressure dependence by

$$E_{min}(p) = E_{min}(p_0)\frac{p}{p_0}\,, \tag{7.26}$$

where p_0 is a reference pressure. Then the equation for the amplification as a function of the tube voltage, named the Diethorn formula, becomes

$$\ln G = \frac{V\,\ln 2}{\ln\frac{b}{a}\,\Delta V} \ln\left(\frac{V}{a\,\ln\frac{b}{a}\,E_{min}(p_0)\frac{p}{p_0}}\right)\,. \tag{7.27}$$

The so-called Diethorn parameters ΔV and $E_{min}(p_0)$ can be found tabulated for some common gases, see for example [217,616]. They are determined from measurements of the gas amplification as a function of the voltage. This is usually done employing an external radiation source with constant intensity, for example an X-ray tube, for the measurement of the current amplification with increasing voltage. When ramping up the current, care has to be taken that it does not become too high in order to avoid space charge effects that reduce the field in the amplification region.

As an example, fig. 7.9 shows measurements of the gas amplification in two drift tubes with different diameters but otherwise equal properties. In fig. 7.9(a) the Diethorn formula was fit to the measurements with the 10-mm-diameter tubes. The fit yields the parameters $\Delta V = 53.8\,\mathrm{V}$, $E_{min} = 39.2\,\mathrm{kV/cm}$. Similar results were found

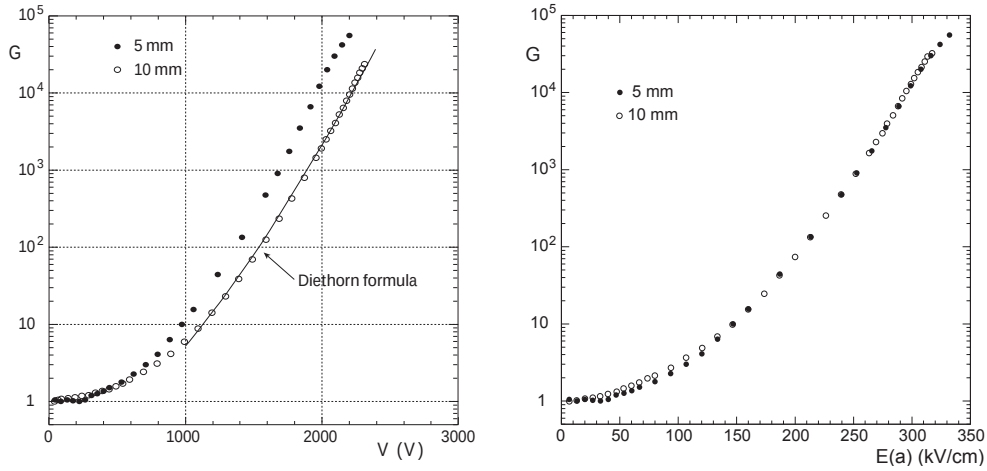


Fig. 7.9 Measurement of gas amplification in drift tubes (honeycomb structure as in fig. 7.35) with 5 and 10 mm diameter, both with 25 µm thick anode wires (adapted from [877]). For these measurements a CF_4/CH_4 mixture in the ratio 80:20 was used as a drift gas. (a) Dependence of the amplification on the anode–cathode voltage for both drift tubes. The fit to the measurement points of the 10-mm-diameter tube was done with the Diethorn formula (7.27). (b) This picture demonstrates that the curves for the drift tubes with the two different diameters fall nearly on top of each other if one plots the amplification versus the field strength on the anode surface. This results from the Diethorn formula (7.27) if only the outer diameter $2b$ of the counting tube varies but the other parameters are kept the same (see explanation in the text).

for the same gas in [477]. The formula describes the measurements above 1 kV quite well; however, at lower voltages the function does not have the flexibility to turn into a constant (=1).

The plot in fig. 7.9(b) shows that the amplification in two counting tubes with different cathodes but equal anode diameters are the same when plotted as a function of the field strength on the anode surface. According to (7.27) the voltage has to be scaled in the ratio of the logarithms $\ln(b/a)$ (a fixed) to obtain the same amplification. In the example of fig. 7.9 this ratio is 1.13, which for $V(5\,\text{mm}) = 2000\,\text{V}$ yields a value of $V(10\,\text{mm}) = 2260\,\text{V}$.

7.4.2 Operation modes of gaseous detectors

Figure 7.10 displays the principal dependence of the gas amplification on the anode–cathode voltage for a counting tube with a thin anode wire. The dependence on the voltage V of the gas amplification G can be very different depending on the electrode arrangement and the used gas. The choice of the operation mode depends on the intended application and various constraints, for example, from technical or safety considerations. Referring to fig. 7.10 the following amplification regimes can be distinguished:

Recombination region, $G < 1$. In the region of low field strengths the primary electrons and ions recombine with increased probability if they are not sufficiently

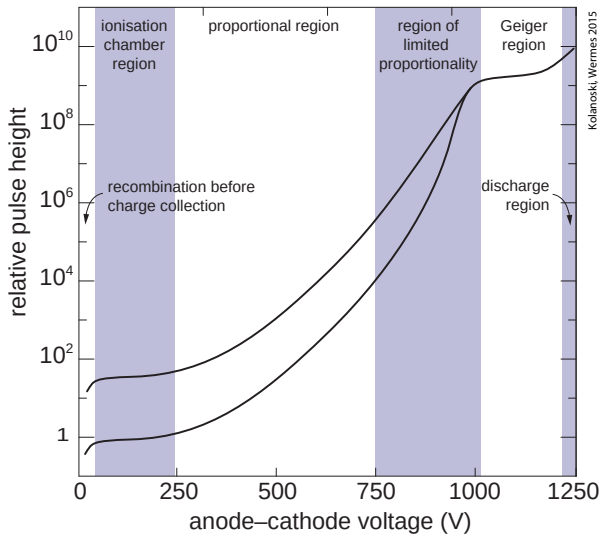


Fig. 7.10 Schematic representation of the dependence of the output signal of a counting tube on the anode–cathode voltage (adapted from [716]). The characteristics are shown for two particles with different primary ionisation. The lower curve is normalised such that the gas amplification can be read off the vertical axis. The numerical values for amplification and voltage are used as examples; in concrete cases they depend strongly on the electrode arrangement and the used gas. The different amplification regions are discussed in the text.

quickly separated (section 7.3.2). Therefore the output signal increases with the voltage until a saturation of the charge collection sets in at which ideally the whole charge reaches the electrodes.

Ionisation chamber region, $G \approx 1$. The saturation of the output signal without amplification defines the regime in which ionisation chambers are operated. This operation mode is suited for measurements of particle fluxes, as for example done with dosimeters (see fig. 7.3 in section 7.2.1). However, it is not suited for the detection of single particles since without amplification the signal charge of single particles is too small for detection (in the range of 0.01 fC/cm for minimum-ionising particles).

Proportional region, $G \approx 10^3$–10^5. If in a strong field electrons gain sufficient energy between two collisions to produce secondary electrons, a charge avalanche develops, as described in section 7.4.1. With increasing operation voltage the amplified charge remains over a wide voltage range proportional to the primary charge. Therefore, as can be seen in fig. 7.10, in this proportional regime the curves for different primary ionisations have equal distances, meaning that the amplification is independent of the primary ionisation. In this operation mode a charge measurement in combination with a momentum measurement can be used for particle identification (see chapter 14, section 14.2.2).

Region of limited proportionality, $G \approx 10^5$–10^8. With increasing voltage the proportionality of the output signal to the primary ionisation will be limited by space charge effects. At sufficiently high amplification an ion cloud builds up at the anode which only slowly drifts away because of the relatively small mobility of the ions. This charge cloud diminishes the field locally around the anode wire. The regime of limited proportionality will be reached earlier the higher the primary ionisation.

Saturation and Geiger region, $G \gtrsim 10^8$. With further increasing gas amplification the output signal finally becomes independent of the primary ionisation. A detector which works according to this principle 'counts' ionising particles or quanta

independent of their type. The positive and negative charges of an avalanche, which drift away from each other, start to neutralise in the inner region of the avalanche because of the higher recombination probability at high charge density. The recombination leads to the emission of photons which can produce further electrons by the photoelectric effect in the counting gas or the chamber walls and thus initiate new avalanches. How far this formation of avalanches extends depends on the cross section for photoionisation in the counting gas. In the Geiger mode the charge avalanche spreads out over the whole wire of a counting tube. At the time of invention of the Geiger counter at the beginning of the twentieth century the resulting high current signal was essential for the electronic signal registration. The discharge at the anode wire will be stopped by the screening effect of the avalanche charge and/or by a voltage drop over a shunt resistor connected in series to the power supply (*self-quenching*).

However, the disadvantage of a high current signal is a correspondingly long dead time after each pulse, about 50–100 μs, which limits the use of Geiger counters to rates below 10^4–10^5 Hz. In proportional counters one tries to suppress avalanche expansion by adding to the counting gas (typically argon or CO_2) an organic gas as a quenching gas which absorbs UV photons. Such quenchers are for example methane (CH_4), isobutane (C_4H_{10}) or methylal ($(OCH_3)_2CH_2$). It is also possible to tune the photoabsorption with the choice of the mixture such that the expansion of the avalanche remains restricted to a fraction of the anode wire, say about 1 cm, in order to decrease the dead time while keeping a relatively high signal. This operation mode is called the *limited Geiger mode*.

Discharge region, above $G \approx 10^8 - 10^9$. At very high voltages self-sustaining discharges occur, either spontaneously or triggered by ionising particles. In a discharge both electrodes become connected through a conducting plasma tube. Before the plasma tube reaches the electrodes a spatially limited *streamer* builds up in the region between the electrodes (see next paragraph). The transitions between a saturated avalanche, the formation of a streamer and discharges, like glow, corona or spark discharges, are floating and strongly dependent on the electrode geometry, the gas properties (composition, pressure, temperature, humidity, ...) and the resistance of the electrical current circuit. The discharge regime can also be interesting for particle detection because very high signals can be obtained. However, with modern readout electronics this is only a criterion in special cases.

Operating a detector in the discharge regime always requires a mechanism for the controlled termination of a discharge (or for stopping a streamer). For example, discharges can be disrupted by:

- increasing space charge which limits avalanche formation;
- active interruption of the high voltage;
- applying a voltage pulse onto the electrodes for a short time only, usually initiated by a particle trigger;
- passive reduction of the high voltage below the amplification threshold by conducting the signal current through a resistor causing a voltage drop.

These mechanisms, or combinations thereof, also serve when operating in the proportional regime as a protection of the detectors against destructive sparking.

The transition avalanche–streamer–spark. There is a wealth of investigations of discharge phenomena in gases with a lot of related literature, often in view of the high voltage safety of electrical apparatuses and of electrical phenomena in the

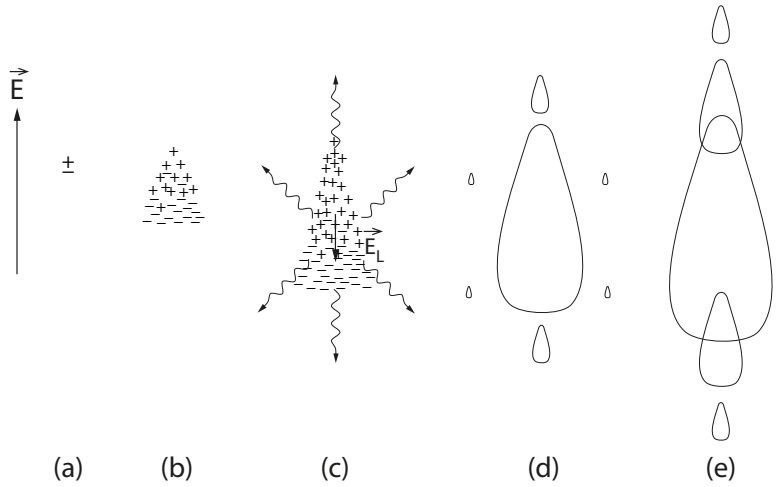


Fig. 7.11 Evolution of a charge avalanche in a strong electric field into a streamer (adapted from [816, 817], with kind permission of IOP): (a) primary ionisation; (b) formation of an avalanche by gas amplification; (c) build-up of a counter-field in the avalanche, recombination in the neutral region of the avalanche and subsequent photoemission; (d) generation of new avalanches by photoionisation; (e) fusion of the avalanches to a streamer.

atmosphere (see e.g. [704, 804, 803]). Here we want to concentrate on the transition from an avalanche to a discharge spark which is important for detector operation.

This transition is sketched in fig. 7.11 based on a simplified model [816,817]. Starting with the generation of an electron–ion pair (a) in a strong electric field initially an avalanche develops (b). The avalanche quickly spreads out in the direction of the anode leaving the slower ions behind. Because of the stronger diffusion of the electrons (section 4.6) a drop-shaped charge distribution forms with a positive charge tip and a broader negative front, as already displayed in fig. 7.7. Between the separating charges an electric field arises which is opposed to the external field, while simultaneously the field between the charge fronts of the avalanche and the electrodes becomes stronger. At gas amplification in the range of 10^8 the field within the avalanche becomes so strong that it compensates the external field. In the thus forming neutral region inside the avalanche the probability is increased that electrons and ions recombine, thereby emitting a photon (c). The photons generate new charge pairs in the surroundings of the avalanche which create new avalanches (d). The avalanches form particularly strongly in the regions with high field between the original avalanche and the electrodes, thereby favouring the expansion towards the electrodes. The velocity of the expansion of about 10^6 m/s is determined by the photons and is thus about an order of magnitude faster than that of the original avalanche, which is determined by the electron drift. The growing avalanches begin to merge (e) and form a so-called *streamer* which quickly grows towards the electrodes while glowing due to photoemission. When the streamer reaches the electrodes they become connected via a conducting plasma through which the electrodes discharge in the form of a spark.

The condition under which an avalanche, after having reached saturation (see (7.17)), develops into a streamer and then into a discharge is called the Raether or

also Raether–Meek condition [803, 704].[2] It states that without a discharge the gas amplification in an avalanche is limited to about 10^8–10^9 which at constant electric field corresponds to

$$\alpha\, d \approx 18\text{–}20 \tag{7.28}$$

secondary ionisations (α is the Townsend coefficient (7.10) and d is the length of the avalanche). These values can only be regarded as a rough rule of thumb because more precise values depend on various factors, like the diffusion of the charge cloud, photoabsorption in the gas and the nature of the electrodes [804].

By early shutting-off or lowering the voltage, the possibility that the streamer reaches the electrodes can be avoided. Used in this way, the *streamer mode* for detectors offers some advantages as compared to operation as a spark chamber (see section 7.7). Detectors using the streamer mode are streamer tubes, streamer chambers and resistive plate chambers, as described in the sections 7.6.3, 7.7.2 and 7.7.3, respectively. In homogeneous fields, for example between parallel plates, streamers can be initiated at arbitrary positions if the field is sufficiently strong. In cylindrical counting tubes they develop from the avalanche near the wire. In this geometry it is possible to stop streamer evolution by the developing space charge if a suitable gas with sufficient photon absorption is chosen. This operation mode is referred to as the *self-quenching streamer* or *limited streamer* mode. In this case the streamer mode differs from the limited Geiger mode (see above) by the fact that the highest charge density of the streamer is distinctly separated from the wire because the streamer grows into the direction of the cathode [295, 551].

7.5 Choice of chamber gases

In this section we will discuss general criteria for the choice of chamber gases following the corresponding discussion in [849]. The main criteria for the gas choice are:

- high ionisation density;
- little charge loss;
- low voltage at the required amplification;
- stable operation, safety against spark discharge;
- proportionality between ionisation and output signal;
- low diffusion (in particular for detectors with position resolution);
- suitable drift velocity (again for detectors with position resolution);
- rate tolerance and low or minor dead time (low space charge accumulation);
- radiation resistance;
- safety: non-inflammable, non-toxic, environmentally acceptable;
- costs.

Ionisation and charge loss. The ionisation density in table 7.1 on page 177 is very good for the noble gases argon, krypton, xenon as well as for the molecular gases CO_2, the hydrocarbons (except for methane), CF_4 and DME and best for krypton and xenon. Because of charge losses, electronegative gases, in particular oxygen, cannot be considered as main components of a chamber gas.

[2] Note that what we called the 'Raether limit' given by (7.17) is the limitation of amplification due to saturation caused by space charge. Here we are discussing the transition to a discharge.

High voltage and gas amplification. In order to minimise sparking and leakage currents, which occur in particular near the mounts of electrodes, the potential differences between electrodes should be as low as possible. A high amplification at a relatively low voltage is provided by the noble gases helium, neon, argon, krypton and xenon, whose thresholds for gas amplification lie at field strengths of about 70 kV/cm. Krypton and xenon have the disadvantage of high costs and for track detectors the short radiation length is also disadvantageous because of the increased multiple scattering affecting the reconstruction precision. On the other hand, exactly because of the short radiation length, both gases are employed for the detection of photons, for example for the detection of X-ray photons in a transition radiation detector (chapter 12). In most cases helium and neon are less suited because of their low ionisation density. However, because of their long radiation length, they are employed for the detection of particles with low momenta in order to reduce multiple scattering. The deficit in ionisation density can be balanced by an admixture of DME (table 7.1) which also has a long radiation length but a high ionisation density. The light noble gases are also employed for detectors operating in the discharge mode because there the ionisation density is less important. For example, helium-neon mixtures are used in spark chambers.

But in general argon is the preferred gas for detectors with gas amplification in the proportional regime. Pure argon, however, exhibits some unfavourable properties, in particular a very strong diffusion (high characteristic energy, see section 4.6.4.5 and fig. 4.12 on page 117) and an unstable behaviour of the gas amplification. Therefore argon is practically only used together with admixtures, as will be further discussed in the following paragraphs.

Photoabsorption and quenching gases [849]**.** Pure noble gases exhibit an unstable behaviour in detectors with gas amplification because the photons which are produced in the amplification process cannot be absorbed. In contrast to polyatomic molecules which can absorb UV photons over a wide spectral band because of the multitude of excitation levels, noble gases have only few discrete levels (see e.g. fig. 4.7 on page 110).

In the gas amplification process atoms and molecules are ionised and excited. For example, an excited argon atom can only fall back into the ground state by emitting a photon with a minimal energy of 11.6 eV. This energy lies far above the ionisation potential of copper of 7.7 eV so that such a photon, once it hits a copper cathode, can extract an electron which initiates another amplification avalanche. A new avalanche can also be initiated when an argon ion becomes neutralised at the cathode. The released energy will be either radiated by a photon which can again ionise, or can be used for the extraction of an electron from the cathode. In all cases this leads to a delayed avalanche or at higher gas amplification to a permanent discharge.

The solution of the problem consists of adding to the noble gas, mostly argon, a quencher. Usually one uses the hydrocarbons from methane to isobutane and, though with worse quenching properties, also CO_2 (the influence of these gases on the drift velocity and other transport parameters is discussed in section 4.6). These molecules have a broad spectrum of rotational and vibrational states which can be excited by absorption of photons but which release the excitation energy preferentially in non-radiative processes (collisions). The quenching properties of gases improve with an increasing number of atoms per molecule.

Frequently used mixtures of argon with a quenching gas are for example:

Argon–ethane	50 : 50,
Argon–methane	90 : 10,
Argon–isobutane	75 : 25,
Ar–CO_2–CH_4	90 : 9 : 1.

The mixing ratios are given for guidance only. They can be varied in order to optimise different detector properties.

Under high radiation load the problem arises that hydrocarbons cause undesired depositions of polymerisates on the electrodes, which can lead to a complete failure of a detector (see details in section 7.11). In this regard mixtures of argon with CO_2 are much safer though more unstable at high amplification. In this case, however, often a third component is added to strengthen the quenching, as for example in the combination Ar–CO_2–CH_4 in the list above.

The gases CF_4 and DME (table 7.1) also possess quenching properties which, however, are also used as the main component of a mixture for other reasons. For CF_4 the reason can be the high drift velocity and for DME the high ionisation density at large radiation length and little diffusion (see the detailed discussion of gases for drift chambers in section 7.10.5).

Diffusion and drift velocity. The polyatomic gases are primarily employed as quenchers for the absorption of photons. At the same time the molecules have the capability of damping the heating of the electrons through the electric field by absorbing energy in inelastic collisions from the electrons and distributing that energy preferentially in non-radiative processes to other molecules. A high cross section with large inelasticity leads, as explained in section 4.6.4.3, to a small characteristic energy (eq. (4.112)) and a small diffusion coefficient (eq. (4.111)); see also fig. 4.12. The drift velocity is in some range tunable by choosing the relative fraction of the quencher in the gas (fig. 4.10) and reaches with increasing field strength a maximum followed by a decline. This saturation behaviour is exploited for drift chambers in order to keep the drift velocity, which is a function of the reduced field strength E/p or E/n (see eq. (4.113)), stable against fluctuations of field strength, pressure and temperature. Therefore the mixing ratio of noble gas and quencher is an essential optimisation parameter for drift chambers (section 7.10) which use the drift time for position measurements.

Electronegative additions. In order to stop unwanted discharges, in particular at high amplifications, one adds to the chamber gas electronegative gases, such as halogenated hydrocarbon (e.g. $CBrF_3$ or C_2H_5Br), sulfur tetrafluoride (SF_4) or also oxygen. These molecules adsorb electrons thus becoming negative ions which do not contribute to the amplification because of their much lower mobility. In detectors with longer drift paths these additions cannot be employed because this could cause inefficiencies in particle detection.

Additions to prevent polymerisation. A main reason for a limited lifetime of a gas-filled detector are deposits of polymerisates on electrodes caused by the hydrocarbons which are employed as quenchers (section 7.11.2). Additives which do not tend to polymerise but still have good quenching properties are alcohols, for example propanol (C_3H_8O) and methylal ($C_3H_8O_2$). The problem with alcohols is, however, that the saturation vapour pressure at normal temperatures is very low (for propanol 20 hPa at 20 °C), which means that high concentrations cannot be reached under regular conditions. But also with small additions of alcohol, from a tenth of a per cent to some per cent, a positive effect for the lifetime of a chamber can be reached which

is due to a charge exchange mechanism between the ions of the main gas component and the additives. Since the additives have a lower ionisation potential the charge is very effectively transferred from the quencher to the alcohol molecules. Thus the non-polymerising molecules finally neutralise at the cathode, leading to a longer lifetime of the chamber.

In order to increase the conductivity of the deposits on the electrodes one can add minor portions of water to the gas (see also section 7.11). Low water concentrations can be obtained by bubbling the gas through a water bottle (*bubbler*) at constant temperature. For example, a 0.7% water concentration is obtained at a temperature of 0 °C and a pressure of 1 bar, corresponding to the saturation vapour pressure at this temperature.

7.6 Operation of counting tubes

Counting tubes with cylindrical geometry, as in fig. 7.4, are the basis for most detector types which measure ionisation with gas amplification (see also section 7.2.2). A landmark development was the Geiger–Müller counter about a century ago [454]. While this counter type employs the Geiger mode, counting tubes employing the proportional mode subsequently became the rule because developments in electronics offered increasingly more sensitive signal processing.

In the following we describe operation and applications of counting tubes in different operation modes: proportional, Geiger and streamer mode. The choice of the respective gas filling follows the criteria discussed in the previous section and specifically the same criteria as for multiwire proportional chambers (section 7.8.3). In most cases the counting gas consists of a main component determining the ionisation density, a quencher gas for the absorption of photons and optionally an electronegative admixture preventing after-pulsing by fast removal of electrons. Typical gas mixtures have argon as main component and methane, ethane or isobutane as quencher, and possibly also an alcohol additive.

7.6.1 Proportional counter tubes

With modern electronics it is well possible to measure the generation of single electron–ion pairs in a gas with proportional gas amplification in a counter tube, as shown in fig. 7.4. A single electron yields at a gas amplification of 10^4 a charge of about 1 fC at the anode, which provides a sufficiently large signal if read out by low-noise electronics (chapter 17). The signal formation in cylindrical counting tubes is described in section 5.3.2. Operated in the proportional mode the detectors have practically no dead time and thus work at much higher counting rates than those operated at saturated amplification.

In contrast to the operation in a saturated mode (discharge, Geiger or streamer mode) the operation with proportional gas amplification offers the possibility to determine the energy deposition in the gas, which can be exploited for particle identification (section 14.2.2). In dosimetry a proportional counter can distinguish different kinds of radiation, for example β and α radiation. Also the characteristic lines of γ radiation can be made visible. Figure 3.41 shows the two lines of an ^{55}Fe source corresponding to the photoeffect and the escape peak which are often used for calibration of the gas amplification. With the knowledge of the average energy required for the generation of an electron–ion pair, the measured average charge of a gamma line yields the amplification factor and the line width yields the resolution. In addition, the proportionality

of the amplification can be controlled by observing the ratio of the distance of the lines to the distance from the zero point.

A typical application of proportional counters is radiation measurement. A special example is the measurement of β decays of radioisotopes which are generated by solar neutrinos in the corresponding neutrino detectors; see details in section 16.6.1. Proportional counters are also used for the detection of slow neutrons; see details in section 14.5 and fig. 14.28(b) on page 574. The principle of proportional counters has opened the way for the development of multiwire proportional chambers with which position measurements can be performed by arranging parallel wires in a plane (section 7.8).

7.6.2 Geiger–Müller counter

The Geiger–Müller counter, or simply *Geiger counter*, operates in the Geiger mode (corresponding to the plateau at high amplification in fig. 7.10, see also description on page 185) as a trigger counter, meaning that the output signals are independent of the deposited energy. The signals are large enough that they can be registered without further amplification, for example as sound from a loudspeaker.

The Geiger discharge, which spreads over the whole wire, will only be stopped when the ion cloud moving towards the cathode has sufficiently reduced the electric field in the amplification region. A renewed ignition of the gas discharge due to the ions hitting the tube wall is suppressed through the addition of a quencher to the counting gas. The amplification in the Geiger mode produces so much charge that the neutralisation takes a relatively long time during which the counter tube has no or reduced efficiency. This dead time, in the range of milliseconds, is an essential reason why the Geiger mode is not used much nowadays in the research area (see also section 17.9 on detector dead times).

7.6.3 Streamer tube

Streamer tubes and *limited streamer tubes* are counter tubes run in the streamer mode [551], as described in section 7.4.2. Applying a very high voltage to the electrodes the streamer mode is evolving as sketched in fig. 7.11. Relatively thick anode wires (50–100 µm) and a quenching gas (e.g. isobutane in the usual mixture argon–isobutane), which inhibits the expansion of the UV photons, distinguishes this operation mode from the Geiger mode. The streamer will be quenched by the space charge of the ion cloud (*self-quenching*, *limited streamer* mode).

Streamer tubes are often employed as relatively cost-efficient, robust solutions for large detector areas at relatively low counting rates, for example for muon detectors (see section 14.3). The large areas are achieved by tightly concatenating many parallel tubes to form a plane. Figure 7.12 schematically shows the structure of a typical streamer tube system. The plastic walls of the tubes form the cathode, which is made conductive by a resistive coating. The resistive coating is made of a soot-doped epoxy resin or another synthetic material. The required area resistivity can be adjusted by the amount of soot addition. The quadratic cross section of the cathodes permits cost-efficient manufacture from extruded plastic: first, the upper side of the tubes is left open such that a comb structure is obtained which will be closed after the wires have been strung.

Usually the signals which are induced onto the conductive strips attached to the outside of the tubes are read out. In the fig. 7.12 two independent coordinates are read

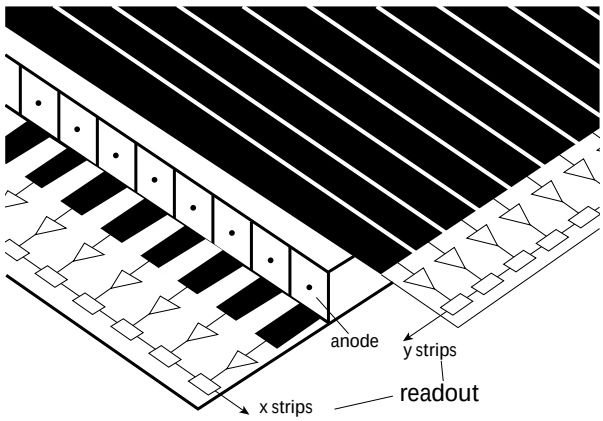


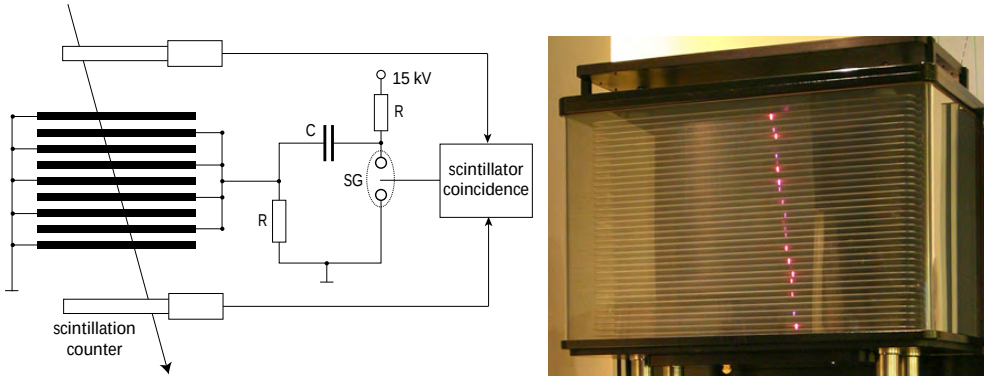
Fig. 7.12 Typical layout of a streamer tube system (adapted from [551], with kind permission of Elsevier). In plastic profiles with quadratic cross section and 1–2 cm edge length anode wires with a diameter of 50–100 μm are strung. The inner walls of the plastic tubes are made conductive by a resistive coating in order to serve as cathodes. As a rule, both electrodes are not read out. Instead conductive strips are attached to the outside of the tubes which pick up the induced streamer signals and pass them to amplifiers.

out via strips at the top and bottom of the tubes, but none of the electrodes (sometimes the anode is read out as well). Because of the high signals at amplifications of up to 10^{11} the strips can be relatively long, several metres, with a resulting high capacitance (the effect of the detector capacitance on the readout electronics is discussed in chapter 17).

7.7 Sparks and streamers in parallel-plate systems

Until the 1940s the only detectors with which particle tracks could be made visible were cloud chambers and photoemulsions (chapter 6) which, however, cannot be read out electronically. The possibilities to register particle tracks electronically, for example by arranging many Geiger counters together, was at that time still very limited. For the coverage of large detection volumes, the principle of gas amplification was applied for the first time at the beginning of the 1950s when spark chambers were developed (section 7.7.1), which provided photographic pictures of sparks along a particle track. Compared to bubble chambers, which were introduced at about the same time, they have the advantage that they can be triggered by traversing particles with the aid of external counters. But they have the disadvantage that information on the ionisation density, which is exploited in bubble chambers for particle identification, is lost in the spark formation. The most important disadvantage of spark chambers, however, is the long recovery time after a discharge. In the 1960s this was partly remedied by the introduction of detectors working in streamer mode (section 7.7.2). However, both detector types have been replaced nearly everywhere by gaseous detectors which operate in the proportional regime and are thus virtually dead-time-free. Higher gas amplification modes are now only employed for specific applications, such as spark chambers for demonstration purposes (fig. 7.13) or streamer discharges over short distances with high time resolution (as in resistive plate chambers, section 7.7.3).

In the following, spark and streamer chambers will only be dealt with briefly because they are mostly of historical interest. However, spark chambers are still frequently used for demonstration purposes (as in fig. 7.13(b)) because of their impressive optical and acoustic presentation of particle tracks and because of their relatively



(a) (b)

Fig. 7.13 (a) Typical layout of a spark chamber system with several parallel plates between which the sparks form (adapted from [816,817]). In the initial state the capacitor C is charged up by the applied high voltage. At the passage of an ionising particle a coincidence of scintillation counters triggers the discharge of the capacitor onto the plates by firing a spark gap (SG). (b) Between the plates sparks become visible along the ionisation trace. This detector was set up for demonstration purposes in order to visualise particles from cosmic radiation, mostly muons. Source DESY.

simple construction. Resistive plate chambers will be discussed in some more detail because they are employed in modern experiments for fast triggering and for time measurements.

7.7.1 Spark chambers

Figure 7.13(a) shows a typical layout of a spark chamber system with several parallel plates between which high voltage is applied leading to sparks between the plates when ionising particles have passed. The chambers are filled with noble gases, typically a helium–neon mixture (30:70) resulting in red-violet glowing sparks. The particle passage is registered by fast detectors, typically plastic scintillators (see chapter 13) connected to a coincidence circuit. The coincidence signal triggers a switch which initiates the injection of a charge pulse stored on a capacitor at high voltage (about 10–20 kV) onto the plates. For the fast switching of very high, short current pulses, triggerable spark gaps[3] are employed. Figure 7.13(b) shows how a particle track becomes visible through the sparks.

For scientific applications the sparks are photographed from different directions, allowing for three-dimensional reconstruction of the track. Another readout method exploits the fact that a spark discharge always generates a heated plasma from which sound waves emanate. The track coordinates can be reconstructed by electronically measuring the arrival times of the sound by two or more microphones at different positions. Compared to photography this acoustic method has the advantage that fully electronic registration of events is possible.

Another possibility for electronic registration is the so-called magnetostrictive readout which is schematically depicted in fig. 7.14. The method implies that the plates

[3]Spark gaps are electronic devices which are used for fast switching of high voltage pulses. They work like a spark plug in a combustion engine.

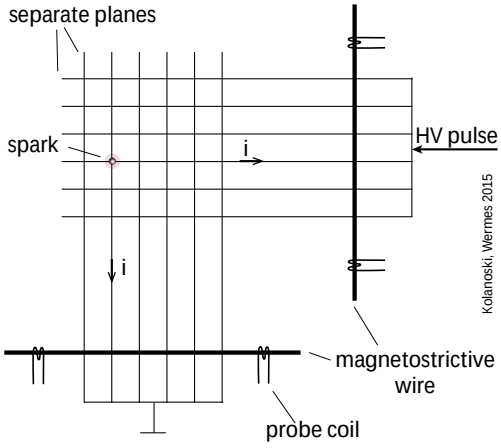


Fig. 7.14 Magnetostrictive readout of a spark chamber. The high voltage planes of the spark chamber are here formed by wires. A spark generates current pulses (i) on both affected wires which excite acoustic waves on the (electrically insulated) magnetostrictive wires. The acoustic waves propagate to the ends of the wire where they are converted back to current signals by pick-up coils. The propagation time difference between both ends is a measure of the spark position.

where the sparking occurs are replaced by wire planes (*wire spark chamber*) with typical wire separations of 1–2 mm in the plane and about 1 cm between the planes. A spark generates a current pulse on the wire which will be routed over an electrically insulated magnetostrictive wire[4] stretched along the border of the chamber. In this wire a mechanical wave is generated by the magnetic field of the current pulse. In the reverse process current pulses will be generated in pick-up coils at the wire ends. The comparison of the arrival times of the pulses at both ends of the magnetostrictive wire yields the localisation of the hit signal wire and thus of the passing track in one coordinate. Through the orientation of the signal wires in different directions in the various planes the particle trajectory can be spatially reconstructed. This readout method avoids the electronic complexity connected to a readout of individual wires, although at the expense of a reduced resolution.

The time constant for the charging of a spark chamber lies in the microsecond range, a time span in which the ionisation charges between the plates are still remaining. In contrast the time constant for recharging the capacitor by which the chamber is charged (C in fig. 7.13(a)) lies in the order of milliseconds leading to correspondingly large dead times. Another disadvantage of spark chambers is the loss of information on the primary ionisation due to the operation in discharge mode.

7.7.2 Streamer chambers

In a streamer chamber [358, 817] the development of a spark is avoided (see fig. 7.11 with the corresponding description in the text) by applying very short high voltage pulses to the electrodes with duration in the range of 10 ns. Thus streamers can be kept so short, with lengths of less than a millimetre, that they effectively provide points in three dimensions on the particle track. Because of this feature the geometry of a streamer chamber is basically different from that of a spark chamber. In streamer chambers a high voltage is also applied between parallel plates, in this case with sensitive volumes of up to several cubic metres. However, here the preferential direction of the particles is parallel to the plate electrodes and thus perpendicular to the electric field. Hence the streamers preferentially develop perpendicularly away from the

[4] Magnetostriction: change in length in a magnetic field.

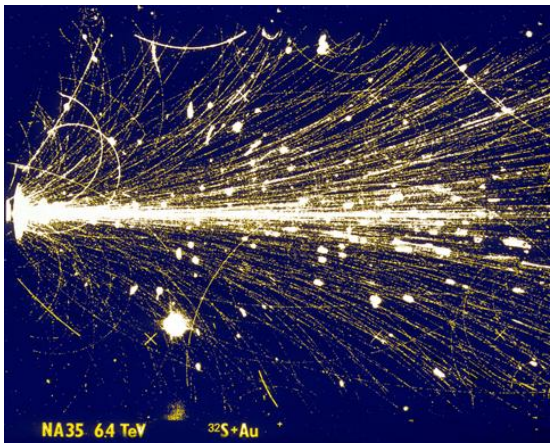


Fig. 7.15 Recording of a reaction between a sulfur and a gold nucleus at a total energy of 6.4 TeV [720]. The picture was taken with the streamer chamber of the heavy ion experiment NA35 at the SPS (CERN). Source: CERN.

track. The registration of streamers is only possible optically, similar to bubble chambers (section 6.2). In contrast to a bubble chamber, however, the photographs can be selectively triggered.

A streamer chamber is filled with a gas that has mostly neon as main component, often with a fraction of helium (typically 80–90% neon, 10–20% helium), and a small admixture of electronegative gas, for example freon or SF_6 (see explanations in section 7.5). In neon the streamers are particularly bright [358] with a maximal intensity at wavelengths around 640 nm. For example, the streamer chamber of the heavy ion experiment NA35 at CERN (fig. 7.15) used the gas mixture He–Ne (20:80) with 0.25% isobutane and 0.05 ppm SF_6. With the small additions of isobutane as quencher and SF_6 as electron absorber the chamber achieved dead times of only about 10 µs.

The photograph of a heavy ion reaction in fig. 7.15 demonstrates that streamer chambers are able to image very high particle densities, which is not achievable to this extent in spark chambers because there the discharge current is concentrated on the spark that developed first. The charge of a single streamer is saturated and hence carries no information on the initiating primary ionisation. However, the streamer density along a track is proportional to the primary ionisation density and can thus be used for particle identification. This is a further advantage of the streamer mode as compared to the spark discharge mode.

7.7.3 Plate electrodes with high resistivity

Discharges in a homogeneous electric field between parallel plates with a narrow gap have very small time fluctuations, in the order of only few multiples of 10 ps. While in a counting tube the signal only develops after the electrons have drifted into the high field regions near the anode, in a homogeneous field the avalanche can be initiated at any point, thus avoiding fluctuations arising from the drift time. In normal spark chambers, however, this good time resolution cannot be exploited because they are operated in a pulsed mode where the start time of the spark development depends on the externally triggered high voltage pulse. In continuous operation with constant high voltage, on the other hand, uncontrolled discharges with subsequent dead times would make a chamber very inefficient. Therefore, in order to be able to exploit the potentially good time resolution in parallel plate geometries, methods of discharge limitation (see list on page 186) have been developed, as we will discuss in the following.

7.7.3.1 Pestov counter

In the early 1970s the so-called *Pestov counter* was introduced [761] which could be operated in a continuous mode, indeed reaching time resolutions of 25 ps [780]. A Pestov counter consists of two parallel plates about 100 µm apart in a gas volume. The gas is kept under a pressure of about 10 bar in order to increase the ionisation probability of traversing tracks. Voltages of about 10 kV are applied to the plates such that a spark or streamer develops when an ionising particle passes the gas volume. It is essential for continuous operation that at least one of the two electrodes has a high resistance such that a discharge current over the resistor leads to a voltage drop, thereby terminating the discharge. See also the corresponding discussion of (7.29) on page 199. Pestov counters are constructed using glass plates with a resistivity of 10^9–$10^{11}\,\Omega\,\mathrm{cm}$ [761,211] (*Pestov glass*, a glass with semiconductor properties). Thanks to the high resistivity of the plates the signals can be extracted by induction onto conducting strips outside of the plates (as for the streamer tubes, fig. 7.12).

7.7.3.2 Resistive plate chambers

Pestov counters are very delicate to construct and operate, not least because of the small tolerances for the plate distance and the operation at high pressure. A further development has led to *resistive plate chambers* (RPCs) [847] with reduced requirements on construction and operation. The key point of operation is a sufficiently high electric field such that instant avalanche multiplication sets in at any point of the gas volume with negligible drift before, thus fulfilling the conditions for high time resolutions and fast trigger decisions. Meanwhile RPCs are established as a cost-efficient solution for large-area particle detection with good time resolution. They are used in many experiments at accelerators, for example at the LHC [4, 298, 10, 87], in detectors for cosmic radiation [138] and in neutrino experiments [186,175]. In particular, they are employed for the measurement of muons and for triggering with them (e.g. in ATLAS [4] and CMS [298]) as well as for time-of-flight measurements (e.g. in ALICE [10]).

Different to Pestov counters, RPCs mostly operate at ambient pressure. The gaps between the plates are larger, typically ranging between 0.3 mm for time critical applications and 2 mm for applications requiring a high detection efficiency. Bakelite, a phenol formaldehyde resin, has proven to be a robust and easy to machine plate material with high resistivity.[5]

Single-gap RPC. We begin with the discussion of an RPC with a single gas gap (single-gap RPC) for triggering applications. The principle is shown in fig. 7.16. The plates are made of a high-ohmic material with a resistivity of 10^8–$10^{12}\,\Omega\,\mathrm{cm}$ (e.g. glass or bakelite) and have a typical distance of about 2 mm. A high voltage of about 10 kV is applied between the plates. On the sides facing away from the gas volume the plates are coated with low-conductivity graphite with a surface resistivity[6] of $R_\square \approx 10^5\,\Omega$.

[5] To avoid sparking the surface has to be microscopically smooth which is achieved by a treatment of the surface with linseed oil.

[6] The surface resistivity (also called sheet resistance) $R_\square$ is the resistivity (also called specific resistance) ρ divided by the thickness of a (homogeneous) layer: $R_\square = \rho/d$. This quantity can be measured as the resistance of a quadratic area between two parallel edges. Since the resistance is proportional to the distance of the edges and reversely proportional to the lengths of the edges, which for a square have the same lengths, the measurement is independent of the size of the square. Therefore the surface resistivity has the dimension ohms, sometimes also expressed as 'ohms per square' ($\Omega/\square$) in order to indicate that surface resistivity is meant.

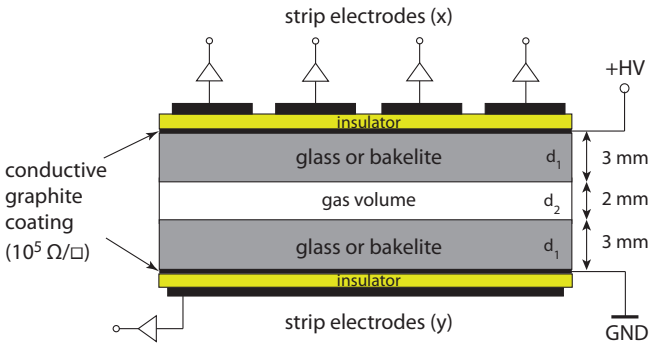


Fig. 7.16 Layout of an RPC. The high voltage is applied between the two high-resistive electrodes (glass or bakelite) via conductive graphite layers. The signals are induced onto readout strips with different orientation on top and bottom. The strips are insulated from the voltage electrodes. Typical dimensions are given for illustration (adapted from [664]).

The readout planes on the outside are separated from the conducting graphite coating by insulating layers. The readout plane picks up the induction signals of the discharges in the gas volume. It is usually subdivided into parallel conducting strips, which can be oriented on both sides in different directions in order to obtain a position measurement. The induction signals reach through to the readout strips because of the relatively low conductivity of the graphite layer. The readout electrodes are decoupled from the high voltage and can be put, for example, on ground potential such that the subsequent electronics can be operated at low voltages.

RPCs can be operated in streamer mode or, with somewhat reduced voltage, in avalanche mode (see section 7.4.2 on page 186) with different performance properties, in particular what their rate and timing capability is concerned. Gas additives can additionally suppress streamer formation when operation in avalanche mode is desired. In streamer mode operation, the signals typically range from 100 pC to some nC; in avalanche mode they are typically one order of magnitude smaller. Due to the smaller signal charge generated in avalanche mode the rate capability is distinctly better than in streamer mode (see discussion below). However, in the homogeneous electric field an avalanche can start at any point in the gap, which makes gas amplification in this mode dependent on the distance of the primary ionisation from the anode according to (7.15). The typically small numbers of ionisation clusters in the narrow gap correspondingly lead to (large) signal fluctuations and possible efficiency loss. In streamer mode this is less of a problem since the charges of the signals are saturated anyway.

Regarding the choice of RPC gas mixtures, high ionisation density and sufficient quenching power of the avalanche or the streamer are very important. Initially, RPCs were operated mostly with argon gas, optionally with an electronegative addition [847]. For the RPCs of the LHC experiments, mixtures based on the freon tetrafluoroethane ($C_2H_2F_4$, code name R134a) with isobutane as quencher and SF_6 as electronegative addition are common (e.g. in ATLAS: 94.7% $C_2H_2F_4$, 5% i-C_4H_{10}, 0.3% SF_6 [4,418]). In [856] the primary ionisation density of tetrafluoroethane is estimated to be $n_p = 8.2$/mm and thus much higher than that of argon with $n_p = 2.9$/mm (table 7.1). Tetrafluoroethane is not flammable at abient temperatures and does not attack the ozone layer of the atmosphere, but it is a relatively strong greenhouse gas. Since this gas may be banned in the future because of its environmental impact, intensive

Table 7.3 Properties of materials used as electrodes in Pestov counters or RPCs (ρ = resistivity, ε = relative permittivity, τ = recovery time).

Material	ρ (Ω cm)	ε	τ (s)
glass	$\approx 10^{12}$	4.4–5.4	≈ 0.5
bakelite	$\approx 10^{10}$	7.6–8	≈ 0.01
Pestov glass (Schott S8900) [211]	$\approx 10^{11}$	6–8	≈ 0.1

searches for 'eco-friendly' replacements are underway (see e.g. [856, 123]). See also the discussion of ageing effects for RPCs in section 7.11.1.

For streamer quenching, as well as to prevent streamers occuring in avalanche mode, the following mechanisms contribute [272]:

- local 'switching off' of the field around the point of discharge achieved from the voltage drop over the high electrode resistance;
- absorption of UV photons by gas additives to the gas (often isobutane) to prevent secondary discharges;
- capture of electrons from the discharge by an electronegative component in the gas, usually sulfur hexafluoride (SF_6), in order to confine the extent of the discharge.

Recovery time after a discharge. Because of the high resistivity of the (bakelite or glass) electrodes the charge drain is delayed and therefore a counter-field builds up, limiting the discharge in time and strength, an effect which is referred to as *self-quenching*. The duration of the discharge is typically about 10 ns. The time for the restoration of the electric field in the region of discharge, however, can last up to a second, that is, the time for the charge created near the glass or bakelite electrode to 'decay away'. It is characterised by the electric properties of the electrode material (ρ and ε) and the gas gap.

The development of the charge 'decay' between the initial charging at a point on the resistive layer and its final spreading over the entire layer has been studied in [667, 821]. An estimate of the characteristic time constant τ can also be obtained from the capacitances and resistances involved. With $R = \rho\, d_1/A$, where ρ, d_1 and A are resistivity, thickness of the electrode layer and (charged) area, respectively, and with the capacitance $C = C_1 + C_2$, formed by the electrode layer and by the gas gap, $\tau = RC$ becomes

$$\tau = \rho\, \varepsilon_0 \left(\varepsilon_1 + \frac{d_1}{d_2} \varepsilon_2 \right) , \tag{7.29}$$

where $d_{1,2}$ and $\varepsilon_{1,2}$ are thicknesses and permittivities of electrode (bakelite or glass) and gas gap, respectively. Note that the area A cancels out in the product RC so that it need not be specified. Assuming typical values of $d_1 = 3$ mm and a gap width $d_2 = 0.3$ mm, $\varepsilon_1 = 8$, $\varepsilon_2 = 1$, and $\rho = 10^{12}\,\Omega$cm one obtains $\tau = 1.6$ s. More detailed calculations in [821] yield 'decay times' between 0.8 s and 1.6 s.

For some materials, table 7.3 shows the recovery times during which the electrical field around the position of large charge deposit is reduced. This introduces a local low efficiency area with a typical size of about 0.1 cm^2 for a streamer signal [194]. Since the other regions of the RPC remain fully efficient, the detector can thus tolerate rates of the order of 10 Hz/cm^2.

The rate tolerance is determined by the average current $\langle I \rangle$ through the electrodes leading to a drop of the external voltage V_{ext} over the resistive layer:

$$\Delta V = V_{ext} - V_{gas} = R\,\langle I\rangle\,, \tag{7.30}$$

which should remain negligible as compared to the voltage over the gas gap V_{gas}. Obviously, this voltage drop depends on the electrode resistivity ρ (for two electrode layers) and the discharge current, given by the average charge $\langle Q\rangle$ produced per avalanche or streamer and the particle rate per unit area f_A:

$$\Delta V = \rho\,(2d_1)\,\langle Q\rangle\,f_A. \tag{7.31}$$

The rate tolerance, definable as $f_A/\Delta V$ [66], can hence be increased by lowering the electrode resistivity (table 7.3) and thickness, but also by reducing the gas amplification, that is, changing the operation mode.

Originally RPCs were operated in streamer mode with correspondingly large signals and fewer demands on the readout electronics and the accuracy of the electrode distance. The operation in avalanche mode offers better rate tolerance and less ageing problems (see section 7.11) but requires electronic amplification and more elaborate readout electronics as well as fine-tuning of the gas amplification. Dedicated developments—amongst others of special gas mixtures—accomplished RPCs operating in avalanche mode with gap widths in the millimetre range that could reach near 100% efficiencies at particle rates of up to some kHz/cm^2. Examples are the RPCs of the LHC experiments ATLAS [4] and CMS [298].

In the ATLAS experiment single-gap RPCs are employed as trigger detectors for muons covering an area of 3650 m^2 in three layers, each with two single gap planes [4]. For trigger detectors the detection efficiency is the essential criterion, whereas the time resolution is usually less important. In collider experiments the latter must only be better than the time difference between beam crossings (25 ns at LHC). Operated in avalanche mode the ATLAS RPCs reach detection efficiencies of 98.5% at particle rates of about 1 kHz/cm^2. The position resolution is about 1 cm, achieved by subdividing the electrodes into strips.

Time resolution of RPCs. The time stamp of a signal is set when it exceeds the threshold of a discriminator. The time until which the signal amplitude reaches the threshold varies with the time fluctuations of the avalanche formation. While in cylindrical geometries an avalanche will always develop at the wire, in the homogeneous field of an RPC an avalanche can form anywhere in the gas gap. The gas amplification $G = \mathrm{e}^{\alpha s}$ (eq. (7.15)) depends on the distance s of the ionisation from the anode. The Townsend coefficient α (ion–electron pairs per path length, eq. (7.10)) is constant in a homogeneous field. The gas amplification is adjusted such that primary ionisation occurring up to a certain distance from the cathode delivers a signal above the threshold. Hence at least one charge cluster has to lie within this distance for a signal to be registered. For example, if a detection efficiency of 98% is aimed at, the average number of clusters in the region near the cathode has to be four.[7]. The cluster density of 3–4/mm in typical chamber gases can be increased by employing respective gases with higher ionisation density or by increasing the pressure (as done in Pestov counters). For example, the cluster density of Freon 13B1[8] (CF_3Br) is about 10/mm [286].

[7] According to Poisson statistics the probability to have no cluster in this region is $p(0) = \mathrm{e}^{-\lambda}$ for an expectation value λ of the number of clusters. For an efficiency of 98%, that is, $p(0) = 0.02$, one obtains $\lambda \approx 4$.

[8] In the past, Freon CF_3Br was also used, in particular as an electronegative additive. However, it is no longer produced because of its high potential for ozone reduction in the atmosphere.

Tetrafluoroethane (R134a), which is the main component of the RPC gas used in LHC experiments, has a cluster density of 8.2/mm [856].

In an avalanche the characteristic time $\tau_\alpha = 1/(\alpha\, v_D)$ between multiplication steps is given by the Townsend coefficient α and the drift velocity v_D of the electrons. Effectively, α is reduced by attachment of electrons to the molecules of the chamber gas. The relevant quantity is the *effective Townsend coefficient* $\alpha - \eta$, where η is the attachment coefficient (see eq. (7.9) on page 178).

The time evolution of RPC signals is treated in [690] and in [825] on the basis of the physical properties of avalanche formation in gases for different chamber parameters (see section 7.4.1). Each electron created by ionisation in the chamber gap can start an avalanche with some probability or can be attached to a gas molecule and is hence lost. If amplified, the induced signal, which determines the timing precision, is given by the avalanche cloud started from an initial charge size Q_0 of a given ionisation cluster. If multiple clusters are generated within a gap thickness their contribution must be folded in. The dominant contribution to the time evolution of an (induced) signal is governed by the avalanche development until its arrival at the electrode plane ($t < x/v_D$). With (7.15) for the gas amplification, the exponential growth of the avalanche charge for a given drift velocity v_D can be written as

$$Q(t) = Q_0\, \mathrm{e}^{\,(\alpha - \eta) v_D t}\,. \tag{7.32}$$

Other time contributions to the signal development, like for example the time needed for the (position-independent) development of the induced signals on the electrodes, can be neglected in comparison. For a first order estimate one can adopt the view that the initial charge size Q_0 in (7.32) is given by primary ionisation charges which are immediately multiplied by avalanche processes initiated from every electron. The main fluctuations of $Q(t)$ at some fixed time t arise from primary fluctuations at the beginning of the avalanche process when the number of electrons is still small. Later in the avalanche, when much larger numbers of electrons govern the exponential growth, the development of $Q(t)$ becomes increasingly smoother and the initial fluctuations can be absorbed in the charge starting value Q_0 which itself follows a statistical distribution. It is shown in [825] that the probability distribution of Q_0 follows a decaying exponential function with an average value Q_0^{av}:

$$P(Q_0) = \frac{1}{Q_0^{\mathrm{av}}} \exp\left(-Q_0/Q_0^{\mathrm{av}}\right)\,. \tag{7.33}$$

In order to examine the time resolution we look for fluctuations in time occurring when a signal crosses a given threshold Q_{thr}. With (7.32) we obtain

$$Q(t) = Q_{\mathrm{thr}} \quad \Rightarrow \quad t(Q_0; Q_{\mathrm{thr}}) = \frac{1}{(\alpha - \eta)\, v_D} \ln \frac{Q_{\mathrm{thr}}}{Q_0}\,, \tag{7.34}$$

with the characteristic time $\left((\alpha - \eta)\, v_D\right)^{-1}$ for avalanche formation (average time between secondary ionisations). Using (7.33) that Q_0 is exponentially distributed around some average, one can derive the time-stamp distribution as [825]:

$$P(t) = (\alpha - \eta)\, v_D f\left((\alpha - \eta)\, v_D t\right) \qquad \text{with} \qquad f(x) = \mathrm{e}^{-(x\, +\, \exp(-x))}\,, \tag{7.35}$$

given here such that the maximum of the distribution has been shifted to $t = 0$.

Surprisingly, the shape of the distribution is independent of the threshold and the average signal, since a different threshold merely corresponds to a time shift. The function $f(x)$ is asymmetric, resembling in shape the Landau distribution function (see eq. (3.55)). The variance of $f(x)$ about the mean value is $\sigma^2 = (1.28)^2$, that is, in the order of one. Thus the time resolution is

$$\sigma_t \approx \frac{1.28}{(\alpha - \eta)\, v_D} \,. \tag{7.36}$$

RPC efficiency. One can estimate the hit efficiency on an RPC by making some simplifications. Following [825] we assume that the RPC 'fires' if the first ionisation cluster creates an avalanche exceeding the threshold, if the first cluster is lost (attached) and the second cluster exceeds the threshold or if both first and second cluster are lost and the third fires the RPC, and so on. We further assume that the clusters contain only one electron and neglect avalanche fluctuations. Then, a primary electron created at position x in a gap starting at $x = 0$ with thickness d will induce a charge

$$Q_{ind}(x) = E_w \frac{e}{\alpha - \eta} \left(\mathrm{e}^{(\alpha-\eta)\,(d-x)} - 1 \right) \tag{7.37}$$

on the readout electrode, where E_w is the weighting field as introduced in section 5.3.1 (for unstructured electrodes $E_w = 1/d$ is a constant) and α and η are the respective coefficients for amplification and attachment, as used above. For a set threshold Q_{thr} an avalanche is detected if for some x_{thr} the induced charge is larger than the threshold: $Q_{ind}(x) > Q_{thr}$. This means that the distance $d - x_{thr}$ to the collecting electrode must be large enough for sufficient amplification; solving (7.37) this is for

$$d - x_{thr} > \frac{1}{\alpha - \eta} \ln \left(1 + \frac{1}{E_w} \frac{Q_{thr}}{e} (\alpha - \eta) \right) \,. \tag{7.38}$$

We can assume that a primary ionisation cluster does not depend on the previous ionisation. Hence the distance δx between ionisation clusters for a given mean ionisation length λ is exponentially distributed:

$$P(\delta x) = \frac{1}{\lambda} \mathrm{e}^{-\frac{\delta x}{\lambda}} \,. \tag{7.39}$$

The probability that an electron is attached before it can be amplified is η/α and, correspondingly, the probability that the electron is amplified (i.e. not attached) is $1 - \eta/\alpha$. Then the probability that the first electron induces a signal above threshold is obtained by integrating the probability distribution (7.39) from zero to x_{thr}, defined in (7.38). Accounting for the amplification and attachment probabilities we obtain

$$P_1 = \left(1 - \frac{\eta}{\alpha} \right) \int_0^{x_{thr}} \frac{1}{\lambda} \mathrm{e}^{-\frac{x_1}{\lambda}} \, dx_1 \,. \tag{7.40}$$

The probability that the first cluster does not reach the electrode, but the second one does and is above threshold, then is

$$P_2 = \int_0^{x_{thr}} \int_0^{x_2} \left(\frac{\eta}{\alpha} \frac{1}{\lambda} \mathrm{e}^{-\frac{x_1}{\lambda}} \right) \left(1 - \frac{\eta}{\alpha} \right) \frac{1}{\lambda} \mathrm{e}^{-\frac{x_2 - x_1}{\lambda}} \, dx_1 \, dx_2 \,. \tag{7.41}$$

Table 7.4 Typical parameters for trigger and timing RPCs together with the resulting values for the time resolution and the efficiency as derived from (7.36) and (7.43). For the trigger RPC an average cluster size of 2 has been assumed for illustration, halving λ.

Type	E $\left(\frac{\text{kV}}{\text{cm}}\right)$	λ (mm)	α $\left(\frac{1}{\text{mm}}\right)$	η $\left(\frac{1}{\text{mm}}\right)$	v_D $\left(\frac{\mu\text{m}}{\text{ns}}\right)$	d (mm)	Q_{thr} (fC)	σ_t (ns)	ϵ (%)
trigger RPC	50	0.06	13.3	3.5	140	2	100	1	99
timing RPC	100	0.1	123	10.5	210	0.3	20	0.05	78

Continuing the series, the probability for the nth cluster to be above threshold and one before, the (n-1)th cluster, to be attached is

$$P_n = \left(\frac{\eta}{\alpha}\right)^{(n-1)} \left(1 - \frac{\eta}{\alpha}\right) \left(1 - K\left[\frac{x_{thr}}{\lambda}, n\right]\right) \quad \text{with} \quad K[r,n] = \mathrm{e}^{-r} \sum_{k=0}^{n-1} \frac{r^k}{k!}\,. \tag{7.42}$$

The efficiency then is given by the sum of all probabilities P_n:

$$\epsilon = \sum_{n=1}^{\infty} P_n = 1 - e^{-(1-\frac{\eta}{\alpha})\frac{d}{\lambda}} \left(1 + \frac{\alpha - \eta}{E_w} \frac{Q_{\text{thr}}}{e}\right)^{\frac{1}{\alpha\lambda}}\,. \tag{7.43}$$

This formula agrees quite well with Monte Carlo simulations [825], although it underestimates the true efficiencies somewhat since accumulating cluster contributions are not accounted for.

In conclusion, note that the time resolution (7.36) depends only on the difference $\alpha - \eta$ of the gas-dependent Townsend and attachment coefficients and the drift velocity v_D, but not on the detection threshold Q_{thr} because each primary cluster contributing to the time measurement is by definition above the threshold and each avalanche contributes with the same time dependence. The detection efficiency ϵ in (7.43), however, depends on α and η separately as well as on Q_{thr}, d and λ.

As a rule, depending on the particular application, the efficiency ϵ (for trigger RPCs) or the time resolution σ_t (for timing RPCs) are optimised. Table 7.4 shows typical parameters for both cases as well as the resulting time resolutions and efficiencies as calculated using the derived formulae [825]. The good time resolution of RPCs with small gap width is achieved because of the strong electric field leading to large values of αv_D and correspondingly small time-scales which determine the timing precision. However, the efficiency of timing RPCs is distinctly lower than that for trigger RPCs. This results from the small gap and fewer ionisation clusters in combination with a reduction by space charge effects [824, 666].

Multi-gap RPCs. In order to further improve the time resolution and the efficiency of RPCs, so-called *multi-gap RPCs* have been proposed [286] with multiple avalanches inducing signals on the same readout strip. The layout of an RPC with several gas gaps is shown in fig. 7.17. The high voltage is applied through the graphite coating of the two outer plates, while the intermediate plates separating the gas gaps are electrically not connected and their potential is free floating. If no current flows to and from the intermediate layers a linear potential drop over all inner layers will be maintained. Because of current conservation in the stationary case, the incoming and outgoing currents must be equal, which requires equal gas amplification in neighbouring gas gaps thus stabilising the linear potential drop [286]. The time resolution improves

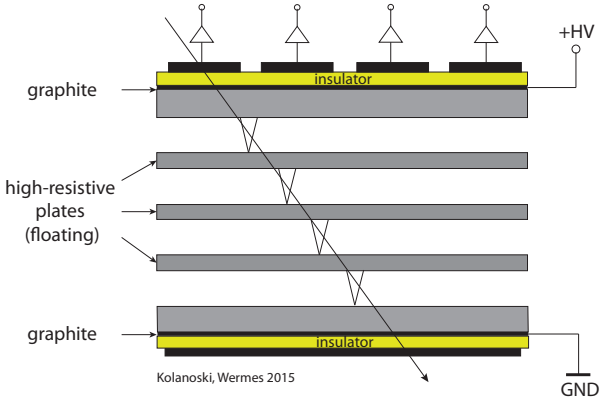


Fig. 7.17 Typical layout of a multi-gap RPC. This RPC variant offers superior time resolution at higher efficiency. See text for details.

with the number of gaps n, although not according to a simple $1/\sqrt{n}$ law since the resolution is rather dominated by the gap with the largest signal. For $n \approx 10$ intrinsic time resolutions of 20 ps are achievable [665].

In the ALICE experiment multi-gap RPCs are employed as time-of-flight detectors for particle identification (see also section 14.2.1 on page 548). With the choice of very thin gas gaps of 250 µm width and precisely positionable glass plates as resistive electrodes the experiment reaches time resolutions of about 85 ps at efficiencies of 99% [80, 10]. Such resolutions can also be reached employing organic scintillators with photomultiplier readout (section 13.2), however, at appreciably higher costs.

7.8 Multiwire proportional chambers (MWPCs)

A multiwire proportional chamber (MWPC) principally functions like proportional counter tubes that are arranged in a plane side by side. Such an arrangement provides spatial resolution perpendicular to the wires for particles passing the wire plane. With several MWPC planes stacked behind each other the trajectories of charged particles can be electronically registered (in contrast, for example, to bubble chambers which are read out photographically). The MWPC was developed by George Charpak at CERN in the late 1960s [296]. This detector type and variants of it, like the drift chamber, have been crucial for progress in particle physics until today. George Charpak received the Nobel Prize in 1992 'for his invention and development of particle detectors, in particular the multiwire proportional chamber' [290].

Important for the development was the realisation that a mechanical separation between the anode wires, for example by grounded walls, is not mandatory in order to clearly separate the signal from the wire where the avalanche occurs from those of the neighbouring wires (see e.g. [291]). Thus every wire can be considered as an individual detector, like a proportional counter tube. This can be understood by noting that the signal charge is only generated very near to the anode wire and that the highest signals are induced by drifting ions near the amplification region (see section 5.3.2). In this region the weighting fields of neighbouring wires are small and lead to positive induced signals, in contrast to the negative pulses on the central wire. The situation roughly corresponds to the weighting fields very near a strip anode in fig. 5.14. For a typical MWPC the positive signals are about 1/5 of the negative signals and can be discarded by appropriate electronic thresholds. Hence the walls of the counting tubes can be omitted thereby saving construction material. The development of MWPCs,

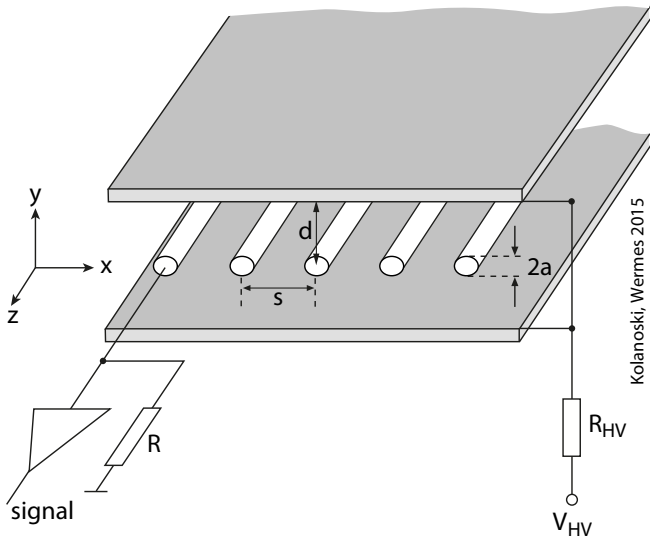


Fig. 7.18 Perspective view of a typical MWPC (schematic). In the shown version the cathodes consist of two parallel conducting planes at a distance of $2d$. The anode wires with radii a (drawn strongly magnified) are stretched parallel in the middle plane between the cathodes with a pitch s. Typical values for a, s and d are given in (7.44).

Table 7.5 Typical properties of wires which are used as anodes or cathodes, respectively. In order to apply a constant, reproducible tensile stress to a wire, a mass (last column) is attached to the wire running over a deflector pulley.

	Material example	Diameter	Tensile stress	Mass
anode:	gold-coated tungsten	20–30 μm	≈ 500 MPa	≈ 20 g
cathode:	Cu–Be alloy	≈ 100 μm	≈ 100 MPa	≈ 60 g

with which large detector volumes can be filled, made the fully electronic detection of complex high energy reactions possible for the first time. An essential advantage compared to spark chambers (section 7.7.1) is the much shorter dead time due to the fact that the small signals in the proportional regime cause little voltage drops and correspondingly short recharging times of the electrodes. In return the signal has to be amplified at each individual wire. The large number of electronic readout channels only became possible when electronic components became more compact.

7.8.1 Layout of MWPCs

A typical multiwire proportional chamber (fig. 7.18) consists of a plane of anode wires which is sandwiched between two cathode planes with a distance $2d$ between the cathodes. The wires are stretched in parallel with a pitch s, typically with $s \approx 2\,\text{mm}$. The distance d between the anode and the cathode planes is typically three to four times the distance s between the anode wires. The cathodes could be made of smooth metal plates, for example copper, or could also be formed by wires. Common wire materials with typical diameters and the applied tensile stress they are tensioned with, are listed in table 7.5. A typical set of parameters which we want to refer to in the following is (see fig. 7.18):

$$a = 0.01\,\text{mm}\,, \quad s = 2\,\text{mm}\,, \quad d = 8\,\text{mm}\,. \tag{7.44}$$

The electrostatic forces which arise due to the voltage between anode and cathode

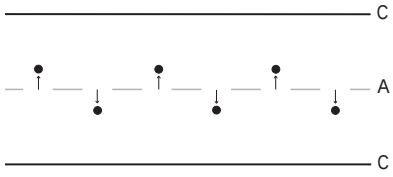


Fig. 7.19 Electrostatic instability in MWPCs. The anodes lie on the same potential and thus repel each other. At a critical length of the wire the mechanical tensile stress does not suffice to keep the wires in their nominal positions. A new stable configuration is found when neighbouring wires are deflected in opposite directions.

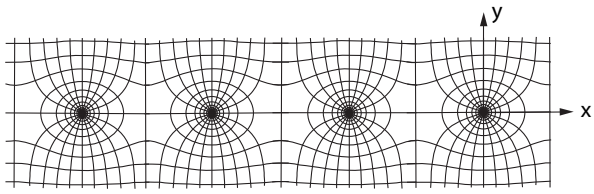


Fig. 7.20 Typical behaviour of potential and field lines in an MWPC (using the approximation for line charges in free space corresponding to (7.46)).

have to be compensated by the mechanical tension of the wires. Between the wires of a plane repulsive forces act which lead to instabilities if the wire tension is insufficient or the wires are too long. As shown in fig. 7.19, the system finds a stable state again if the wires are deflected in alternate directions out of the plane. Since the mechanical tension should not exceed the elastic limit (yield point), a maximal free length of the wire results which for the anodes is about 80 cm. For longer wires intermediate supports are necessary, for example by nylon threads stretched perpendicular to the electrode wires. The support frame for the wires has to be very stable in order to hold the mechanical tension forces. Often this frame also serves as gas container and has therefore to be made gas tight with as little material as possible in the sensitive area where the particles pass the detector.

7.8.2 Electrostatics

Applying a high voltage between the anodes and cathodes leads to an electrostatic field which near the anode almost has the same behaviour as in a cylinder capacitor and turns near the cathode into a nearly homogeneous field, as in a parallel plate capacitor (fig. 7.20). The high voltage is chosen such that in the region near the anode a field strength is reached which is necessary for the required gas amplification (typically $>100\,\mathrm{kV/cm}$ on the anode surface).

Calculation of the electric field of an MWPC. Assuming that the electrodes extend infinitely in the z direction, the direction of the wires, the electric field $\vec{E}(x, y)$ has only components in the (x, y) plane. The field is determined by solving the two-dimensional potential equation

$$\Delta\phi(x, y) = 0 \tag{7.45}$$

in the space outside of the conductors with boundary values ϕ_i on each conductor i. In the following we choose the x-axis in the wire plane and perpendicular to the wires with the origin in the centre of a wire and the y-axis perpendicular to the wire plane (fig. 7.18).

Since the potential problem can only be solved numerically we want to derive an approximate solution which exhibits the essential properties of the exact solution. Initially we consider line charges in free space instead of the wires, that is, without the

cathode planes. Let the arrangement in the plane of the line charges be infinitely extended, thus resulting in infinitely many, infinitely long line charges. The (line) charge per length is $\lambda = dq/dz$ and the grid spacing is s. The potential of this arrangement can be calculated analytically (see e.g. eq. (10.2.30) in [718]):

$$\phi_0(x,y) = -\frac{\lambda}{2\pi\epsilon_0} \ln\left[2\sqrt{\sin^2\frac{\pi x}{s} + \sinh^2\frac{\pi y}{s}}\right]. \tag{7.46}$$

This potential has the following properties which allow us to approximately fulfil the boundary condition of an MWPC electrode arrangement:

(1) For $y \gg s$ the potential ϕ_0 becomes approximately independent of x:

$$\phi_0(x,y) \to \phi_0(y) \approx -\frac{\lambda}{2\pi\epsilon_0} \ln\left(2\sinh\frac{\pi y}{s}\right) \approx -\frac{\lambda}{2\pi\epsilon_0}\frac{\pi y}{s}. \tag{7.47}$$

This means that the planes with constant y are approximately equipotential planes. Placing the cathode planes at $y = \pm d$ ($d \gg s > 0$), we can set their potential to $\phi = 0$ by adding to the potential ϕ_0 a constant:

$$\phi_0(x,y) \to \phi(x,y) = \phi_0(x,y) + \frac{\lambda}{2\pi\epsilon_0}\frac{\pi d}{s}. \tag{7.48}$$

Hence the potential

$$\phi(x,y) = \frac{\lambda}{2\pi\epsilon_0}\left[\frac{\pi d}{s} - \ln\left(2\sqrt{\sin^2\frac{\pi x}{s} + \sinh^2\frac{\pi y}{s}}\right)\right] \tag{7.49}$$

approximately fulfils the boundary condition $\phi(x, \pm d) = 0$.

(2) For small values of y and $|x - m\,s|$ (with $m = 0, \pm 1, \ldots$ defining a specific wire) we find that ϕ becomes azimuthally symmetric around each wire m if in the expansion of sin and sinh only the first term in each case is picked up:

$$\phi(x,y) \approx \frac{\lambda}{2\pi\epsilon_0}\left[\frac{\pi d}{s} - \ln\left(\frac{2\pi}{s}\sqrt{(x - m\,s)^2 + y^2}\right)\right] = \phi(r), \tag{7.50}$$

with

$$r = \sqrt{(x - m\,s)^2 + y^2}.$$

Thus this approximation yields $\phi = \phi(r)$, which means that the anode wire surface $r = a$ constitutes approximately an equipotential surface where a boundary value can be given:

$$\phi(a) \approx \frac{\lambda}{2\pi\epsilon_0}\left[\frac{\pi d}{s} - \ln\left(\frac{2\pi a}{s}\right)\right]. \tag{7.51}$$

Hence with this approximation one can replace the line charge λ by an anode wire with radius a, such that λ obtains the meaning of a surface charge per length on the wire. Then the electric field is azimuthally symmetric close to the wire:

$$\vec{E}(r) = -\frac{\partial\phi}{\partial r}\vec{e}_r \approx \frac{\lambda}{2\pi\epsilon_0}\frac{1}{r}\vec{e}_r. \tag{7.52}$$

If the potential on the cathode surface is predefined, here by $\phi(x,d)=0$, then the boundary value on the anode will be fixed by the voltage V_0 applied between anode and cathode corresponding to the potential difference:

$$V_0 = \phi(x,y)\big|_{r=a} - \phi(x,d) = \frac{\lambda}{2\pi\epsilon_0}\left[\frac{\pi d}{s} - \ln\left(\frac{2\pi a}{s}\right)\right]. \tag{7.53}$$

For a given voltage, the charge induced per length on the anode, λ, depends on the capacitance of the arrangement:

$$\lambda = C\,V_0\,. \tag{7.54}$$

The capacitance C per length for a single wire is obtained by substituting λ into (7.53):

$$C = \frac{2\pi\epsilon_0}{\frac{\pi d}{s} - \ln\left(\frac{2\pi a}{s}\right)}\,. \tag{7.55}$$

For an anode area with a large number n of wires the capacitance $C_n = n\,C$ can be compared with the capacitance C_{Pl} of a respective planar capacitor consisting of a continuous anode plane between two cathode planes:

$$C_n = n\frac{2\pi\epsilon_0}{\frac{\pi d}{s} - \ln\left(\frac{2\pi a}{s}\right)} \;<\; C_{Pl} = n\,\frac{2\epsilon_0 s}{d} = n\,\frac{2\pi\epsilon_0}{\frac{\pi d}{s}}\,. \tag{7.56}$$

The inequality arises because in general $s > 2\pi a$, yielding the logarithm in the left hand side equation to be negative. With the values for a typical MWPC (7.44) one obtains for example for one cell ($n = 1$): $C = 3.5\,\mathrm{pF/m}$. A corresponding planar capacitor has the somewhat larger capacitance per cell of $C_{Pl} = 4.4\,\mathrm{pF/m}$.

The calculation of electrostatic fields in detectors with arbitrary electrode arrangements is in general done using computer programs for the solution of the potential equation with given boundary conditions. An often used program is Garfield [959,958]. Garfield can calculate two-dimensional configurations analytically employing methods of complex potential theory (as also discussed in appendix B); however, in the general case it can resort to numerical methods. Analytical solutions, if available, are preferable because they are less time-consuming, in particular when used in simulations of charge carrier transport.

7.8.3 Operation of MWPCs in experiments

7.8.3.1 Choice of the chamber gas

The choice of the gas for the operation of an MWPC follows the general criteria discussed in section 7.5. As a gas with good amplification and quenching properties the so-called *magic gas*, composed of 75% argon + 24% isobutane + 0.5% freon, has proven successful. The three components fulfil the tasks described in section 7.5: argon takes care of the high gas amplification at a relatively low field strength, isobutane acts as quencher, thereby damping the electron motion and absorbing photons, and the electronegative freon makes for the clean decay of the signal.

7.8.3.2 Detection efficiency

The probability that a particle traversing an MWPC is detected essentially depends on the following parameters:

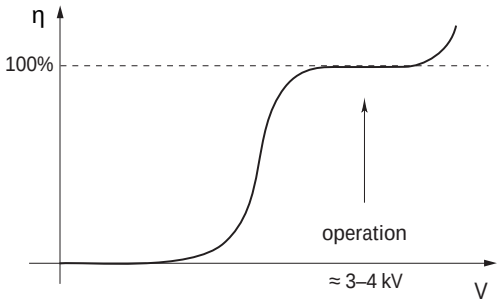


Fig. 7.21 Typical behaviour of a high voltage curve (also called efficiency curve) of an MWPC. Fraction of signals recorded by a chamber plane relative to the number of all passing particles as a function of the high voltage applied to the electrodes (the electronic signal threshold is kept fixed).

- Ionisation statistics: How much charge contributes to the signal depends on the number of electron–ion pairs produced and on possible charge losses on the way to the electrodes.
- Gas amplification and sensitivity of the electronics: Using modern low-noise electronics it is possible to become sensitive to femtocoulombs corresponding to a single electron at a gas amplification of 10^4.
- Dead time of the detector and the electronics: At high particle rates, high primary ionisation or high gas amplification, detector dead times or times with reduced efficiency can arise due to space charges screening the electrodes. The strength of the effect is determined by the delay of the charge equalisation by ion motion and the time constant of the voltage supply.

Multiwire proportional chambers reach efficiencies for minimum-ionising tracks of close to 100% per detector layer. At a typical ionisation density of about 100/cm and a path length in the gas of more than 8 mm the ionisation statistics virtually do not contribute to an efficiency loss. In order to set a sufficiently high gas amplification at minimal voltage necessary for full efficiency one usually measures *high voltage curves* where the fraction of detected particles from all passing particles is plotted versus the high voltage. The passing particles are usually determined by other detectors, for example scintillation counters as in fig. 7.25(a). Figure 7.21 shows a typical efficiency curve of an MWPC. During the measurement the signal threshold is kept fixed at a value which sufficiently reduces background noise. At a characteristic threshold voltage the rate increases relatively fast from zero to a plateau value close to 100% which is retained over a certain range. At further voltage increase the recorded rate can go beyond 100% due to spontaneous discharges and other effects. In order to ensure an efficient and at the same time stable operation the *working point* is set to a value not too far behind the beginning of the plateau.

At each particle passage the released charges are transported to the electrodes so that the voltage between anode and cathode would drop if the charges were not compensated by the high voltage supply (HV). This charge compensation proceeds with a time constant $\tau = R_{HV}\, C_{Det}$ where R_{HV} is the resistance of the voltage source and C_{Det} the detector capacitance (see fig. 7.1). As a protection against spark discharges this time constant is chosen much larger than the time constant for the formation of a discharge. Then a discharge current reduces the voltage for a time long enough to stop the discharge, at least in most cases. For example, for a resistance $R_{HV} = 10\,\mathrm{M\Omega}$ and a detector capacitance of 100 pF the recharge time constant would be $\tau = 1\,\mathrm{ms}$.

In order that the voltage drop due to a current over R_{HV} does not reduce the

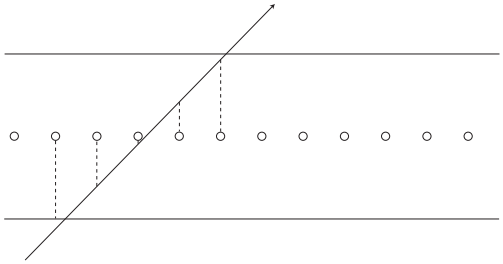


Fig. 7.22 Response of several neighbouring wires at non-perpendicular passage of a particle through an MWPC plane (cluster formation).

detection efficiency the rate has to be small compared to $1/\tau$. The achievable rates can be estimated from the voltage decrease and the subsequent decrease of the gas amplification. The working point set on the high voltage curve (fig. 7.21) determines which voltage decrease is still tolerable.

7.8.3.3 Resolutions

Resolution of the coordinate perpendicular to the anode wire. We consider an anode plane as in fig. 7.18 where the x-axis lies in the anode plane perpendicular to the wires. If upon the passage of a particle exactly one wire with the coordinate x_a fires, then the point of passage must lie within $x_a \pm s/2$, meaning within an interval $\Delta x = s$, the wire spacing. If all passage points near a wire occur with equal frequency (uniform distribution) then the standard deviation, corresponding to the resolution of the measurement device, is given by (see appendix E, eq. (E.4)):

$$\sigma = \frac{\Delta x}{\sqrt{12}}\,. \tag{7.57}$$

At a wire spacing of $s = 2\,\text{mm}$ the resolution then becomes $\sigma \approx 0.6\,\text{mm}$, a typical value for MWPCs.

Often several wires respond, in particular if a track does not pass perpendicular to the detector plane (fig. 7.22). In order to assign an x coordinate of the crossing point of the track with the anode plane, the centre of gravity of the x coordinates of the hit wires, called cluster centre of gravity, is calculated. With appropriate algorithms even better resolutions than with a single wire are reached (see also appendix E).

Two-dimensional position resolution. There are different methods to determine the second coordinate (z coordinate in fig. 7.18), for example:

- Combination of anode planes with different wire orientations, referred to as *stereo layers*;
- readout of the induction signal on the cathodes which are then usually segmented as strips or pads (the strips with another orientation than the anode wires);
- *charge division*: The anodes are read out from both wire ends, comparing the measured charges yields the position along the anode wire (the discussion of charge division in the context of drift chambers on page 227 is also here applicable).

Stereo layers. With the arrangement of stereo layers, ambiguities in the reconstruction of space points are avoided. While for the spatial resolution perpendicular wire orientations are most favourable, this leads to ambiguities if more than one particle is registered in the same event. In fig. 7.23(a) this is demonstrated for two tracks. The combinations of the two hit wires lead to four possible space points and in general, n

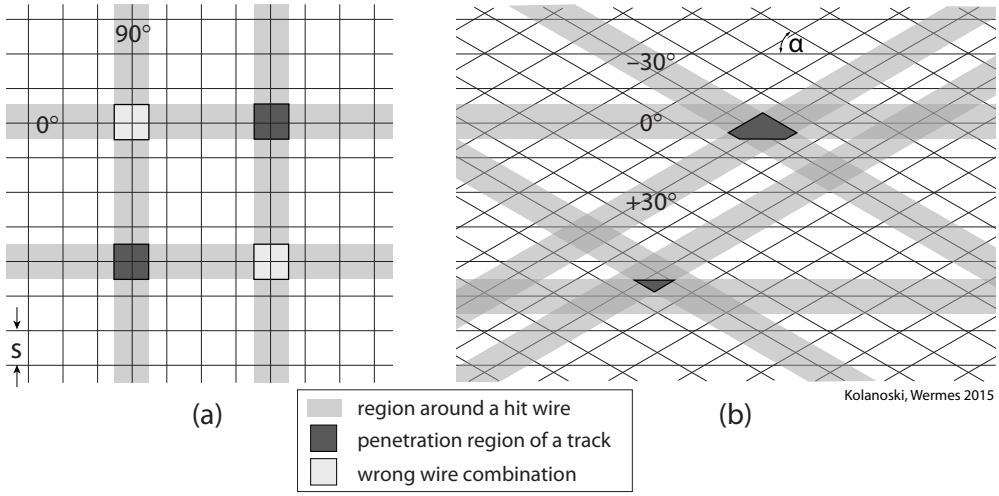


Fig. 7.23 The pictures show perpendicular top views of two and three anode planes, respectively, with different wire orientations stacked behind each other. The grey stripes are the $\pm s/2$ regions around hit wires. the dark areas mark the area containing the actual crossing point of a track, the light areas mark the false wire combination. (a) Ambiguities arising from the spatial reconstruction of the crossing points of particles if only orthogonal coordinates are measured. (b) Example of three stereo layers ($0°$, $\pm 30°$) arranged in a way to prevent ambiguities in track reconstruction. The crossing points of the tracks lie in the overlap region of $\pm s/2$ regions. This example does not contain any false assignment.

tracks lead to n^2 space points. Hence, for example, four tracks already lead to 12 false combinations.

In principle these ambiguities can be removed with the aid of a third, independent wire orientation. For example, in fig. 7.23(a) an additional wire plane, rotated by 45° would largely resolve the ambiguities. Figure 7.23(b) shows three anode layers with relative wire orientations $0°$, $\pm 30°$. Because of construction considerations this variant, with its symmetry relative to the $0°$ layer, is often preferred. Small stereo angles can also be advantageous for the construction, for example if the tensile stress of the wires can only be supported from two sides or if the electronic readout can only be placed on one side of the chamber.

In fig. 7.23(b) the passing points of tracks lie in regions in which the $\pm s/2$ stripes around the hit wires cross. The resolution is in general given by a polygon which is formed by the overlap of three stripes (fig. 7.23(b)). The coordinate along a hit wire is determined by a hit on a crossing wire with a resolution proportional to $1/\sin\alpha$, where α is the cutting angle between the two wires. For perpendicular wires $1/\sin\alpha = 1$ and thus the best resolution is obtained. But for small wire rotations $1/\sin\alpha$ can be relatively large as compared to the resolution perpendicular to the wires (for $\alpha = 30°$ it is a factor of 2). This is acceptable if the resolution requirements for the different coordinates are not the same. For example, the coordinate which determines the deflection of a track in a magnetic field, and thus the momentum resolution, should generally be measured more precisely than the coordinate orthogonal to that. If the resolution should be equal in all directions the optimum is reached by rotating three wire planes by $60°$ to each other.

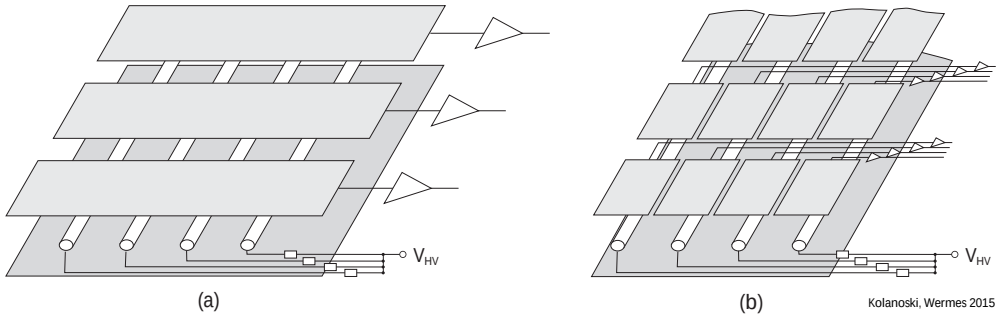


Fig. 7.24 Principle of a readout of (a) cathode strips and (b) cathode pads.

Cathode readout. A *cathode strip* readout (fig. 7.24(a))provides a further coordinate in addition to the one provided by the anode readout. In this case the signal induced by the charge avalanche on a strip is recorded.[9] For typical widths of cathode strips in the millimetre range a resolution of fractions of a millimetre can be reached. The resolution can be improved if the cluster centre of gravity can be determined by measuring the signal heights on each strip (see appendix E).

The direct measurement of space points is possible if the cathode is formed by small conducting areas called *cathode pads* (fig.7.24(b)). A pad structure is beneficial if the space point is to be used for a fast trigger or if the background is very high so that a space point reconstruction from projections would yield too many false combinations (see discussion above). For example, pad structures are employed in photosensitive MWPCs for the determination of Cherenkov rings (RICH detectors, see chapter 11) and in MWPCs for the readout of *time projection chambers* (section 7.10.10).

Time resolution. Because of the varying distances of the tracks from the anode wire and because of the statistical distribution of the electron drift times, the signal times lie typically in an interval of 100 ns after the particle has passed the layer. This time is particularly important for the employment of MWPCs in a trigger logic (see section 7.8.4) which possibly already includes a coarse track recognition. If necessary the time span estimated above could be shortened by forming a logical *OR* of signals from neigbouring layers and using as time stamp the time of the earliest hit. In this way resolutions around 10 ns have been reached, but a more typical value is about 60 ns.

7.8.4 Example of an MWPC detector arrangement with electronics

A simple arrangement of planar MWPCs with three chambers, each with three stereo layers, is shown in fig. 7.25(a). Together with the scintillation counters (Sci1,2 in the diagram) such an assembly could be employed for measurements of cosmic radiation. The scintillation counters provide signals for fast coincidences (coincidence widths $<$10 ns) for the suppression of background. For this set-up fig. 7.25(b) shows the principle of an MWPC readout and a scheme how the chambers can be incorporated into the trigger logic.

[9] This is similar to the readout of the streamer tubes in fig. 7.12. There, however, the readout strips are independent from the cathodes which have a high surface resistivity making it transparent for the induction signals.

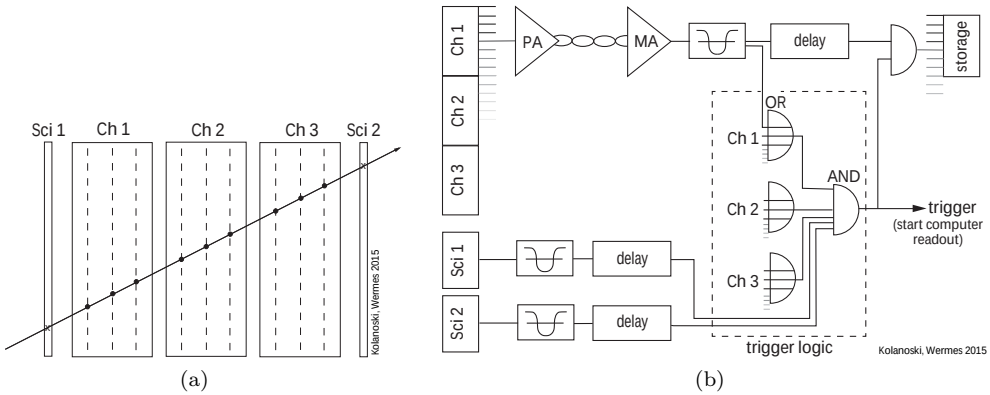


Fig. 7.25 (a) Simple arrangement of MWPCs as is often used in test experiments. Three chambers, each with three layers (with different stereo angles), are positioned between two scintillation counters. The scintillators provide a fast coincidence signal indicating the passage of a particle. (b) Schematic of simple readout and trigger electronics for the example of the MWPC set-up in (a) (description in the text).

Readout. The signals from the anodes are fed into a *preamplifiers* (PA) which are mounted directly at the detector, thereby avoiding long cables which would add additional capacitances and noise sources. As a rule, bipolar output signals of the preamplifiers are sent via *twisted pair cables* (see section 17.8.1) to a *main amplifier* (MA). The (analog) output signals of the main amplifier are converted by a discriminator into logical signals which are registered in a storage unit upon the occurrence of a trigger. At the same time the logical signals can be deployed for further processing in a trigger logic. With modern integrated electronics the functions of pre- and main amplifier can be combined in a chip directly at the chamber.

Trigger. An example for a simple trigger is depicted in the dashed rectangle in fig. 7.25(b). The logical *OR* of all signals of a chamber is routed to an *AND* coincidence together with the signals of both scintillation counters. Thus this schematic incorporates an *OR* of the three layers of a chamber leading to a time resolution of better than about 60 ns (see above). In order to exploit the better time resolution of the scintillation counters the circuit can be set up such that the time stamp of the trigger (e.g. the leading edge of the signal) is always formed by the same scintillation counter.

7.8.5 Applications of MWPCs

In particle physics the MWPCs have largely been replaced by drift chambers (see section 7.10) which offer the advantage of having a better position resolution at less readout channels. However, there are special applications, also outside particle physics, where MWPCs provide the optimal detector choice:

– Because of their fast response MWPCs are employed as *trigger chambers* in particle experiments (e.g. in the H1 experiment [34]).
– MWPCs are employed as spatially resolving photodetectors for the detection of photon radiation in particle physics, medicine or material research and control (see also chapter 10). Important examples in particle physics are the detection of Cherenkov

radiation ranging from the visible light to the UV region (chapter 11) and of transition radiation in the X-ray region (chapter 12). The photon conversion takes place either in the gas (e.g. heavy noble gases like xenon or krypton for X-ray detection) or at a photosensitive cathode (e.g. coated with CsI). The choice of the converter depends on the absorption in the relevant spectral region. For the detection of Cherenkov radiation organic gas mixtures are also used (chapter 11).

– In apparatus for atomic physics MWPCs can be run at low gas pressure for the detection of slow ions which would have insufficient range at normal pressure (see e.g. [797]).

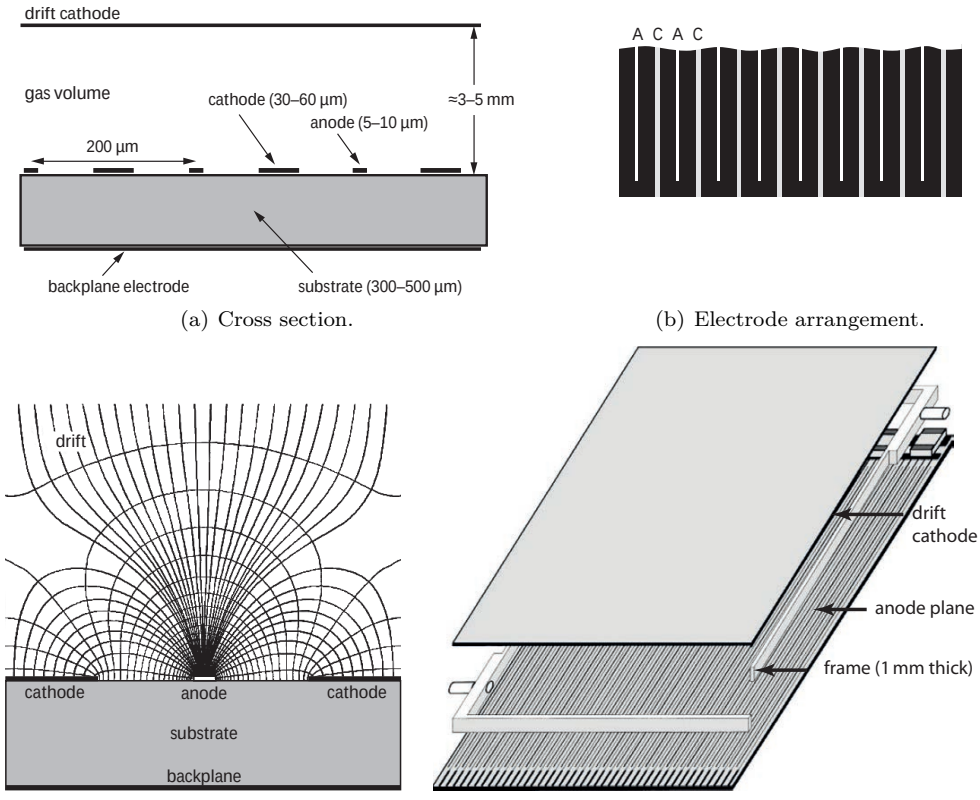
7.9 Micro pattern gas detectors

In experiments with high particle fluxes, as for example at the LHC, care has to be taken that the hit rate per readout channel does not become too high. A measure for the hit rate is the *occupancy* defined as the average probability to register a hit in a channel within the readout window. With increasing occupancy a single hit increasingly loses information; in the extreme case of 100% occupancy a registered hit does not contain any information. For a given particle flux the occupancy is less the smaller the sensitive area of a readout channel. In gaseous detectors with signal wires, however, the sensitive area cannot be made arbitrarily small (a) because the generation of sufficient ionisation charge requires a sufficiently long path of the particles through the gas and (b) because the wires cannot be arbitrarily short. Therefore the minimal cell diameters are in the range of millimetres, the wire lengths in the centimetre range and the ion collection times in the millisecond range. In contrast, semiconductor detectors (chapter 8) can reach electrode dimensions in the 10 µm range and charge collection times in the few-nanoseconds range.

Since the 1980s, gas-filled detectors with micro-structured readout planes, *micro pattern gas detectors* (MPGDs), have been developed which are able to exploit—even at high particle fluxes—the favourable properties of gaseous detectors, like relatively low cost and little disturbing material. By a smart arrangement of the electrodes the ion collection times can also be shortened. The micro-structuring of gas-filled detectors was introduced for the first time in 1988 [743], exploiting technologies known from silicon microstrip detectors. In these devices, instead of using wires, the gas amplification is achieved with microstrips on printed boards with widths in the range 10–100 µm. With such detectors particle rates of up to 2.3 MHz/cm^2 have been measured.

Subsequently, other structures generating high field strengths for gas amplification were also successfully developed, as will be discussed in the following sections for three examples. The concepts and techniques for MPGD construction as well as the signal readout with highly integrated electronics using custom-made chips (see section 17.6) have largely been adopted from the development of semiconductor microstrip detectors (section 8.5). An overview of the different MPGD variants can be found for example in [854,787]. While MPGDs are well suited as autonomous detectors for charged particles, they are also of interest for the measurement of the drift of electrons in time projection chambers (see section 7.10.10).

Gas mixtures for MPGDs. Conceptually MPGDs usually have small gas volumes so that the generated charges can be quickly collected. In order not to compromise the efficiency by fluctuations of the ion statistics the mean number n_p of primary ionisation clusters should be as high as possible. A frequently used gas is dimethyl



(a) Cross section.

(b) Electrode arrangement.

(c) Field lines.

(d) Drawing of an MSGC module.

Fig. 7.26 Microstrip gas chamber (MSGC): (a) cross section, (b) electrode arrangement on the anode plane, (c) electrical field lines (from [853], with kind permission of Cambridge University Press), (d) perspective view of a chamber module (from [225], with kind permission of Elsevier).

ether (DME) with a primary ionisation density more than twice as large as that of argon (see table 7.1). DME also has good quenching properties and shows no or minor ageing effects under radiation exposure (in contrast to isobutane which otherwise offers an even higher ionisation density). It is also not sensitive to background from X-ray and gamma radiation because of the relatively long radiation length. In order to reduce the anode voltage required to obtain the necessary gas amplification, the DME filling in MPGDs is usually mixed with noble gases, mostly argon or neon. As described in section 7.9.3, mixtures of DME with CF_4 (10–20%) [292] are also used which are 'faster', that is, the drift velocity is higher.

7.9.1 Microstrip gas chamber

The first realisation of micro-structured detectors with gas as detector medium was the *microstrip gas chamber* (MSGC) [743]. The principle is shown in fig. 7.26. The gas volume with a thickness of some millimetres is enclosed between a thin cathode plane ('drift cathode' in fig. 7.26(a)) and an insulating carrier substrate (e.g. glass) (fig. 7.26(d)). The narrow, gas amplifying anode strips are photolithographically applied to the substrate. In order to quickly suck off the ions generated in the avalanche

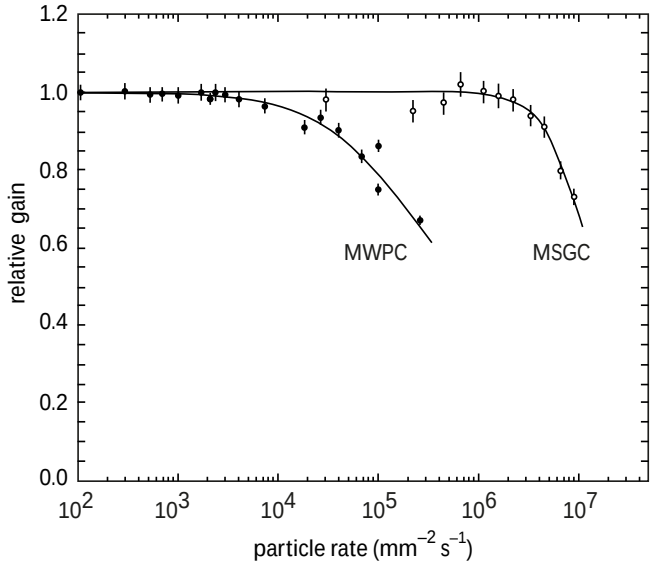


Fig. 7.27 Comparison of the rate capabilities of an MSGC and of a conventional multiwire chamber (MWPC). Shown are measurements of relative gas gain as a function of the particle rate [788].

near the anode, additional cathode strips (C) are placed between the anode strips (A) (fig. 7.26(b)).

The majority of the ions drift along the field lines (fig. 7.26(c)) to the cathode strips which are only 50–60 µm away from an anode strip. With this electrode geometry much higher particle rates can be coped with than with conventional multiwire chambers, about two orders of magnitude more (fig. 7.27). Space charges caused by the slowly moving ions, which would change the electric field for the subsequently arriving electrons, are strongly reduced by the cathode strips being close to the anodes.

In order to cover larger detector areas, MSGC systems are built in a modular way, usually assembled with overlaps to avoid dead regions. Modularity offers a simplified production of the microstructures. Moreover, a limitation of the strip lengths to about 10–20 cm helps to minimise the detector capacitance and thus the electronic noise (section 17.10.3). Low noise allows one to minimise the gas amplification, hence potentially prolonging the lifetime of an MSGC (see the discussion below on uncontrolled discharges in MSGCs).

MSGCs obtain position resolutions in the range of 30 µm, hence about 10–20 times better than with conventional MWPCs. Deploying additional readout strips on the back side of the substrate, orthogonal or tilted with respect to the anode strips, two-dimensional position information is obtained. Alternatively the readout electrodes can be structured as pads two-dimensionally.

In test set-ups gas amplifications of more than 10^5 have been reached with MSGCs. However, this performance usually cannot be obtained at high radiation levels, as was discovered during the development work for high-rate experiments, for example for CMS and HERA-B. Although some of the problems could be solved, they finally prevented the deployment of MSGCs in these experiments, at least using the original MSGC layout.

Part of the problem is related to the insulating substrate surface between the anode and the cathode strips. At high rates, positively charged ions are collected on the surface which cannot be neutralised fast enough because of the insufficient surface

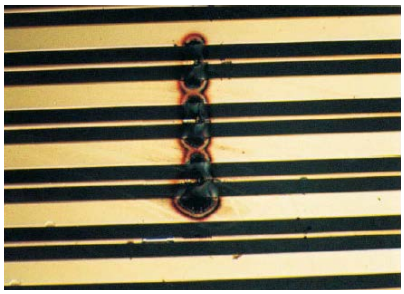


Fig. 7.28 Effect of a discharge on the metal electrode of an MSGC [446].

conductivity. This leads to local field variations and consequently to non-uniform gas amplification. The ion charges can only be neutralised if the surface has a small electron conductivity. A solution is the deployment of semiconducting glass, for example Pestov glass [779, 780] (see also section 7.7.3.1) with a resistivity of 10^9–10^{10} Ω cm. Alternatively a high-resistive coating, for example *diamond-like carbon* coating (DLC), as described in [140], with a surface resistance of 10^{14}–10^{16} Ω/□, can be applied (concerning the unit Ω/□ see footnote 6 on page 197).

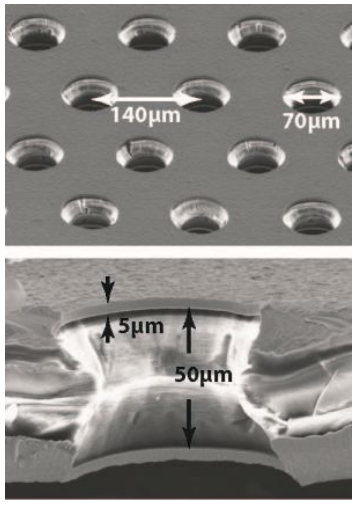
An even more serious problem for MSGC operation (in the configuration discussed here) is the uncontrollable discharges, which can lead to damage of the electrodes. Even if the charge generated by a highly energetic particle might be unproblematic for stable operation, it is unavoidable that from time to time very highly ionising atomic nuclei or nuclear fragments are also produced. For example, a gold nucleus, knocked out of an electrode, can have a 20 000 times higher ionisation density than a minimum-ionising particle. At such high field strengths this leads to energetic discharges which can distroy the electrodes, as shown in fig. 7.28. Attempts to limit the energy going into these discharges, for example by means of series resistors or of a special coating of the electrode edges, was only moderately successful.

The final conclusion is that for MSGCs in the original layout a stable operation at the necessary gas amplification with lifetimes of many years cannot be assured, at least not in high-rate experiments. Therefore, from the beginning of the 1990s solutions for micro-structured gaseous detectors with higher operational reliability were looked for. An option is the distribution of the amplification over several stages, as will be described in the next section. Another one is avoiding sharp edges at the electrodes causing high fields; for MSGCs such edges actually have curvature radii in the range of micrometres. An example for such a solution will be presented in section 7.9.3.

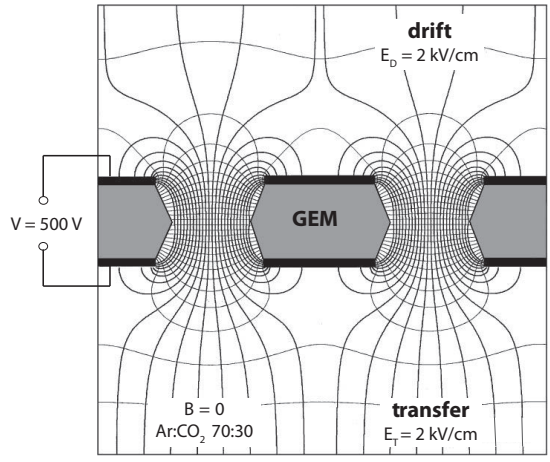
7.9.2 Gas electron multiplier (GEM)

A very successful variant of an MPGD is the *gas electron multiplier* (GEM) [851]. Originally conceived as a preamplifier stage for MSGCs to solve the problems with electrical spark discharges, meanwhile GEMs combined with appropriate readout units are used both as autonomous particle detectors and as elements of readout planes for time projection chambers (TPCs, see section 7.10.10).

Mechanically, the GEM construction is comparatively simple and very robust. Typically a GEM consists of a 50 µm thick Kapton foil with copper coating on both sides in which holes with a diameter of about 50–70 µm are etched (fig. 7.29(a)). Between the copper coatings of the foil a voltage of about 400 V is applied creating a strong field in the holes (fig. 7.29(b)). A GEM foil is purely a gas amplification device without its own readout layer. Therefore a separate electrode layer for signal readout can be



(a) GEM foil (top), hole cut-out (bottom).



(b) Electric field.

Fig. 7.29 *Gas electron multiplier* (GEM): (a) GEM foil with 70 µm large perforations [848]; (b) field distribution of a GEM foil (from [139], with kind permission of Elsevier).

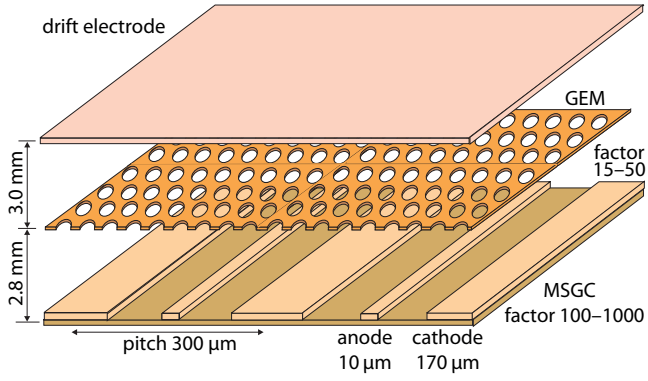


Fig. 7.30 The GEM MSGC: Due to the additional gain of a factor 15–50 provided by the GEM foil the amplification at the anode strips of the MSGC can be lowered to about $< 10^3$ in order to reduce the danger of discharges (from [1019], with kind permission of Elsevier). The respective gain factors are indicated in the drawing.

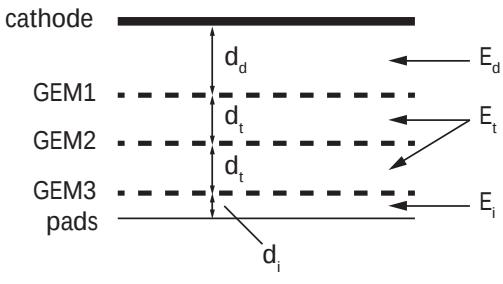


Fig. 7.31 A triple-GEM assembly with three GEM layers and a pad readout (from [176], with kind permission of Elsevier). The indices d, t, i attached to the fields E and the electrode distances d denote: 'drift', 'transfer' and 'induction'. In [176] the following values are given for a tested set-up: $E_d = 3\,\mathrm{kV/cm}$, $E_t = 3\,\mathrm{kV/cm}$, $E_i = 5\,\mathrm{kV/cm}$ and $d_d = 3.3\,\mathrm{mm}$, $d_t = 2.2\,\mathrm{mm}$, $d_i = 1.5\,\mathrm{mm}$.

optimised for a chosen application. Figure 7.30 shows the combination of an MSGC with a GEM foil splitting the gas amplification into two stages. With such an arrangement the problem of discharges leading to the destruction of electrode strips in MSGCs could be solved.

Several GEM amplification stages can be deployed in series. The induction signal of the last stage can then be read out via a non-amplifying electrode, for example with a pad structure. In such a scheme the region of high fields is spatially separated from the readout electrode, protecting the electronics from spark discharges more efficiently. Since GEM foils are flexible they can be built, for example, in cylindrical form suited for storage ring experiments [79].

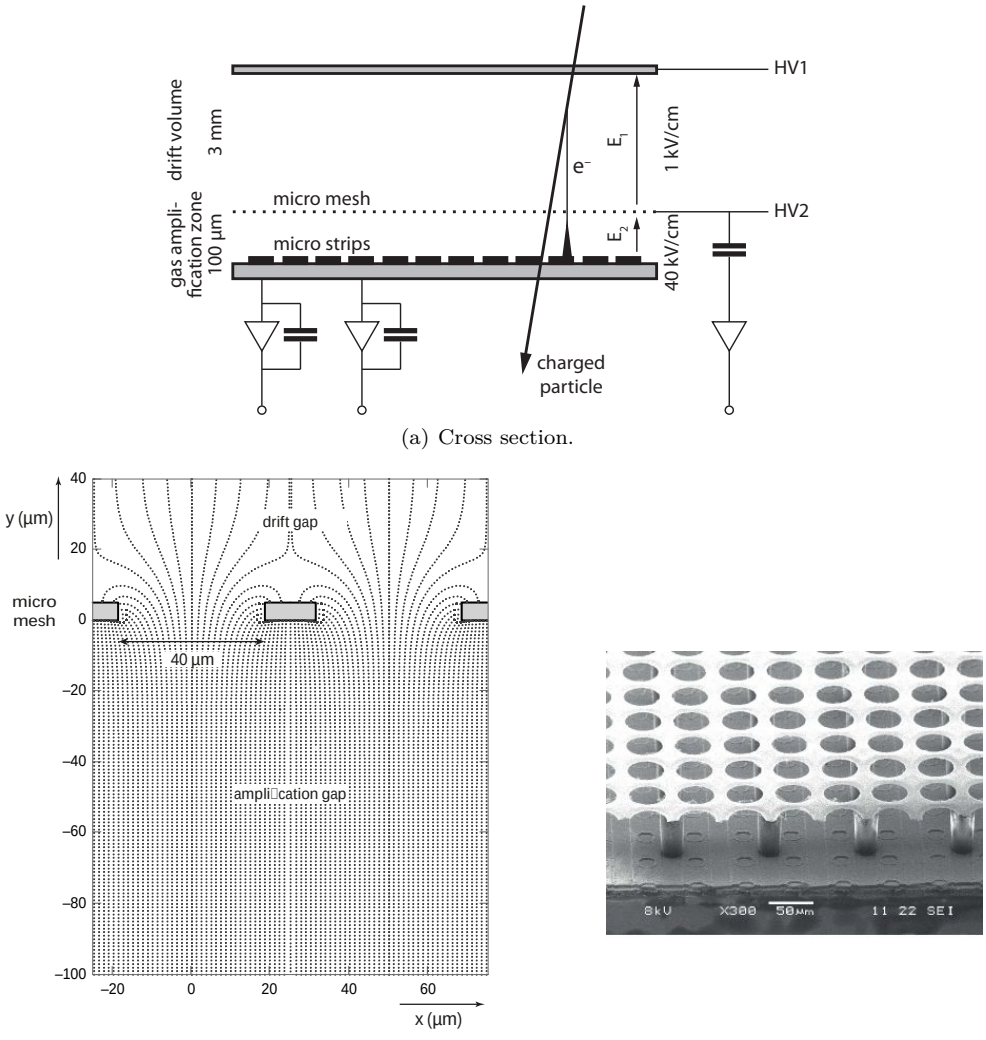
Triple-GEM devices (fig. 7.31) have become a certain standard configuration. The length of the drift distance before the first GEM foil is a compromise between the demand for as much ionisation charge as possible and a fast charge collection. The distance between the last foil and the readout electrode should be as small as possible to obtain a high induction signal. A voltage of 350–500 V over the GEM holes leads to field strengths in the order of 100 kV/cm in the holes and a total amplification of a triple-GEM device in the range of 10^4. With these parameters it can be assured that the gain factor remains below 100 for each stage, which is important for reliable operation.

7.9.3 Micromegas

Another MPGD variant is the *MICRO-MEsh GAseous Structure* (micromegas) [460] which combines the principle of parallel-plate detectors (section 7.7) with the separation of the amplification from the ionisation–drift region, as just described for GEM devices. A typical layout is shown in fig. 7.32(a). The about 3 mm thick ionisation and drift volume is separated from the about 100 µm thick gas amplification region by a micro-structured mesh.

The electrons generated by charged particles in the ionisation–drift region drift through the mesh into the amplification zone (fig. 7.32(a)). The field for the gas amplification (fig. 7.32(b)) is typically 35–45 kV/cm, corresponding to a voltage on the mesh relative to the electrode plane (anode) of 350–450 V. The corresponding drift velocities are 10–20 cm/µs. The secondary electrons are collected on the electrodes, segmented as strips or pads, within not more than about 1 ns and the major part of the ions are collected on the mesh within about 100 ns. This leads to a fast induction signal from the electron and ion movement. The gas filling has to be optimised for low spark rates, suitable gas amplification and low diffusion. Common gas mixtures are for example argon–DME with 10–20% CF_4 admixture, see [292], or neon/C_2H_6/CF_4 in the ratio 80:10:10, see [734]. Neon as the noble gas base was found favourable over argon in terms of space and time resolutions, as well as in reducing the discharge probability upon ionisation, a common phenomenon with micropattern detectors. In addition neon is a 'fast' gas. Applying high fields (≈40 kV/cm) time resolutions for muon tracks in the order of 25 ps are possible [236]. Micromegas detectors with these specifications are deployed as tracking detectors, for example, in the COMPASS experiment at the CERN-SPS [734], featuring a rate capability of some 100 kHz/mm^2 and a position resolution of 100 µm.

The mesh is supported by insulating micro-columns, thus keeping a constant distance between the mesh and the readout plane. An adjustment of the mesh as precise as possible is essential for the optimisation of the detector properties. By us-



(a) Cross section.

(b) Electric field.

(c) Micro-mesh structure.

Fig. 7.32 Micromegas detector. (a) Cross section through the chamber structure with drift and gas amplification volumes [292], (b) electric field distribution (from [294], with kind permission of Elsevier), (c) detail of an *InGRID* structure which is directly placed on a pixel readout chip by means of micro-mechanical postprocessing on CMOS wafers (from [470], with kind permission of Elsevier).

ing modern micro-mechanical processing on fabricated electronic wafers very homogeneous grid structures can be produced which are directly mounted on a pixel readout chip [470, 635]. The so-called *integrated grid* (InGRID) structure is depicted in fig. 7.32(c) [470]. This technique allows for a precise alignment of the grid holes with the readout structure and ensures a constant distance of the grid from the readout plane. Because of the precise alignment the induced signal of a single amplified electron is essentially concentrated onto one pixel, thereby achieving a sensitivity of the device to single electrons [471, 675].

A constant distance between grid and readout plane should make the gas amplification independent of the position. However, the requirements are alleviated by a compensating effect. To first approximation small distance variations do not change the amplification [292] for two reasons: (a) the traversed potential difference remains the same and (b) to reach the ionisation threshold the same portion of the potential difference is needed, thus keeping the number of multiplications (and hence the gas amplification) about the same (see the eqs. (7.22) and (7.27)).

7.10 Drift chambers

Drift chambers [968] are a further development of MWPCs. In a drift chamber the position where a particle traversed the detector is determined by the drift time of the electrons measured from the generation of ionisation at the time t_{start} to the arrival at the anode at the time t_{stop}. The basic formula is an integral over the drift velocity v_D (which in general may vary with the local field corresponding for moving charges to a time dependence of v_D):

$$x = \int_{t_{start}}^{t_{stop}} v_D(t)\, dt\,. \tag{7.58}$$

Hence for a drift chamber the position resolution is no longer directly dependent on the spacing of the signal electrodes but rather on the precision of the time measurement of the electron drift until the arrival at the anode. Since drift distances can be much larger than the wire distances in MWPCs, in general the number of readout channels and the material per area that a particle has to traverse are reduced. At the same time the position resolution improves as compared to MWPCs. While typical MWPC resolutions are about 600 µm (see page 210) the best resolutions reached with drift chambers lie at about 20 µm. However, for large volume detectors (some cubic metres as in collider experiments) the typical values lie more in the range 100–200 µm.

7.10.1 Working principle of a drift chamber

In order to explain the working principle of a drift chamber we consider the detector in fig. 7.33 where the drift and the amplification regions are separated. Such a layout is employed, for example, as monitor for the gas parameters of drift chamber systems by measuring drift velocities and signal heights (see e.g. the use by the TOPAZ [430] and CMS experiments [902]). The particularly long drift path gives rise to a high sensitivity to changes of the drift velocity and to changes of the pulse height by charge losses, for example due to contaminations. The separation of drift and amplification regions is also the principle of the time projection chamber employing a particularly large drift volume (see section 7.10.10). Apart from that, however, such a layout is rather untypical for larger drift chambers.

In fig. 7.33 the electrodes arranged on the sides of the drift volume are connected in series via a voltage divider chain yielding a uniformly decreasing potential and thus a very homogeneous drift field. The homogeneous field lines end at a grid which is kept at a fixed potential and which separates the drift region from the amplification region. The corresponding potentials are also depicted in the figure. Depending on the drift gas the field strengths in the drift region is chosen according to the conditions to be monitored, typically around 500 V/cm. With typical drift velocities of 50 µm/ns the

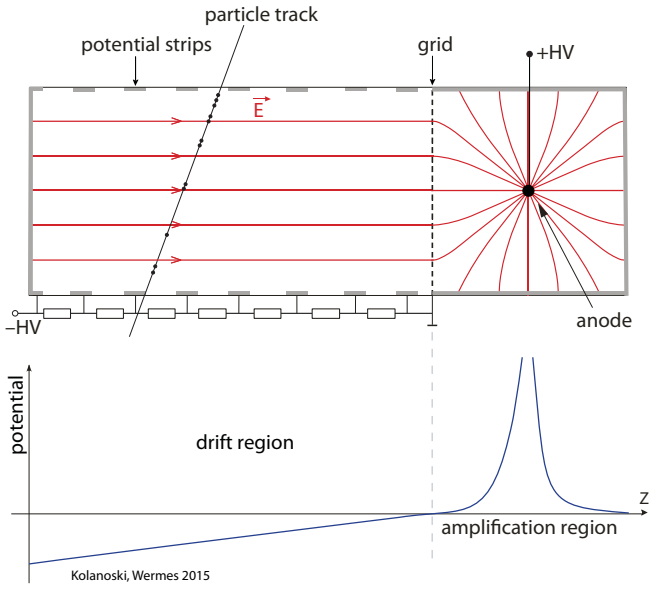


Fig. 7.33 Detector arrangement as used for monitoring drift velocities. It does not really correspond to a typical drift chamber but illustrates quite clearly the principle. The drift volume with a drift field as constant as possible is separated from the region where the amplification takes place.

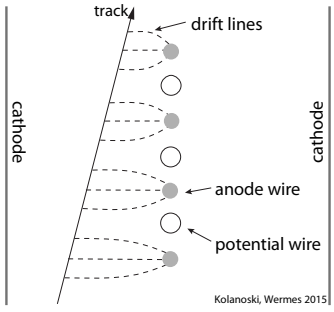


Fig. 7.34 Drift cell of a *jet chamber*. The anodes are positioned in the midplane between cathode planes. In order to improve the separation of the drift paths additional *potential wires* are strung between the anodes. The cathodes can be made from conducting planes or from wires. The principle of the jet chamber was developed for the JADE experiment [367] and then employed by various experiments, for example, for the central drift chamber of the OPAL experiment [202].

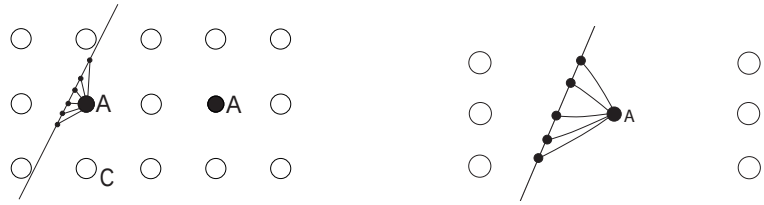
time measurements should have a resolution in the order of 1 ns if position resolutions of about 100 µm are to be reached.

7.10.2 Geometries of drift cells

The measurement and reconstruction of tracks from charged particles requires the coverage of large volumes, which can be achieved by different methods. In most cases the chamber volume is filled with many small drift cells, each with a signal wire (anode). To keep the design simple, in a usual drift cell geometry the drift and amplification regions merge into each other and the drift field is no longer homogeneous (in contrast to the design in fig. 7.33). In general the relation between the distance of the ionisation from the anode and the drift time is then no longer linear, a situation which can, however, be dealt with by appropriate calibration (section 7.10.7). The central anode wire is surrounded by electrodes lying on cathode potential. The cathodes can be formed by a conductive surface on a carrier material, as for example for the cells of *jet chambers* (fig. 7.34) or for *drift tubes* with cylindrical, rectangular or hexagonal cross sections (see fig. 7.35).

However, often the drift cell boundaries are approximated by wires arranged in

Fig. 7.35 Cross sections of round (*straw*) and hexagonal (*honeycomb*) drift tubes.



(a) (b)

Fig. 7.36 Examples for drift cells which are built from anode (A) and cathode (C) wires. (a) Two neighbouring drift cells of the central drift chamber of the ARGUS experiment [339]. (b) Drift cell of the central drift chamber of the TASSO experiment [222].

various configurations (see examples in fig. 7.36). Typical properties of anode and cathode wires correspond to those of MWPCs as listed in table 7.5 on page 205. The advantage of drift cells formed from wires is the potentially lower material density of the detector. However, the wire tension often requires quite massive constructions as, for example, in the case of the end plates of cylindrical drift chambers (fig. 7.38).

The advantage of self-supporting drift cells, as for example drift tubes made from aluminium or conductive polycarbonate foil, is the saving of carrier constructions and the electrical decoupling of the cells (e.g. a broken wire remains inside the cell containment and does not cause a short circuit in other cells). Metallic drift tubes, for example made of aluminium, are used in applications where multiple scattering is not a major concern, for example in muon detectors between absorber layers (see section 14.3).

7.10.3 Drift chamber types

The drift cells are arranged in a regular pattern covering the entire detector volume. The pattern depends on the respective application. Normally the drift cells are arranged side by side in layers and several layers are stacked one after another thereby allowing for the reconstruction of particle directions. Most often the layers have planar or cylindrical shapes.

A planar layout is used if the tracks of interest have a preferred direction (fig. 7.37) as in fixed-target experiments or in the forward and backward directions in collider experiments. In a planar drift chamber system the direction of the anode wires usually changes from layer to layer in order to achieve a spatial resolution (fig. 7.37).

Cylindrical drift chambers are often employed as central tracking chambers of collider experiments (fig. 7.38) with the chamber axis pointing in the beam direction. The cells are arranged on cylinder surfaces, as shown in fig. 7.39(a). Usually these experiments have a solenoidal magnetic field (section 9.2.2) with the field axis pointing in the beam direction.

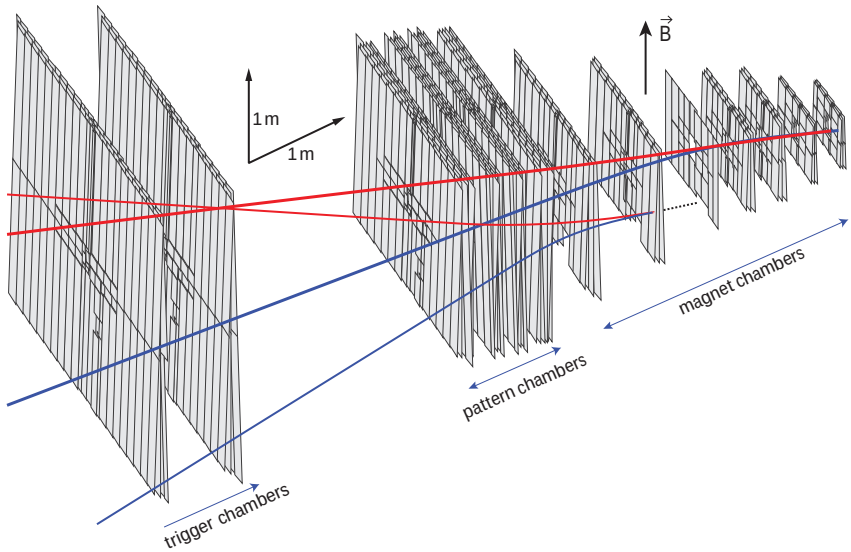


Fig. 7.37 Planar drift chamber of the *Outer Tracker* of the HERA-B detector [76]. The chambers are arranged in *super-layers*, each comprising several layers of drift tubes with different orientations (0°, ±5°). The different orientations make a three-dimensional reconstruction possible. The first seven chambers are located inside a magnetic dipole field where the principal field component has the direction of the 0° wires. In the projection perpendicular to the field, which is the projection of momentum deflection, the best position resolution is obtained and thus an optimal momentum resolution.

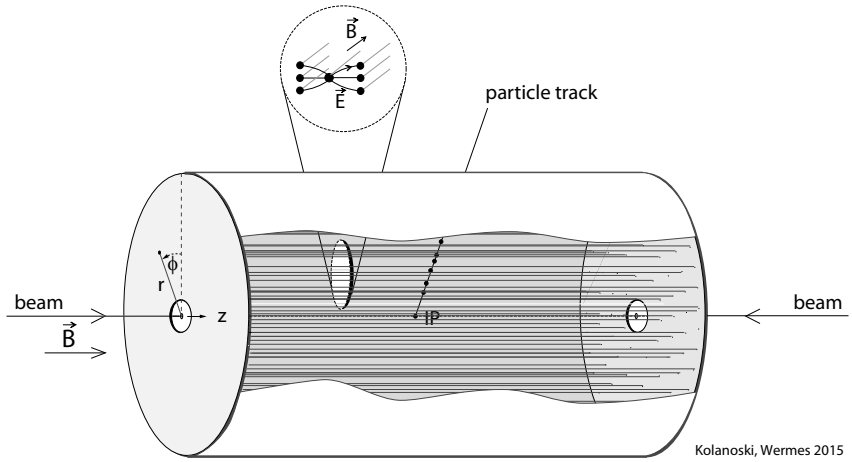
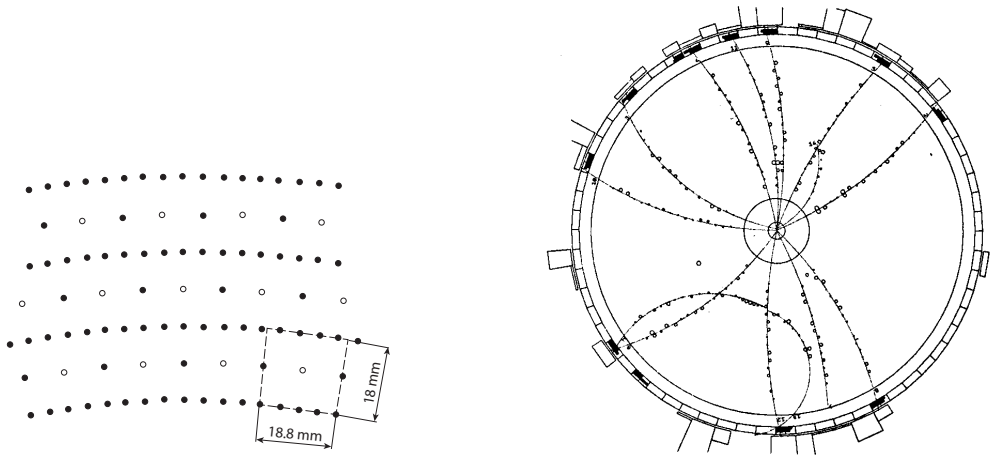


Fig. 7.38 Principle of a cylindrical drift chamber with cells formed by wires. The cathode and anode wires are strung parallel to the beam and to the magnetic field. The magnified cut-out shows the cross section of a drift cell, which corresponds here to the cell in fig. 7.36(b). Normally cylindrical coordinates (r, ϕ, z) are used (r = perpendicular distance from the beam, ϕ = azimuth angle in a plane perpendicular to the beam, z = coordinate in the direction of one of the beams with the origin at the nominal beam crossing point).



(a) (b)

Fig. 7.39 (a) Drift cells arranged on cylinder shells in the central drift chamber of the ARGUS experiment as an example for a cell structure formed with wires (projection onto a plane perpendicular to the beam). The cell structure corresponds to the one depicted in fig. 7.36(a). (b) Tracks in the central drift chamber of the ARGUS experiment ($r\phi$ projection: wires and magnetic field are perpendicular to the image plane). The measured isochrones correspond to the little circles to which the tracks run approximately tangentially (see fig. 7.40). The drift chamber is surrounded by scintillation counters and calorimeter modules for which the hits or energy deposits, respectively, are also indicated. From [339], with kind permission of Elsevier.

In a frequent variant of cylindrical drift chambers the wires forming the drift cells are strung parallel to the magnetic field. The electric field is then perpendicular to the magnetic field ($\vec{E} \perp \vec{B}$). With this field configuration a particularly good position resolution is achieved in the plane in which the particle trajectories are deflected ('$r\phi$ plane', as in fig. 7.39) yielding a good momentum resolution.

Another cylindrical chamber variant is the time projection chamber where the drift field is parallel to the magnetic field. This chamber type will be discussed in section 7.10.10.

7.10.4 Determination of the track coordinates

The coordinates of a track passing a drift chamber are given by the combination of the electrode arrangement and the measured drift time. In the following we explain different the procedures for obtaining a set of coordinates necessary for the reconstruction of a track.

Measurement in the plane perpendicular to the anode wires. A curve around an anode wire connecting all points from which electrons have the same drift time is called *isochrone*. When a particle crosses a drift cell the drift time measurement determines the isochrone to which the track is tangential. The tangent points of a track are determined by fitting a track to the measured isochrones in different anode layers (fig. 7.40). The tangent points can lie on both sides of the hit wire. In general, such a left-right ambiguity can be resolved by fitting to hits in several anode layers (fig. 7.40). However, the example in fig. 7.41 shows that wire arrangements exist where

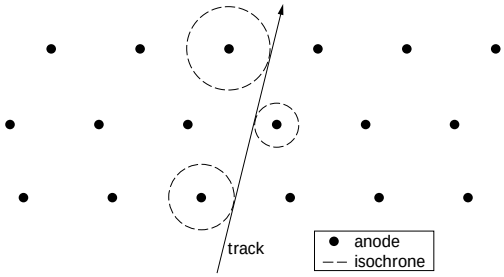


Fig. 7.40 Reconstruction of a track by fitting to the isochrones, here idealised as circles, which are determined by the measured drift times. Only the anode wire of each drift cell is shown. The anode wires are staggered in sequential planes in order to be able to resolve the left-right ambiguities (see text and fig. 7.41).

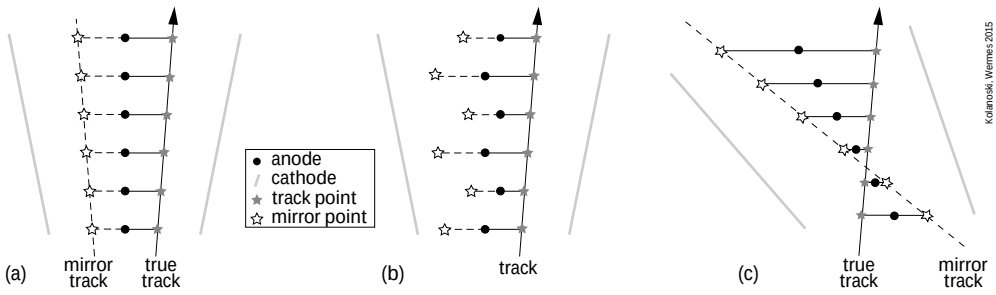


Fig. 7.41 Left-right ambiguities in a *jet cell* of a cylindrical drift chamber. The plotted (true) tracks come from the interaction point which lies on the cylinder axis (equal to the beam axis). (a) Track reconstruction for radially uniform arrangement of the anode wires ('radial' relative to the beam axis). In this case the correct reconstruction cannot be distinguished from the incorrect one of a mirror image. (b) When *staggering* the anode wires the mirror points do not lead to a solution for the track reconstruction. This anode layout was chosen for the jet chamber of the OPAL experiment [202]. (c) A linear anode arrangement can be tilted against the radial direction so that the mirror tracks do not point back to the interaction point and therefore can be rejected. Such a design was chosen, for example, for the central jet chamber of the H1 experiment [261] and for the central jet chamber of the CDF experiment [58].

this is principally not possible. This holds in particular for an alignment of the anode wires along the preferential direction of the tracks as demonstrated in fig. 7.41(a). Figures 7.41(b) and (c) show how ambiguities can be avoided: (b) by *staggering* of the wires or (c) by an anode arrangement which does not point back to the interaction point (in the projection onto a plane perpendicular to the anodes). The arrangement of equal-sized drift cells on concentric cylinder surfaces, as in fig. 7.39, leads inevitably to an azimuthal staggering of the cells thus avoiding ambiguities.

Three-dimensional track reconstruction. With parallel anode wires only the projection of the track onto a plane perpendicular to the wires can be determined. Often a preference direction of the wires is already given by the chamber layout. For example, in cylindrical drift chambers the wires have to be strung between end plates and thus essentially parallel to the cylinder axis (usually the beam direction). The coordinate along the chamber axis, which is necessary for a spatial reconstruction as well, can only be determined with much more effort.

If a magnetic field is employed for momentum analysis (see section 9.3) the wires usually run parallel to the magnetic field, which in cylindrical drift chambers at colliders is also the beam direction. Since in this case the momentum deflection occurs

perpendicular to the anode wires, the momentum resolution is good (more precisely, the resolution of the momentum component perpendicular to the magnetic field). In this configuration the resolution along the anode wires is less critical for the track reconstruction. In most cases an accuracy in the range from 1 mm to about 1 cm in the z direction is sufficient for the required direction and momentum resolution.

The third coordinate can be determined with different methods (see also section 7.8.3):

Charge division on the wire: The position along the wire is determined from the ratio of the signals at both ends of the anode wire [801]. For this purpose the wire has to have a higher resistance (typically about 10 times higher) than otherwise normal for anode wires ($<100\,\Omega$/m for a 25 µm thick tungsten wire). Typical resolutions are about 1% of the wire length (see e.g. [148, 202]).

Time-difference measurements: The position is determined from the difference of the arrival times of a signal at both ends of the anode wire. Using this technique in the central drift chamber of the ZEUS experiment, for example, position resolutions of about 1.5% of the wire length, corresponding to about 3 cm, have been reached [420].

Stereo wire layers: In special layers, *stereo layers*, the wires are strung under an angle α (usually only a few degrees) with respect to the preferential direction (0°). In planar drift chamber systems detectors with different wire direction can be combined in a modular way. For example, for the planar chambers in fig. 7.37 the detector production was largely simplified by building only one module type per super-layer and then rotating the module in different layers by the chosen stereo angle.

An example for stereo wires in cylindrical drift chambers is depicted in fig. 7.42. With 0° layers alone only the coordinates in the $r\phi$ projection can be determined; the z coordinate remains undetermined (the coordinate system is defined in fig. 7.38). The drift time to a 0° wire fixes the ϕ coordinate while the drift time to a stereo wire delivers the ϕ coordinate as a function of the z coordinate of the track, as one can see in fig. 7.42. The ϕ value in a stereo layer with radius r is with respect to the ϕ value measured in a 0° layer displaced by $\Delta\phi = z\,\alpha/r$. This yields the z coordinate and the corresponding polar angle θ relative to the z-axis:

$$z = \frac{r\,\Delta\phi}{\alpha}\,, \qquad \theta = \arctan\frac{r}{z} = \arctan\frac{\alpha}{\Delta\phi}\,. \tag{7.59}$$

The resolution in z direction is

$$\sigma_{rz} = \frac{\sigma_{r\phi}}{\alpha} \tag{7.60}$$

and thus by a factor $1/\alpha$ worse than the resolution $\sigma_{r\phi}$ in the $r\phi$ plane. For example, for a stereo angle of $\alpha = 3°$ and an $r\phi$ resolution of 150 µm, the rz resolution is 3 mm.

Cathode signal readout: If the cathodes are not formed by wires, but for example by conductive boards with z segmentation, the induced cathode signals can be read out in order to obtain the z coordinate (see e.g. [280]).

Independent z chambers: Cylindrical chambers can be combined with externally installed planar chambers that allow for arbitrary wire directions and thus also for a direct measurement of the z coordinate. An example is the OPAL detector where outside of the central drift chamber special z chambers were deployed [63].

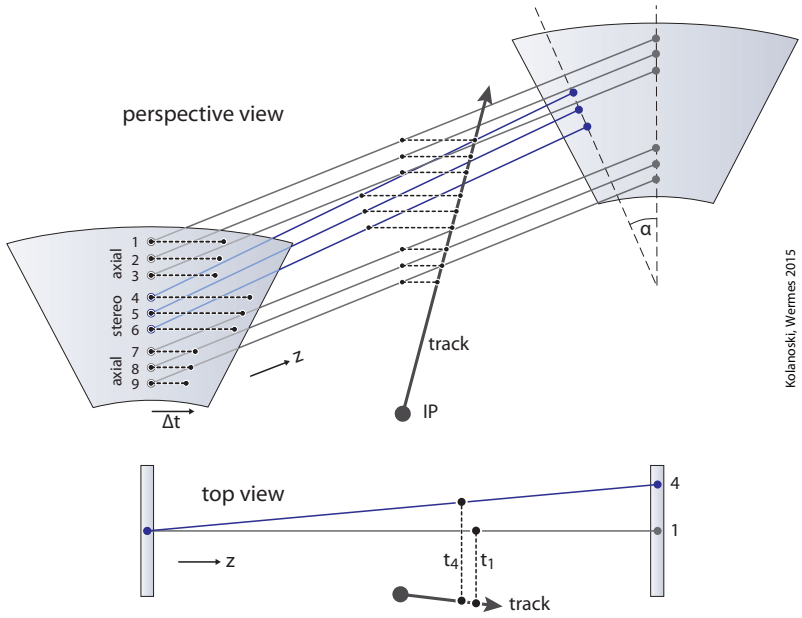


Fig. 7.42 Cutout of a cylindrical drift chamber (as in fig. 7.38) with stereo layers for the determination of the coordinate parallel to the beam direction (z) (adapted from [798]). The beam direction is also the direction of the magnetic field and of the preference direction of the wires. The upper part of the figure shows the wires which are strung in axial direction (parallel to the beam) and those which are strung under a *stereo angle* α (only the anode wires are shown). The lower part of the figure shows a top view of the assembly with the outermost wires for both orientations. Looking at the delineated particle track which starts at the interaction point (IP) one realises that the distance of the track from a stereo wire varies with the z coordinate (for a given track projection in the $r - \phi$ plane). This feature allows for a three-dimensional fit of a track to the drift times.

7.10.5 Choice of the operation parameters of a drift chamber

The *working point* of a drift chamber is defined by the gas mixture, the gas pressure and the applied high voltage. Usually this determines both the drift velocity and the gas amplification, except if the drift and amplification regions are separated as in figs. 7.33 and 7.52.

7.10.5.1 Gas mixture

In section 7.5 the general criteria for the choice of a detector gas are discussed. For drift chambers the position resolution is particularly important, requiring a gas with high ionisation density and little diffusion. A high ionisation density is also important if at the same time particle identification by dE/dx measurements should be performed (see section 14.2.2). A particularly good dE/dx and position resolution can be reached by increasing the chamber pressure (typically 2–5 bar) because the ionisation density is proportional to the pressure. The gain in dE/dx resolution has to be balanced against the increased multiple scattering and absorption because of the additional material for the pressure vessel.

The drift velocities lie typically at about 50 µm/ns. Thus a time measurement with digitisation steps of about 1–2 ns is appropriate if resolutions of about 100 µm should be reached. For better position resolutions lower drift velocities are advantageous, which can be achieved, for example, with admixtures of DME (table 7.1). A gas with high drift velocity is for example CF_4 (table 7.1), which is relevant for high-rate experiments. More generally the tuning of drift velocities of electrons is discussed in section 4.6.4.3.

In the following we want to explain the requirements for the chambers and the resulting choice of the gas using the parameters of drift chambers of different experiments listed in table 7.6:

Argon with ethane, methane or isobutane. These argon mixtures with hydrocarbons are standard for experiments with not too high rates. The mixing ratios account for the requirement that at a gain of some 10^4 the corresponding drift fields lead to a saturation of the drift velocity; see the examples in fig. 4.10 on page 116. Saturation implies that the drift velocity and hence the position determination becomes less sensitive to fluctuations of the E fields as well as of temperature and pressure. Due to the damping of the (random) electron velocities these so-called 'cool' gases exhibit moderate drift velocities (about 50 µm/ns), with correspondingly low demands on the precision of the time measurement, and little diffusion (see e.g. fig. 4.12 on page 117).

At high hit rates, however, detectors employing these gases tend to accumulate deposits on the electrodes, deteriorating the detector (see the related sections 7.5 and 7.11.2).

Argon–CO_2. This gas is, for example, used for the muon drift tubes of the ATLAS experiment, in this case with a small addition of water. The replacement of a hydrocarbon by CO_2 renders this gas mixture very stable against polymerisation, and with a 30% portion of CO_2 the quenching effect is sufficient. The admixture of water yields an additional protection against insulating deposits on the electrodes.

Propane–methylal. This mixture was employed for the central drift chamber of the ARGUS experiment [73]. Propane was chosen because of a larger radiation length and thus less multiple scattering as compared to argon (table 7.1). The admixture of methylal and a small addition of water suppresses polymerisation deposits on the electrodes (see section 7.11).

Ar–CF_4–CO_2. This gas mixture was employed in the Outer Tracker of the HERA-B experiment. The CF_4 component leads to drift velocities about twice as fast as for the standard mixtures of argon with a hydrocarbon. Large drift velocities are advantageous for experiments with very high rates (like HERA-B or the LHC experiments). The CF_4 portion also acts as a quencher in addition to CO_2.

However, in various tests CF_4 turned out to be not safely controllable under radiation exposure; see section 7.11. Therefore the LHC experiments have abandoned the initially foreseen CF_4 component from the gas mixtures (e.g. for the detectors LHCb and ATLAS listed in table 7.6).

Ar–CO_2–O_2 and Xe–CO_2–O_2. These gas mixtures are used in the drift tubes of the Outer Tracker of LHCb and the Transition Radiation Tracker (TRT) of ATLAS. As just mentioned, in both cases the deployment of CF_4 was discarded because of the ageing problems. Given the high radiation load at the LHC the addition of the strongly electronegative oxygen should reduce space charge effects. Although oxygen

Table 7.6 Operation parameters of some selected examples of drift chambers employing different gas mixtures.

Experiment	TASSO	ARGUS	OPAL	H1	HERA-B
detector	DC (central)	DC (central)	jet chamber	jet chamber	outer tracker
reference	[222]	[339]	[202]	[261,736]	[77]
gas	argon–ethane	propane–methylal (+H_2O)	Ar–methane–isobutane	Ar–ethane–ethanol	Ar–CF_4–CO_2
mixture	50 : 50	97 : 3 (+0.2)	88.2 : 9.8 : 2.0	50 : 50 : 0.8	65 : 30 : 5
pressure (bar)	1	1	4	1	1
B field (T)	0.5	0.8	0.435	1.15	0–0.8
HV (kV)	3.0	2.7	2.5–25	≈ 7	1.95
gain	$\approx 10^4$	$\approx 10^4$	10^4	$\approx 10^5$	$\approx 3.3 \times 10^4$

Experiment	LHCb	ATLAS	ATLAS	ALEPH	ALICE
detector	outer tracker	TR tracker	muon tubes (MTC)	TPC	TPC
reference	[87]	[4]	[4]	[126]	[85]
gas	Ar–CO_2–O_2	Xe–CO_2–O_2	Ar–CO_2 (+H_2O)	argon–methane	Ne–CO_2–N_2
mixture	70 : 28.5 : 1.5	70 : 27 : 3	93 : 7(<0.1)	91 : 9	85.7 : 9.5 : 4.8
pressure (bar)	1	1	3	1	1
B field (T)	0–0.85	2	0.15–2.5	1.5	0.5
HV (kV)	1.55	1.53	3.08	24	100
gain	5×10^4	2.5×10^4	2.0×10^4	3–5 $\times 10^3$	2.0×10^4

should rather be avoided in drift chambers, here it is tolerable because of the short drift lengths of the considered detectors. In the TRT xenon is employed instead of argon, making for an efficient detection of the transition radiation photons (see chapter 12 on transition radiation).

Ne–CO_2–N_2. This very special mixture was chosen for the TPC of the ALICE experiment (see section 7.10.10, table 7.8). The main component is neon instead of argon in order to minimise multiple scattering. The disadvantage compared to argon is its low ionisation density (only about 40%) which has to be compensated by a relatively high gas amplification. The addition of nitrogen improves the quenching properties of the gas mixture, which is accredited to the improved absorption of the dominant excitation line of neon at 16.8 eV [441]. A more practical reason for choosing nitrogen as additive is the fact that it is unavoidable to have nitrogen as contaminant in a gas system thus changing the gas properties. Since nitrogen is very difficult to filter out it is more convenient to have a portion of it to begin with and keep the proportions by adding the other components. On the other hand, because of the long drift paths of up to 2.5 m contamination by oxygen has to be kept below 5 ppm.

7.10.5.2 Choice of the high voltage

Efficiencies and resolutions can be optimised by choosing a gas amplification as high as possible which, however, increases the danger of spark discharges and ageing effects. Noble gases are particularly favourable because the voltage necessary to reach a certain amplification is comparatively low. For high rate experiments space charge effects can play a role so that a rather low gas amplification is required, in particular to keep the number of slowly drifting ions low.

In the examples of drift cells in figs. 7.34 to 7.36 the drift and amplification regions are not separated by electrodes. In this case the relation between the amplification field and the drift field is determined by the thickness of the anode wire. In a given chamber gas one obtains for different wire thicknesses the same amplification if the field on the anode wire surface is kept at the same value (more accurate conditions can be inferred from (7.27)). Since the field near the wire has a $1/r$ behaviour, increasing the radius requires the voltage to also be raised by the same factor if the same amplification is to be obtained. That is the reason to preferably use thin wires (typically 10–30 μm).

7.10.6 Measurement of the drift time

The arrival time of electrons generated by an ionising track is measured relative to a signal which has a fixed time distance to the ionisation process. There are different methods to obtain such a signal:

- a signal from a detector providing a fast time stamp, as for example from an organic scintillator (section 13.5.1);
- a time mark of an accelerator at which the particle reaction is recorded (e.g. at a collider the time of the beam crossing).

After electronic processing the position of this signal relative to the drift chamber signals is usually not known, at least not with sufficient precision, so that the time zero point, called t_0, has to be determined from the measured data. In general corrections for different flight times of the particles from the production to the ionisation process also have to be applied.

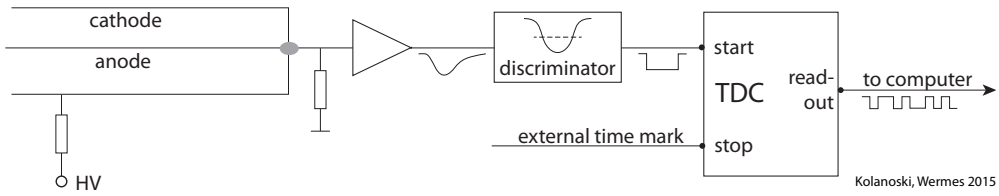


Fig. 7.43 TDC readout of a single drift chamber channel. The anode signal is passed via a connector (grey oval) to an amplifier and then to a discriminator which generates a digital signal when the input signal crosses a certain threshold. In this sketch the TDC is started by the drift chamber signal and is stopped by an external signal which is common to all channels (*common stop*).

For the measurement and digitisation of drift times there are two principally different methods:

– digitisation of the time interval between a start and a stop signal by a *time-to-digital converter* (TDC);
– digital recording of the analog signals with a so-called flash ADC (FADC = flash analog-to-digital converter) and determination of the drift time by analysing the signals.

A detailed description of digitisation methods is presented in section 17.7.1.

Drift time measurement with a TDC. A TDC (see section 17.7) digitises the time between a start and a stop signal (fig. 7.43). The analog signal of the anode is converted into a digital signal yielding the time marker by using a discriminator (section 17.5). The term *digital signal* refers to a pulse with a fixed amplitude (the amplitude height depends on the used electronic standard) whose leading and falling edges are in the ideal case infinitely sharp.

A *leading-edge discriminator* delivers a digital pulse at the output when the leading edge of the signal exceeds a preset threshold v_T or, respectively, falls below the threshold for negative signals (see the related fig. 17.18 on page 737). If the time constant of the signal shaping is long compared to the variation of the arrival times of the electrons at the anode, the threshold voltage corresponds to a certain number k of electrons necessary to generate a signal. The threshold determines the position resolution, as discussed in section 7.10.8.

The characteristics of a TDC (section 17.7.1) are the smallest time unit Δt corresponding to the *least significant bit* (LSB) and the number of bits available for the coding of the time measurement. For example, if the TDC has 10 bit at its disposal one says that the TDC has 1024 channels (these are 'channels' of the time units, not to be confused with 'readout channels'). If the LSB corresponds, for example, to a time unit of 1 ns this TDC can record drift times up to about 1 µs.

The natural way to run a drift chamber with TDC readout would be to start all TDCs by one fixed time marker (*common start*) and then to separately stop each channel through the individual signal of the respective drift chamber channel. Frequently, however, the *common stop* variant is used. The TDC is started by the drift chamber signal and stopped by a fixed time marker generated after a trigger has been formed. The common stop is delayed such that the full drift time range is covered. If no trigger and hence no stop signal is formed the TDC runs up to the maximal drift time and is then reset. This mode has the advantage that time is gained for a trigger

decision which then delivers the common stop signal. In common stop mode the times are inverted: the signals belonging to the shortest drift time are furthest away from the stop signal.

Drift time measurement with a flash ADC. A flash ADC (FADC) measures the momentary pulse heights of an analog pulse sequence with a fixed sampling frequency and stores the measured values consecutively in a memory (see section 17.7.1 and fig. 17.34). This memory contains at each point in time the pulse height sequence of a fixed recent time interval, corresponding to the depth of the memory (principle of a ring or stack memory). The memory is read out upon a trigger signal.

The characteristics of an FADC are the sampling frequency, the number of bits per sample and the depth of the memory. Typical values for the deployment of FADCs as drift chamber readout are sampling frequencies of 100–200 MHz with about 8 bit per sample and a memory depth of some 10 µs. The memory depth is matched to the time necessary for a trigger decision.

From the digitised signal sequence and a reference time the drift time can be determined using an appropriate software algorithm. Usually such an algorithm includes an interpolation between the bins of the leading edge of the signal in order to smooth the digitisation steps. Based on the interpolation function the time mark providing the optimal resolution is determined. This can, for example, be the time when the signal height reaches half of its maximum.

7.10.7 Space–drift-time relation

The central task in the analysis of drift chamber data is the determination of the relation between the measured drift time and the location where the particle traversed the cell.

For the uniform collection of the drift electrons it is optimal if the isochrones, that is, the positions of equal drift time (section 7.10.4), run parallel to the preferred direction of the tracks. This is the case, for example, for the monitor chamber in fig. 7.33, the TPC (section 7.10.10) and approximately also for the cells of the jet chamber (fig. 7.34). In general, a particle trajectory in a drift cell traverses a range of different isochrones thereby smearing the arrival times of the electrons at the anode (see the discussion of ionisation statistics in section 7.10.8).

In the following we initially assume that the position of the particle trajectory in a cell is the point where the track touches the isochrone with the shortest drift time and that the electrons drift nearly radially towards the anode, that is, the isochrones are approximately circles. If the second assumption is not fulfilled the space–drift-time relation (SDR) has to be separately determined for different angles of incidence.

The drift time from the position r to the anode is

$$t(r) = \int_{t_0}^{t} dt' = \int_{0}^{r(t)} \frac{dt'}{dr'} dr' = \int_{0}^{r(t)} \frac{1}{v_D} dr' \,, \tag{7.61}$$

where $v_D = v_D(r)$ is the drift velocity. The inversion of (7.61) yields the position r if the time t is measured:

$$r(t) = \int_{t_0}^{t} v_D \, dt' \,. \tag{7.62}$$

Usually the space–drift-time relation $r(t)$ is determined by a fitting procedure based on the measured tracks. The principal method will be discussed in the following.

Linear space–drift-time relation. The so-called *saturating gases* (see section 4.6.4.3 and figs. 4.9 and 4.10), such as argon–methane or argon–ethane mixtures, yield constant drift velocities over a wide range of field strengths ($v_D = const$). Then the position just depends on the measured drift time linearly (fig. 7.45(a)):

$$r(t) = v_D(t - t_0)\,. \tag{7.63}$$

An example of a nearly linear SDR for an argon–ethane mixture as drift gas is shown in fig. 7.44(a). For a linear SDR the TDC spectrum, that is, the distribution of the measured drift times expressed in TDC channels, is uniform between t_0 and t_{max} (fig. 7.45(a)). Isochrones with the same time difference correspond to the same difference of the distances from the anode.

Nonlinear space–drift-time relation. In general, the drift velocity is field-dependent yielding a nonlinear SDR. For example, the drift velocity in 'cool' gases (CO_2, DME, propane, etc., see section 4.6) increases proportionally to the electric field for field values which are interesting for drift gases (that means the mobility is about constant, as for ions). In such a case the isochrones no longer have a constant distance for a constant time interval but become denser with increasing distance from the anode wire (fig. 7.45(b)). Consequently, the TDC spectrum has more entries at small times (if the tracks are uniformly distributed in r) because more events fall into a given time interval at small times than at large times. An example of a distinctly nonlinear SDR (drift gas: Ar–CF_4–CO_2) is shown in fig. 7.44(b).

In the nonlinear case the SDR has to be determined iteratively. The iteration procedure uses at each step the SDR approximated in the previous step to determine from the measured drift times the hit coordinates to which a track is fitted. The deviations of the measured points from the fitted track, called residuals, are collected in bins of r, the drift distance. In the next iteration step the SDR will be corrected such that the mean values of the residuals for each TDC channel approach zero. The procedure is continued until convergence is reached, which means that the changes of the SDR remain smaller than a preset value. As starting values for the iteration one could use, for example, a linear SDR approximation, an SDR calculated from a measured drift velocity as a function of $E(r)$ or a previously determined SDR.

The final residual distributions can be used to determine the position resolution as a function of r. Usually, a Gaussian distribution of the residuals is assumed and the resolution is defined as the standard deviation of that distribution.

An elegant method is the derivation of the SDR from the frequency distribution of the TDC channels as we will explain in the following. For this purpose we consider the TDC spectrum in fig. 7.45(b) using the following notation:

- t_i: time assigned to the ith TDC channel (in TDC units);
- ΔN_i: counts in the ith TDC channel;
- $r_i = r(t_i)$: the distance from the anode related to the time t_i.

Usually the spectra of several similar drift cells are superimposed, thus yielding a higher statistical precision for the determination of the SDR. To begin with, however, it has to be assured that all cells have the same time zero point t_0. This is achieved by separately calibrating t_0 for each single cell (e.g. by moving the individual TDC spectra on top of each other).

If the cell is irradiated uniformly in r then the counting rate ΔN_i in the interval $\Delta r_i = r_i - r_{i-1}$ is proportional to the interval length:

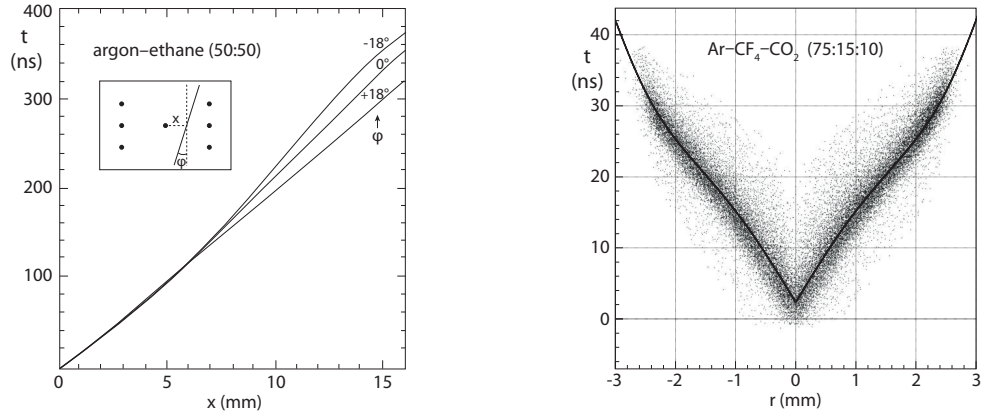
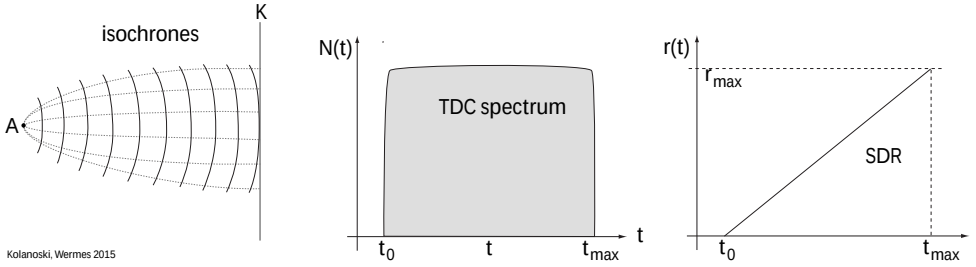
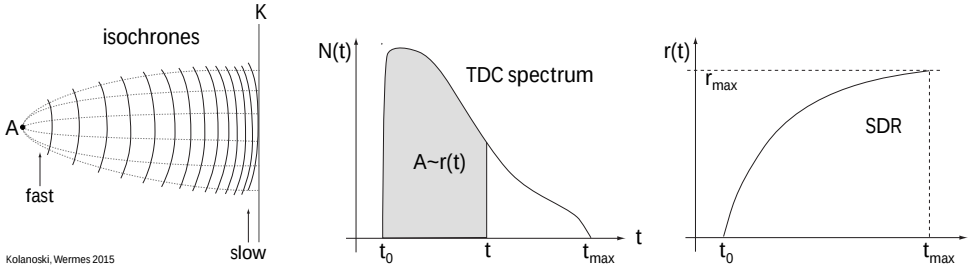


Fig. 7.44 Examples for space–drift-time relations (SDRs): (a) nearly linear SDR of a drift chamber operated with argon–ethane (from [222], with kind permission of Elsevier). The cells have the geometry as in fig. 7.36(b). For larger distances from the anode a correction due to the dependence on the angle of incidence φ of a track has to be applied (see insert in the figure). (b) SDR of the drift tubes of the Outer Tracker of the LHCb experiment (5-mm-diameter tubes) operated here for testing purposes with an Ar–CF_4–CO_2 gas mixture (source: CERN [151]); the gas actually used in the experiment is given in table 7.6. The figure shows the distribution of the r, t values measured with tracks and the derived averaged, nonlinear curve.



(a) Linear SDR: equidistant isochrones (left) lead to a uniformly distributed TDC spectrum (middle) and a linear space–drift-time relation (right).



(b) Nonlinear SDR: non-equidistant isochrones (left) lead to a TDC spectrum (middle) with relatively more entries at large distances of the isochrones (fast drift) and a nonlinear space–drift-time relation (right).

Fig. 7.45 Schematic representation of isochrones and TDC spectra for (a) linear and (b) nonlinear space–drift-time relations (SDRs).

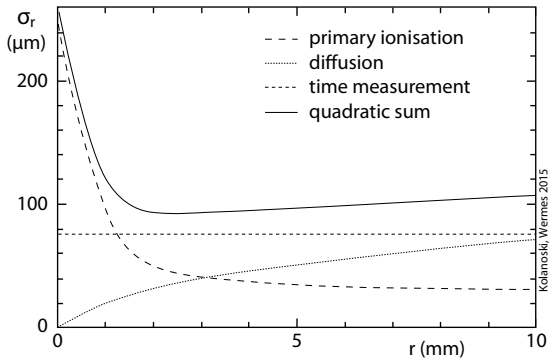


Fig. 7.46 Typical behaviour of the position resolution in a drift cell as a function of the distance of the track from the signal wire. The individual terms can be assumed to be statistically independent and can thus be quadratically added.

$$\frac{\Delta r_i}{r_{max}} = \frac{\Delta N_i}{N_{\rm tot}}\,, \tag{7.64}$$

where $N_{\rm tot}$ is the total number of entries in the TDC spectrum and r_{max} is the maximal radius in the cell (fig. 7.45). The distance r_i belonging to a TDC channel i is then obtained from the sum of the TDC entries up to channel i:

$$r_i = \sum_{k=1}^{i} \Delta r_k = \frac{r_{max}}{N_{\rm tot}} \sum_{k=1}^{i} \Delta N_k\,. \tag{7.65}$$

In the case of a continuous time distribution, (7.64) and (7.65) turn into their differential or integral form, respectively:

$$\frac{dr}{r_{max}} = \frac{dN}{N_{\rm tot}} \quad \Rightarrow \quad r(t) = \frac{r_{max}}{N_{\rm tot}} \int_{t_0}^{t} \frac{dN}{dt'} dt'\,. \tag{7.66}$$

An SDR obtained in this way may serve as starting value for an iterative improvement, as described above.

7.10.8 Contributions to the position resolution of drift chambers

The position resolution in drift chambers is limited by various effects which have different dependences on the drift path (fig. 7.46). At small distances from the anode the ionisation statistics is important, and at large distances the diffusion of the charge cloud moving with the velocity v_D towards the anode is the dominant resolution contribution. In addition there are effects which are rather independent of the drift path, such as mechanical and electronic uncertainties. In general one can assume that these contributions are statistically independent so that the total position uncertainty could be obtained by quadratic addition of the individual contributions:

$$\sigma(r) = \sqrt{\sum_j \sigma_j^2(r)}\,. \tag{7.67}$$

In the following the most important effects contributing to the position uncertainty are itemised according to their origin.

Drift gas. The properties of the drift gas influence in different ways the position resolution:

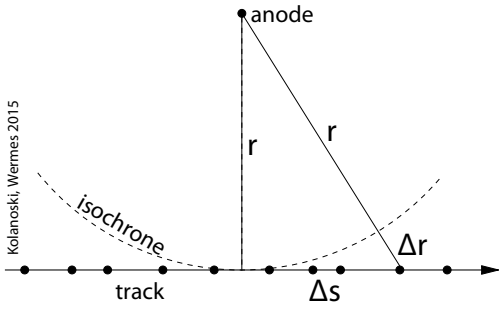


Fig. 7.47 Schematic representation of the effect of ionisation statistics for curved isochrones. See the explanations in the text.

(1) Diffusion: The influence of diffusion on the position uncertainty is determined by the width of the diffusion distribution in the drift direction. According to (4.73) on page 103 this width evolves proportionally to the square root of the drift time. If the ratio of diffusion coefficient to drift velocity is constant, which is always the case if the field is constant (see section 4.6.4.3), this yields a width proportional to the square root of the drift length:

$$\sigma_{\text{diff}}(r) = \sqrt{2Dt} = \sqrt{\frac{2D\,r}{v_D}} \propto \sqrt{r}\,. \tag{7.68}$$

In contrast, if the characteristic energy (see eq. (4.85)), which is the ratio D/μ of the diffusion coefficient and the mobility, is constant and the field has approximately the $1/r$ behaviour of a counting tube, then a linear dependence on the drift distance is obtained:

$$\sigma_{\text{diff}}(r) = \sqrt{2Dt} = \sqrt{\frac{2D\,r}{\mu E}} \propto r\,. \tag{7.69}$$

This behaviour can be observed for 'cool' gases, for example CO_2.

If N electrons with the same mean drift time contribute to the determination of the position and the mean of the charge cloud is measured, then the resulting position uncertainty is $\sigma_{\text{diff}}/\sqrt{N}$. In reality, however, the conditions for the application of this formula are usually not fulfilled. If the time of the crossing of a discriminator threshold is measured and if this requires k electrons, then actually the drift time of the kth electron is determined. In (C.9) in appendix C the position uncertainty of the kth cluster is given for a large total number N:

$$\sigma(r|N,k) = \sigma_{\text{diff}}\sqrt{\frac{1}{2\ln N}\left(\frac{\pi^2}{6} - \sum_{i=1}^{k-1}\frac{1}{i^2}\right)}\,. \tag{7.70}$$

In appendix C further details are discussed about the influence of diffusion on the position uncertainty.

(2) Ionisation statistics: The distances between primary ionisation clusters follow a random distribution along the track. If the tracks do not run parallel to the isochrones the signals of different ionisation clusters arrive at different times yielding a contribution to the position uncertainty which becomes smaller with increasing distance from the wire. This can be understood from fig. 7.47 for a circular or, more generally, a curved isochrone. The track touches the isochrone

Table 7.7 Position uncertainty through ionisation statistics in argon with a cluster density of $n = 29.4/\text{cm}$. The isochrones are here assumed to be circular. The contributions to the position resolution are given for different distances r of a particle trajectory from the anode and for different numbers k of clusters necessary for reaching the threshold.

r (mm)	σ_{ion} (μm)		
	$k = 1$	$k = 3$	$k = 5$
0	170	295	380
0.2	110	274	370
0.5	55	210	328
1.0	29	134	246
1.5	19	95	188

with the curvature radius r. If the next ionisation cluster which delivers the time signal is displaced on average by Δs from the tangent point, then the cluster lies on an isochrone with radius $r + \Delta r$ for which the following holds:

$$(\Delta s)^2 + r^2 = (r + \Delta r)^2 = r^2 + 2\Delta r\, r + (\Delta r)^2 \,. \tag{7.71}$$

Neglecting the term $(\Delta r)^2$ thus yields a $1/r$ proportionality of the position uncertainty:

$$\Delta r \approx \frac{\Delta s^2}{2r} \propto \frac{1}{r} \,. \tag{7.72}$$

In appendix D the variance of this expression and hence the contribution to the position uncertainty is calculated for the case that the discriminator threshold is passed with the kth ionisation cluster:

$$\sigma(r|k)^2 = \frac{k^3}{4n^2(4n^2r^2 + k^2)} \,. \tag{7.73}$$

Here n is the number of ionisation clusters per track length. For argon at normal pressure table 7.7 lists the contribution to the position uncertainty for different values of r and k. For example, for $k = 3$ and $r = 0.5\,\text{mm}$ the position uncertainty is $\sigma(r|k) = 210\,\mu\text{m}$. In this case, only beyond $r = 1.5\,\text{mm}$ are resolutions in the order of $100\,\mu\text{m}$ achievable. The table shows that it is advantageous to keep the number k of the clusters necessary for the signal to pass the threshold as small as possible.

(3) Pressure and temperature stability: If the drift velocity is strongly dependent on the particle density n, as is the case for 'cool' gases with $v_D \propto E/n$ (in eq. (4.113) with constant cross section), the position measurement becomes sensitive to pressure and temperature fluctuations. For an ideal gas, for example, we have

$$p = nkT \qquad \Rightarrow \qquad n \propto \frac{p}{T} \,. \tag{7.74}$$

As an example we consider drifting in a constant field with the drift velocity $v_D = 50\,\mu\text{m/ns}$. Then a 1% change of the pressure or the temperature, that is, $\Delta p = 10\,\text{mbar}$ or $\Delta T = 3\,\text{K}$, yields an error in the position measurement (without a correction) of about $100\,\mu\text{m}$ over a 1 cm drift path. However, if the working point lies in the plateau region of a saturating gas, as for example of an

argon–methane mixture (fig. 4.9), the dependence on pressure and temperature can become very small.

Mechanics. The achievable minimal mechanical uncertainties are about 10–50 µm depending on the size of the detector. Although this can be partially corrected for by alignment with reconstructed tracks (*software alignment*) some systematic uncertainty remains yielding a constant term to the position resolution:

$$\sigma_m = const\,. \tag{7.75}$$

Electronics. A detailed discussion of the following items can be found in chapter 17.

(1) Time walk: The point in time when the signal crosses the threshold, thereby determining the time mark, is dependent on the pulse height; see fig. 17.18 on page 737. This leads to a time smearing according to the variations of the pulse height called *time walk*. The contribution to the resolution can be position dependent if the pulse height distributions become broader at larger distances due to diffusion.

(2) Noise: Electronic noise (section 17.10) which is superimposed on the signals can lead to time smearing due to the variation of the pulse shape. In drift chambers this is usually not very critical in contrast to, for example, semiconductor detectors.

(3) Digitisation: The finite time interval Δt of a TDC channel contributes a timing uncertainty

$$\sigma_{TDC} = \frac{\Delta t}{\sqrt{12}}\,,$$

corresponding to the standard deviation of a uniform distribution in the interval Δt (see also eq. (E.4) on page 832). The respective influence on the position resolution increases linearly with drift velocity:

$$\sigma_t = v_D\,\sigma_{TDC}\,. \tag{7.76}$$

For $\Delta t = 1\,\text{ns}$, a linear space–drift-time relation and a drift velocity of $v_D = 50\,\mu\text{m/ns}$ yields a contribution to the position resolution of $\sigma_t = 15\,\mu\text{m}$.

(4) Time stability of the electronics: With increasing system size it becomes increasingly difficult to keep the electronics stable over the entire detector. A dispersion of the time calibration of the TDCs, called *inter-calibration error*, yields a constant contribution to the position resolution. For large systems it is important that the TDCs are equipped with a so-called *auto-calibration* system keeping the time-scale the same in the whole system.

7.10.9 Drift chambers in magnetic fields

The momentum of a charged particle can be determined by deflection in a magnetic field (see section 9.3). In a drift chamber the superposition of the electric drift field with a magnetic field can lead to the following effects (details in section 4.3.2):

– change of the drift direction;
– change of the drift velocity (usually the drift velocity decreases with increasing magnetic field);
– reduction of the diffusion transverse to the magnetic field.

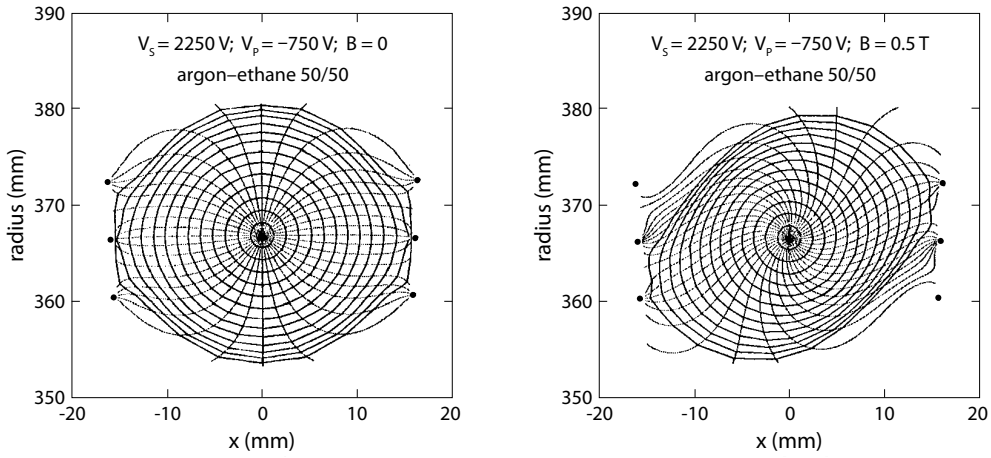


Fig. 7.48 Drift paths and isochrones in a drift cell of a cylindrical drift chamber (the same cell as in fig. 7.36(b); adapted from [234]). Left: without magnetic field, right: with magnetic field (0.5 T) parallel to the anode wires (perpendicular to the electric field).

The actual impact of these effects depends on the drift gas and on the direction of the magnetic field relative to the electric field.

The electron drift in electric and magnetic fields is discussed in section 4.3.2. Equations (4.54) to (4.56) give a general expression for the vector of the drift velocity. In another representation, reflecting better the relation to the field directions, the vector can be decomposed into three linearly independent components in the directions $\vec{E}$, $\vec{B}$ and $\vec{E} \times \vec{B}$. If one neglects the complex averaging described in section 4.3 the following expression can be given (see also eqs. (4.28) and (4.29) on page 96f):

$$\vec{v}_D^{\,B} = -\frac{\mu^{B=0}}{1+\omega^2\tau^2}\left(\vec{E} + \frac{\vec{E}\times\vec{B}}{B}\omega\tau + \frac{(\vec{E}\,\vec{B})\vec{B}}{B^2}\omega^2\tau^2\right). \tag{7.77}$$

Here $\mu^{B=0}$ is the mobility without magnetic field, $\omega = qB/m$ the cyclotron frequency and τ an average collision time of the electrons in the gas. The meaning of the term $\omega\tau$ becomes understandable by looking at the derivation in section 4.2.4.

In drift chambers at colliders $\vec{B}$ and $\vec{E}$ are usually oriented perpendicularly to each other so that the term $\vec{E}\vec{B}$ vanishes. On the other hand, in a time projection chamber (see next section) $\vec{E}$ and $\vec{B}$ are parallel so that in this case the term $\vec{E} \times \vec{B}$ vanishes. For these cases eqs. (4.57) and (4.63), respectively, apply. If the magnetic field is not parallel to $\vec{E}$ the drift lines and isochrones are deformed according to the Lorentz angle, as demonstrated by the example in fig. 7.48.

The space–drift-time relation (SDR; see section 7.10.7) has to account for the B-field. If the SDR determination is done using measured tracks, as described in section 7.10.7, the effect of the distorted isochrones is inherently included. If one or both fields are inhomogeneous the SDR becomes position dependent and its determination correspondingly difficult.

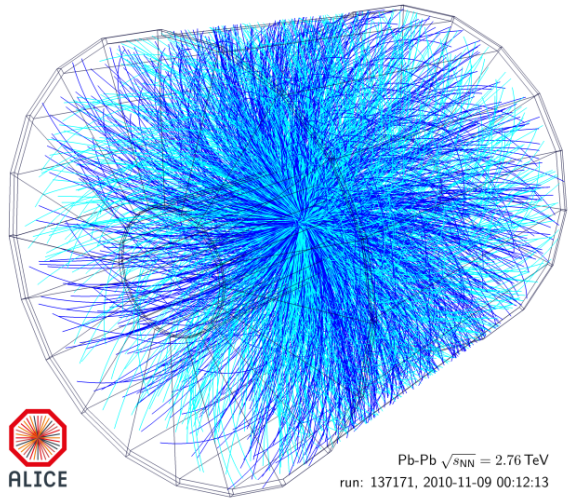


Fig. 7.49 A Pb–Pb collision event with a very high particle multiplicity (about 2500 charged particles) recorded by the TPC of the ALICE experiment at the LHC (Source: CERN/ALICE Collaboration [972]).

7.10.10 Time projection chamber (TPC)

A drift chamber variant with a large sensitive volume is the so-called *time projection chamber* (TPC) [741, 611, 523], which was invented and developed in the 1970s and employed for the first time in the PEP4 experiment at the electron–positron storage ring PEP (see table 2.2 on page 13). The largest TPC currently is used in the ALICE experiment. Each of the more than 100 measurement points along a track is recorded three-dimensionally yielding event pictures similar to those from bubble chambers (fig. 7.49).

7.10.10.1 Principle

The principle of a TPC is similar to the chamber depicted in fig. 7.33. The electrons, produced along a particle trajectory, drift in a homogeneous drift field over large distances (metres) to an electrode grid behind which a signal is formed by gas amplification in a separate structure. The track coordinates are determined along the drift direction through the drift time and perpendicular to the drift direction employing a signal readout with segmented electrodes. Usually, the tracks are deflected by a homogeneous magnetic field aligned parallel to the electric field.

Like the original TPC of the PEP4 experiment, most TPCs at colliders have the shape of a cylinder which encloses the interaction point and is oriented parallel to the beams. Figure 7.50 shows the functional principle. The volume is divided perpendicular to the cylinder axis into two halves by a planar electrode lying on negative high voltage relative to both end plates of the cylinder. In the resulting field the electrons, being produced when an ionising particle passes, drift towards the end plates over distances of the order of metres. The large drift distances pose high requirements on the purity of the gas; in particular the gas has to be free of electronegative contaminations which would capture electrons. For example, the oxygen fraction has to remain below some 10^{-5}.

In contrast to the loss of electrons the loss of ions in the drift space is not relevant

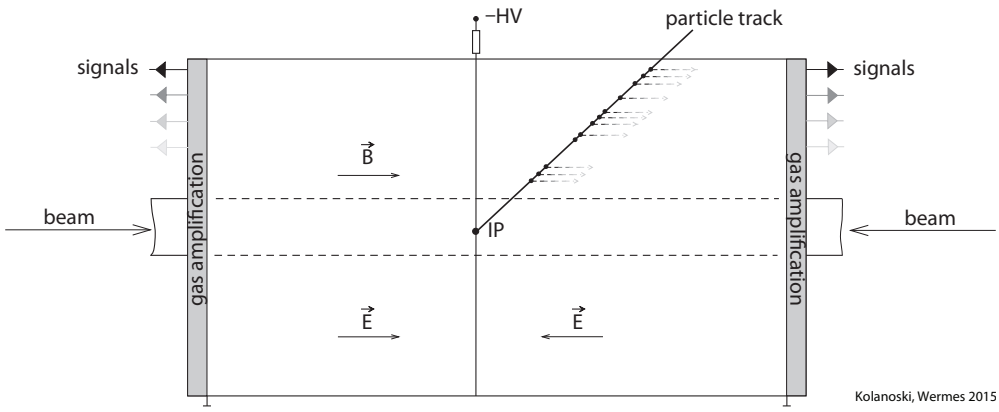


Fig. 7.50 Principle of a time projection chamber (TPC) at a storage ring. Shown is a cut through the cylinder axis, defined by the beams which collide in the interaction point IP. Between the planar electrode in the middle of the chamber and the end caps, an electric field of about 500 V/cm is maintained. The electrons of the ionisation clusters drift up to 1 m and more to the end caps where the $r\phi$ coordinates of the arriving charges are measured by a gas amplification detector, for example a proportional chamber. The z coordinate (along the beam and the E-field) is determined by the drift time. Each $r\phi$ measurement provides additionally a dE/dx value which can be used for particle identification.

for the properties of a TPC (see, however, the paragraph on the prevention of back-drift of ions in the next section).

The longitudinal coordinate (z-axis = cylinder axis) is derived from the drift time; the $r\phi$ coordinates in the projection of the track curvature are measured at the cylinder end plates employing position-sensitive detectors (see below). When collecting the total charge each $r\phi$ measurement delivers in addition a dE/dx value which can be used for particle identification (see section 14.2.2).

Electric field. In order to establish a homogeneous electric field in the whole cylinder volume, a potential is generated along the cylinder walls which uniformly decreases from the end electrode to the middle electrode. This is achieved with the help of a field cage made of annular conductor strips along the inner walls of the cylinder barrel. The decreasing potential on the conductors is provided by a resistor chain, as in fig. 7.33. In fig. 7.51 showing the ALICE TPC the ring structures are visible on the inner walls of the inner and outer barrels.

Magnetic field. Usually the TPC volume is located in a magnetic field which is aligned to the electric field as parallel as possible (other field directions are rather rare, but see the last example in the list in section 7.10.10.4). The component perpendicular to the electric field should be less than 10^{-4} of the parallel component. Then the drift velocity in (7.77) contains in good approximation only the terms in the direction of the E-field, or equivalently of the B-field, and no contribution from the $\vec{E} \times \vec{B}$ term. Besides providing a momentum dependent deflection, the magnetic field also causes a strong reduction of the transverse diffusion of the drifting electrons according to (4.119) on page 118 (see also the footnote 8 on page 118):

$$D_T(B) = \frac{D_T(0)}{1 + \omega^2\tau^2}\,. \tag{7.78}$$

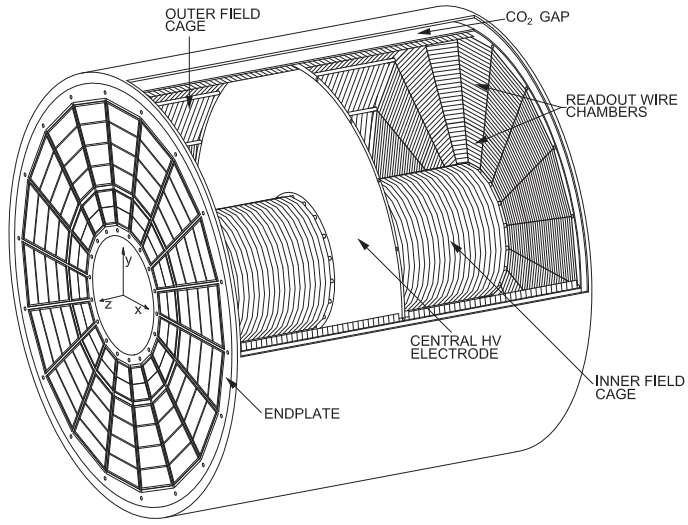


Fig. 7.51 Perspective drawing of the electrode arrangement around the drift volume of the ALICE TPC (from [85], with kind permission of Elsevier). The disk in the middle of the TPC dividing the chamber into two cylindrical volumes are on high voltage. The readout plane at the end plates are subdivided into 18 azimuthal sectors each with two readout chambers.

Table 7.8 Parameters and properties of selected TPCs (data from [523], the ALICE entries are partly updated according to [85]).

Parameter/experiment	PEP4 [741]	ALEPH [126]	ALICE [85]
operation start	1982/1984	1989	2009
inner/outer radius (m)	0.2/1.0	0.35/1.8	0.85/2.5
max. drift length (L/2) (m)	1	2.2	2.5
magnetic field (T)	0.4/1.325	1.5	0.5
gas mixture	Ar/CH_4	Ar/CH_4	$Ne/CO_2/N_2$
	80:20	91:9	85.7:9.5:4.8
gas pressure (atm)	8.5	1	1
drift field ($V\,cm^{-1}\,atm^{-1}$)	88	110	400
e^- drift vel. ($cm/\mu s^{-1}$)	5	5	2.7
$\omega\tau$ (section 4.3.2)	0.2/0.7	7	<1
pad size $w \times L$ (mm^2)	7.5×7.5	6.2×30	4×7.5; $6 \times 10/15$
max. dE/dx measur./track	183	344	159
gas amplification	1000	3000–5000	20 000
pads, total number	15 000	41 000	560 000
resolutions:			
$\sigma_{r\phi}$ (μm)	130–200	170–450	800–1100
σ_z (μm)	160–260	500–1700	1100–1250
double-track separation (mm) (transverse/longitudinal)	20	15	13/30
σ_p/p^2 (GeV^{-1}) (p large)	0.0065	0.0012	0.022
dE/dx (%) single tracks/in jets	2.7/4.0	4.4/–	5.0/6.8

Hence the diffusion constant $D_T(0)$ for the case without magnetic field is reduced by a factor $1/(1+\omega^2\tau^2)$ for drifting in the direction of the magnetic field. This is the essential reason why even for drift distances of more than a metre good position resolutions can be achieved. To obtain a low transverse diffusion the cyclotron frequency ω, and thus the magnetic field, and the collision time τ have to be as large as possible. The parameter $\omega\tau$ is listed for the TPC examples in table 7.8.

7.10.10.2 Signal readout

The signals are usually read out by detectors with gas amplification positioned at the end plates of a TPC. While initially mostly MWPCs with pad readout were used, more recently the newly developed micro-structured gaseous detectors like GEMs (section 7.9.2) or micromegas (section 7.9.3) are preferred.

Figure 7.52 depicts the principle of a pad MWPC integrated into the end plates of a TPC. The cross section in (a) shows the anode plane between two cathode planes. The cathode plane at the side where the drifting electrons arrive is formed as a wire plane ('cathode grid') and on the other side as a pad plane. Typical pad dimensions are in the range of millimetres, see table 7.8. In the anode plane often *field wires* are strung between the anode wires. The field wires are put at a lower potential than the anode wires, thereby achieving a better focusing of the electrons onto the anode (see fig. 7.52(c)).

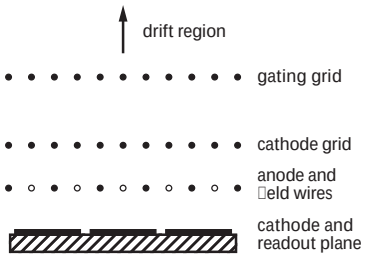
Prevention of back-drift of ions. Before the electrons are amplified in the MWPC they have to pass a gate electrode, *gating grid* (fig. 7.52(c)), which can be opened or closed for drifting charges by putting the grid on different potentials. The purpose of the grid is the necessary reduction of space charge effects due to ions which drift from the amplification region into the drift region towards the middle plane. Since the ions drift much more slowly than electrons they accumulate in the drift volume and thus deform the electric field. The gate will be opened upon an event trigger and closed after about the maximal electron drift time. Obviously this method is only reasonable if the maximal drift time is small compared to the mean time between two events.

With micro-structured gaseous detectors, like GEMs and micromegas (section 7.9), the ion drift can in the first place be reduced such that TPCs can be operated without a pulsed gate even in high rate experiments. With several GEM layers (fig. 7.31) the amplification in the first layer can be kept relatively small. Because of diffusion the ions in the further amplification steps have only little chance to back-drift through the small holes of the GEMs. In micromegas the electrons are channelled through the tight micro-mesh (fig. 7.32) which, however, keeps back the ions generated in the amplification volume, in a way similar to that for GEMs. The deployment of GEMs and micromegas in TPCs is, for example, described in [28, 857].

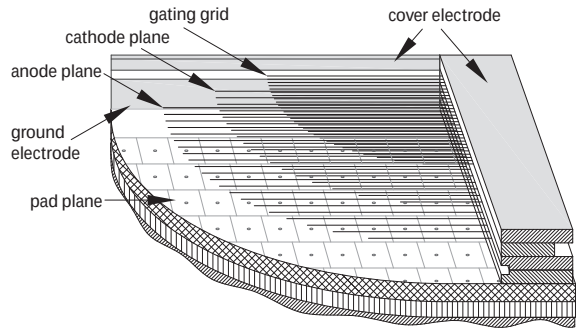
7.10.10.3 Resolutions

***r*ϕ resolution.** The achievable position resolution in the $r\phi$ coordinate (in table 7.8 the entry $\sigma_{r\phi}$) depends on the type of readout. If the charge is distributed over several electrode elements (strips or pads), a position resolution in the order of 200 µm can be achieved by calculating the centre of gravity or using similar methods. The main contributions to the resolution are the transverse diffusion, the pad size and the orientation of the track relative to the anode wire and the pads (for details see [671, 217, 523]).

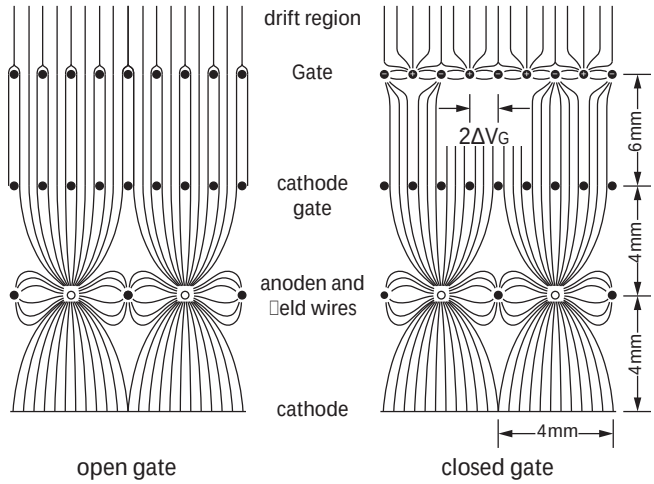
According to (4.73) and (7.78) the width of the charge cloud for a diffusion coefficient D_T and a drift path z is given by



(a) Electrode pattern of a pad MWPC [671].



(b) Structure of the pad and wire layers of the ALICE TPC [85].



(c) Gating electrode, left: open, right: closed [730].

Fig. 7.52 TPC readout with pad MWPC. (a) Scheme of a pad MWPC with anode, cathode and grid layers formed from wires and a layer with pads which record the induced signal of the electrons amplified at the anode. (b) Readout chamber with a pad structure of the ALICE TPC. (c) The potential of the *gating grid* can be switched so that it is open for the drifting electrons or closed to reduce the back-drift of the much slower drifting ions. Reprint of the figures with kind permission of World Scientific (a) and Elsevier (b, c).

$$\sigma_{D_T} = \sqrt{\frac{2D_T(0)\,z}{v_D}}\,\sqrt{\frac{1}{1+\omega^2\tau^2}}\,. \tag{7.79}$$

This contribution to the resolution can be kept small by a smart choice of the gas, according to (7.78) with large $\omega\tau$ (see e.g. the respective parameters of the ALEPH TPC in table 7.8). In addition the measurement uncertainty decreases with the number N of the electrons contributing to the signal about proportionally to $\sigma_{D_T}/\sqrt{N}$ if the arriving charge can be averaged over several pads; see appendix E.

The effect of increasing the gas pressure p on the diffusion contribution to the $r\phi$ resolution can be estimated as follows. The mean free path λ for electron scattering is proportional to $1/p$ leading to $D_T(0) \propto 1/p^{3/2}$ according to (4.111). In a magnetic field this pressure dependence is partially compensated by the collision time τ in (7.79),

which is given by $\tau = \left\langle \frac{\lambda}{v} \right\rangle$ (see section 4.6.4.3). With the random velocity v being proportional to $1/\sqrt{p}$ according to (4.108) and $\lambda \propto 1/p$ we find that τ is proportional to $1/\sqrt{p}$. For the determination of the centre of gravity of the charge cloud in $r\phi$ the uncertainty is roughly proportional to $\sigma_{D_T}/\sqrt{N}$. Since the number N of contributing electrons is also proportional to p we obtain for the case $\omega\tau \gg 1$:

$$\frac{\sigma_{D_T}}{\sqrt{N}} \propto \frac{1}{p^{3/4}}\,. \tag{7.80}$$

Example. As an example we want to consider the transverse diffusion in the ALEPH TPC using the data in table 7.8, as also discussed in [523]. In [523] one finds for the ALEPH TPC:

$$\sqrt{\frac{2D_T(0)}{v_D}} = 600\,\frac{\mu\text{m}}{\sqrt{\text{cm}}}\,. \tag{7.81}$$

Over the maximal drift length of 2.2 m this yields without magnetic field ($\omega\tau = 0$) a width of $\sigma_{D_T} = 8.9\,\text{mm}$ and with a magnetic field of $B = 1.5\,\text{T}$, corresponding to $\omega\tau = 7$, a width of $\sigma_{D_T} = 1.3\,\text{mm}$. For N measured electrons the centre of gravity is by a factor $1/\sqrt{N}$ more accurately localised than the position of a single charge. Since in the ALEPH TPC the 'specific ionisation' is 90/cm, for a radial pad length of 3 cm the diffusion cloud of a straight track projection parallel to the pad can be localised with an accuracy of

$$\frac{\sigma_{D_T}}{\sqrt{N}} = 5\,\mu\text{m}\,\sqrt{z/\text{cm}} \qquad (= 74\,\mu\text{m for } z = 2.2\,\text{m})\,. \tag{7.82}$$

In general the number N depends on the projection of the track onto the pad.

***z* resolution.** In the longitudinal direction the resolution is limited by the longitudinal diffusion; see appendix C. Since for TPCs one can assume a constant drift velocity v_D, after a drift distance z the charge cloud has the width (eq. (C.2) for the coordinate z):

$$\sigma_{D_L}(z) = \sqrt{\frac{2D_L}{v_D}}\sqrt{z}\,, \tag{7.83}$$

where D_L is the longitudinal diffusion constant. If the drift time is determined as the average of all arrival times of the N electrons of a charge cloud, the contribution to the position resolution is

$$\sigma_z = \frac{\sigma_{D_L}(z)}{\sqrt{N}}\,. \tag{7.84}$$

If, however, the time is determined by the crossing of a discriminator threshold upon the arrivial of the kth electron, the corresponding formula in (C.9) holds for the position resolution. The respective resolutions σ_z are listed in table 7.8 for the selected TPCs.

***dE/dx* resolution.** The resolution of a dE/dx measurement (see section 14.2.2) depends on the number of electrons which contribute to a measurement point and on the number of measurement points. The number of electrons depends on the ionisation density, hence on the gas mixture and the pressure, and on the size of a pad corresponding to a measurement point. Larger pads provide a more accurate dE/dx measurement, however, at the expense of a worse $r\phi$ resolution, so that an optimum

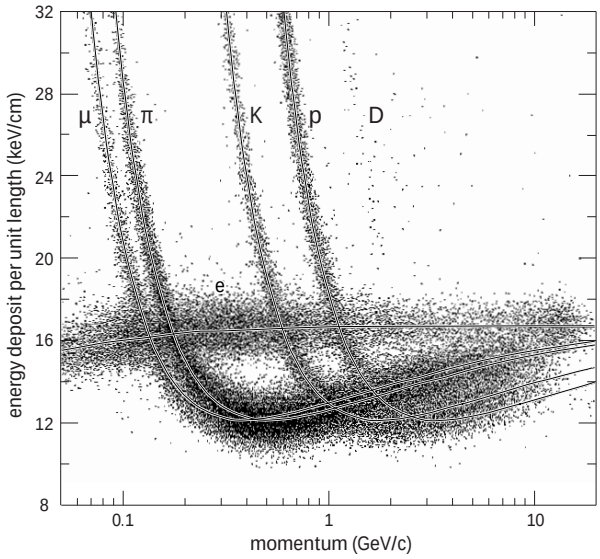


Fig. 7.53 Particle identification by dE/dx measurements in the PEP4 TPC (from [746]). Each point represents the mean value of all energy loss measurements for just one track. See also the description of fig. 14.11 on page 556.

has to be found. A very good resolution of about 3% was achieved with the PEP4 TPC with 183 measurement points per track and a pressure of 8.5 atm (table 7.8 and fig. 7.53). This good result is mainly due to the high pressure because the energy loss dE/dx is according to (3.25) on page 30 proportional to the gas density, which in turn is, in the relevant region, proportional to the pressure.

7.10.10.4 Other TPC variants

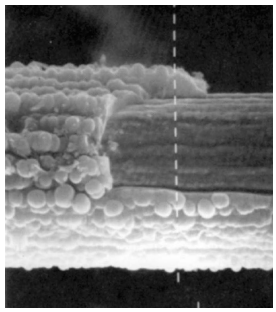
In the preceding part of the section we dealt with the 'typical' TPC: a cylinder volume surrounding the interaction region of an experiment at a collider ring, as displayed in figs. 7.50 and 7.51. However, there are many variants of the TPC principle which are deployed, for example, in non-accelerator experiments.

Examples.

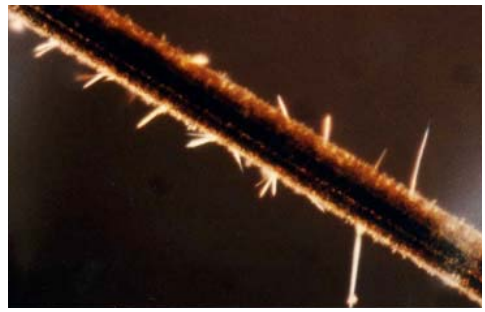
- In section 16.7.2 a detector for dark matter search is described which is based on a TPC filled with liquid xenon.
- The ICARUS experiment in the Gran Sasso laboratory (see section 2.3) deploys a particularly large TPC filled with liquid argon for the detection of neutrino oscillations [91].
- In an experiment at the Paul-Scherrer Institute (Switzerland) the capture of muons by protons is studied using a hydrogen TPC operated at 10 bar [369].
- Cylindrical TPCs also exist where the electric field is oriented perpendicular to the magnetic field and the readout chambers are attached to the cylinder barrel (see e.g. [47]). While this layout has advantages for the track reconstruction the achievable resolutions are in general worse.

7.11 Ageing effects in gaseous detectors

The term *ageing* includes phenomena in detectors which in general lead to a negative change of the detector behaviour after some time of operation. Such negative effects can be:



(a) Deposit of a polymerisate.



(b) Polymerisation filaments ('whiskers').

Fig. 7.54 Microscopic pictures of typical ageing phenomena (from presentation slides related to [852]).

- reduction of the gas amplification,
- growth of dark currents,
- appearance of self-sustaining currents,
- corrosion of electrodes,
- damage of insulating electrode supports and other mechanical structures.

Nearly always such ageing effects are caused by radiation exposure and here especially by the related increased charge collection at the electrodes. In gaseous detectors with high electric fields for gas amplification, ageing effects arise mainly from the decomposition of the gas in the amplification avalanche and the subsequent deposition of the various chemical components on the electrodes (fig. 7.54). Therefore, the substances responsible for ageing are mainly complex molecules, like hydrocarbons, which are employed in gas mixtures for adjustment of the working point or as quenchers (see section 7.5). Often the problems also arise from contaminations of the gas system. The deposits on the electrodes are preferentially polymerisates, either insulating or conductive.

By performing extensive radiation tests on prototypes one tries to ensure the radiation resistance of detectors. During operation the detector parameters are usually continuously monitored for changes which could point to ageing effects. Despite intensive research and an ever-improving understanding of the ageing processes, much in this field is based on pure hands-on experience and often test results are not reproducible. A good overview of the difficulties is given by the reports at the 'ageing workshops' [581, 531].

Radiation damage in semiconductor detectors is discussed in section 8.12. There some numbers are also given for radiation fluences (time integrated particle fluxes) at the LHC.

7.11.1 Measure of 'age' and ageing effects

Integrated charge. The most important measure of the 'age' of a chamber is the charge which was accumulated per length of the anode wires or strips. There are chambers which have been developed to cope with integrated charges in the order of several C/cm. Assuming about 10^5 electrons per signal pulse this means that during its lifetime the chamber has seen more than 10^{14} particles per cm wire length, corresponding

to 1 MHz/cm for a running time of 10^8 s (this is effectively about 10 years). If the drift cell has a width of 1 cm (±0.5 cm drift distance) this corresponds to a particle flux of 1 MHz/cm^2 or a fluence of 10^{14} particles per cm^2 within 10 years. This corresponds roughly to the LHC conditions at radius of about 20 cm from the beam.

In arrangements of parallel plates, such as the RPC (section 7.7.3), the age is given as integrated charge per area. 'Good' chambers can stand charges of about 10 C/cm^2 and more. The same charge per area is obtained for a wire chamber cell with 5 mm diameter at a charge per wire length of about 5 C/cm. However, because of the usually higher gas amplification of RPCs the number of particles corresponding to the charge is for RPCs about a factor 10–100 smaller than for drift cells.

Dependence on the type of radiation. Traditionally the accumulated charge normalised to the length or area of an electrode was considered a sufficient measure for the 'age' of a detector. However, measurements with detectors of high-rate experiments have shown that other parameters, like the type of radiation, the duration of the radiation exposure, the gas exchange rate or the extension of the irradiated region, contribute to the determination of the lifetime. In particular it was shown that radiation with high ionisation density in space and time, as for example generated by slow heavy ions, can have a much more severe effect than the same dose applied, for example, by X-rays. Correspondingly, the damage of microstrips of an MSGC shown in fig. 7.28 was triggered by highly ionising slow particles as dedicated tests with α radiation have proven. In [140] laboratory tests with X-rays are reported where chambers survived several C/cm without damage. Chambers of the same type, however, exposed to radiation generated by a proton beam in a target became unusable after only one thousandth of the charge accumulated with X-rays (the cause was the *Malter effect*, see section 7.11.2).

Amplification and chamber current as control variables. The relative changes of the gas amplification and the chamber current with the accumulated charge are usually recorded during chamber operation in order to recognise ageing effects early; see for example fig. 7.55. The change of the gas amplification G is taken relative to the accumulated charge per unit length, Q:

$$\eta_G = -\frac{1}{G_0}\frac{dG}{dQ}\,. \tag{7.85}$$

In a 'healthy' chamber the current I_{HV}, extracted from the power supply, is proportional to a rate α_r of the radiation level (which has to be measured with an external device). Hence the quantity to be monitored is

$$\eta_I = -\frac{1}{(I_{HV}/\alpha_r)_0}\frac{d(I_{HV}/\alpha_r)}{dQ}\,. \tag{7.86}$$

The current is usually employed as the control variable for triggering the automatic safety shutdown of a detector.

7.11.2 Formation of polymerisates

Methods of plasma chemistry are employed for the industrial production of polymerisates with the aid of discharges in hydrocarbon gases (or gases containing hydrocarbons) [1013]. Hence it might not be surprising that gas amplification avalanches in hydrocarbon gases or mixtures can generate deposits on electrodes. For example,

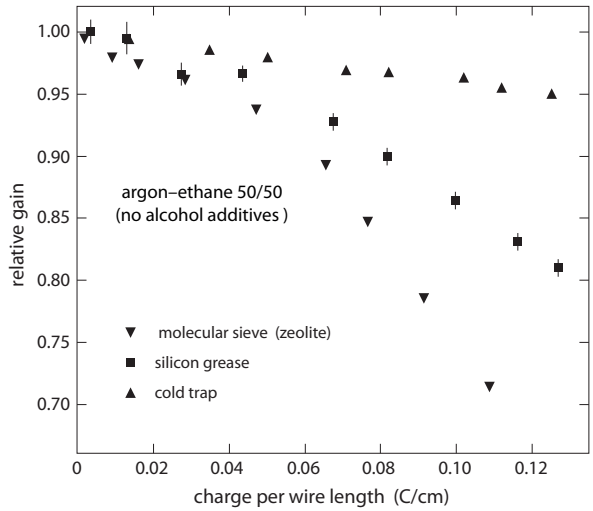


Fig. 7.55 Test measurements on ageing phenomena in the central drift chamber of the CDF detector at the Tevatron (adapted from [204], with kind permission of Elsevier; to keep the figure comprehensible less tests than in the original are shown here). The drift gas is argon–ethane (50:50) without an additive for prevention of deposits in order to be more sensitive to the influence of different contaminations. With the addition of alcohol the ageing rate is at least an order of magnitude lower. For the tests the contaminations have been systematically introduced into the gas system: a molecular sieve (zeolite), silicon grease, frequently used for sealings, and a cold trap for catching volatile contaminations. The ageing effects caused by these contaminations are relatively dramatic and in all cases not acceptable for the experiment.

dissociation of methane through electron collisions has a threshold of 4.5 eV while the ionisation threshold lies at 12.6 eV. For argon, to which often methane is added as quenching gas, the ionisation threshold lies at 15.8 eV. Hence, in order to form an avalanche by secondary ionisation, the electrons have to pass the energy region where dissociation in methane can occur.

In the dissociation processes of hydrocarbons, very reactive radicals, which are molecules with unpaired valence electrons, are produced. These combine to form molecule chains, thus forming polymerisates which can accumulate on the electrodes. For example, the CH_2 radical of methane is produced by electron impact according to the reaction [957]:

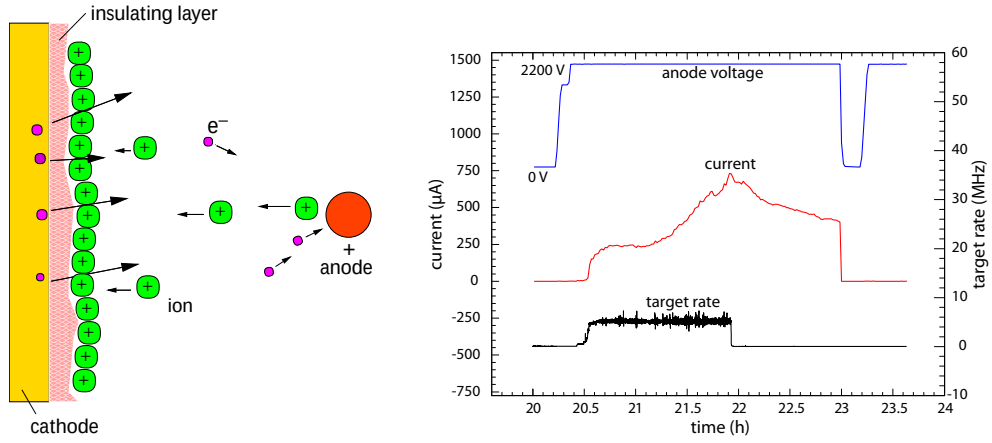
$$e^- + CH_4 \rightarrow CH_2 + H_2 + e^- \tag{7.87}$$

The CH_2 radical is the entry point for the formation of polyethylene with the simplified structural formula $[-H_2C-CH_2-]_n$, which is well known as a good insulator. Since the CH_2 radical has a large dipole moment it preferentially attaches to the polarising electrodes where polymerisation layers build up by chain formation.

7.11.2.1 Deposits on the anodes

Microscopic pictures of deposits on anode wires are shown as examples in fig. 7.54.

Conductive deposits. Deposits of a conductive polymerisate on the anode (fig. 7.54(a)) leads to a reduction of the amplification due to a swelling of the wire. By



(a) Malter effect. (b) Malter current.

Fig. 7.56 (a) Schematic illustration of the Malter effect [870]. Insulating deposits on the cathode lead to accretion of ions on this layer and hence the generation of a field between this charge layer and the cathode. Because of the very small distance of the charge to the cathode the field can become so strong that electrons are emitted from the cathode. The electrons drift towards the anode and generate yet more ions through gas amplification and thus a continuous current, the *Malter current*. (b) Characteristics of a Malter current (from [75], with kind permission of Elsevier). For a constant target rate (here interactions of protons in a nuclear target) the chamber current is initially constant and then rises due to the incipient Malter current. Switching off the target rate, the current no longer returns to zero as it did before the onset of the Malter current.

continuous monitoring of the amplification this effect can be noticed early on (see eq. (7.85)).

Insulating deposits. Deposits of an insulating or only slightly conducting polymerisate on the anodes leads to the build-up of a counter-field by charge accumulation on the insulating layer and hence to a drop of the amplification.

Formation of spiky structures. The formation of polymerisation filaments on the anode, called *whiskers* (fig. 7.54(b)) leads to point discharges, from occasional sparks to permanent dark currents.

7.11.2.2 Deposits on cathodes, the Malter effect

Insulating deposits on cathodes lead to the so-called *Malter effect* [687] schematically depicted in fig. 7.56(a). Due to accretion of ions on the insulating layer a field is generated between this layer and the cathode which can become so strong that electrons are extracted out of the cathode (*thin film field emission*). As long as the potential of the anode is higher than that of the ion layer, the electrons can drift through the ion layer to the anode and generate there yet more ions through gas amplification and thus a permanent current, the *Malter current*. In fig 7.56(b) the characteristic behaviour of the Malter current is presented. The accretion of ions on the insulating layer is ruled by the equilibrium between the neutralisation of the ions by electrons and amplification of the electrons at the anode. An increase of the ion layer decreases the potential

difference between the ion layer and the anode and hence also the amplification at the anode.

7.11.2.3 Measures against deposits

There are some general rules concerning how deposits on electrodes can be avoided:

– Avoid hydrocarbons. However, hydrocarbons are good quenchers and in drift chambers necessary for the adjustment of the drift velocity. Often CO_2, which is not prone to polymerisation, can serve as a substitute though it is not optimal as a quencher. A possibility for improvement of the quenching properties is the admixture of CF_4 which, however, has other problems (see section 7.11.3).
– Avoid contaminations of the gas. Figure 7.55 shows as an example the amplification reduction as an effect of contaminations. Stable detector operation requires scrupulous cleanliness of the gas system and during the detector fabrication the absolute avoidance of out-gassing materials (glue, electronic boards, solder tin, ...). In particular silicon has to be avoided because it forms non-volatile compounds which are prone to polymerisation. Amongst others silicon is contained in silicone oil which is sometimes used for bubbling small amounts of additives into the detector gas through a *bubbler*.

  The gas system, constructed preferentially from stainless steel pipes with oil-free connectors and without out-gassing seals, should be tight against the environment. If in large systems this is not absolutely achievable a slight gas overpressure can help. Contamination can be monitored by gas spectroscopy. However, since this requires a considerable effort it is only implemented in large systems in which the gas is purified and fed back in a closed loop.
– Addition of non-polymerising gases. Even if hydrocarbons cannot be avoided because of their good properties as quenchers, deposits on the electrodes can be suppressed by additives to the gas. In order not to transport dangerous charged molecules and radicals to the electrodes the charges should be transferred to molecules which are not prone to polymerisation. In charge exchange reactions the positive ions can be neutralised by electrons initially bound to molecules with lower ionisation potential. Alcohols (ethanol, propanol), methylal, water and even oxygen have proven to be well suited for such charge transfer processes (see also section 7.5 on the composition of chamber gases).

When deposits have already built up under radiation exposure, a chamber can possibly still be cured by flushing with reactive gases like oxygen or CF_4. Sometimes the addition of water can make an insulating layer conductive, although with the danger that leakage currents over the supports of the electrodes increase. Only extensive tests under conditions as realistic as possible can ensure to some degree that a certain measure promises success and does not lead to even more problems. However, in most cases at least the radiation intensity and/or duration are not realistic since a test can usually not last as long as the respective experiment but has to be carried out in an accelerated mode.

7.11.3 Damage on electrodes and chamber structures

Mechanical structures like electrodes or the insulating mounts of electrodes can also be attacked by radiation. In particular through the decomposition of fluorine-containing gases like CF_4 or freon, etching gases or liquids can build up. For example, hydrogen

fluoride (HF) can be generated which is very aggressive in combination with possibly present water (hydrofluoric acid is industrially employed for etching of glass and silicon). Damages through etching are observed both for wire electrodes (including the peeling of the gold coating) as well as for electrode surfaces, for example of RPCs (section 7.7.3) [957, 852]. Damage of the electrodes or electrode mounts can lead to deformations of the electric field configuration, resulting in increased dark currents and sparking.

In glasses with low conductivity which carry conductive microstructures as electrodes, like for MSGCs (section 7.9.1), the build-up of increasing currents (ion currents) under irradiation has been observed [542]. The currents distort the fields between the electrodes and thus make it more difficult to maintain a constant gas amplification (see also section 7.9.1 on MSGCs).

8
Semiconductor detectors

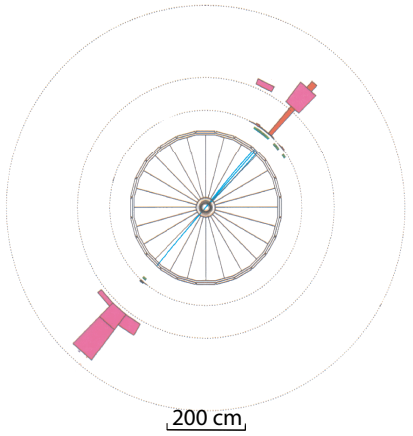
8.1 Introduction

Since the early 1960s semiconductor detectors have been used in nuclear physics, in particular for gamma ray energy measurement (see e.g. [616]). In this chapter, however, we focus on spatially sensitive semiconductor detectors, developed and used in particle physics since the 1980s. Similar to the invention of multiwire proportional chambers twenty years earlier, which revolutionised experiments by electronic recording of particle tracks (see chapter 7), a leap in measurement quality also resulted from the development of position-sensitive semiconductor detectors with electrode structuring in the rang 50–100 μm. With these detectors the precision of position measurements could be improved by up to two orders of magnitude as compared to the then mostly used wire chambers. This, for the first time, made an electronic measurement of secondary vertices and thus of lifetimes of heavy fermions possible. The following two examples demonstrate applications for precision vertex detectors:

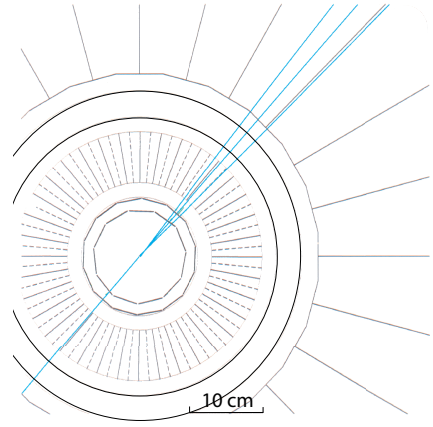
Example 1. τ leptons decay with a lifetime of $\tau_\tau \simeq 290\,\text{fs}$ in final states with one charged particle (e.g. $\tau^- \to e^-\overline{\nu}_e\nu_\tau$) or several charged particles (e.g. $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$). The length of the flight path before the decay is

$$l = \gamma\beta c\tau\,, \tag{8.1}$$

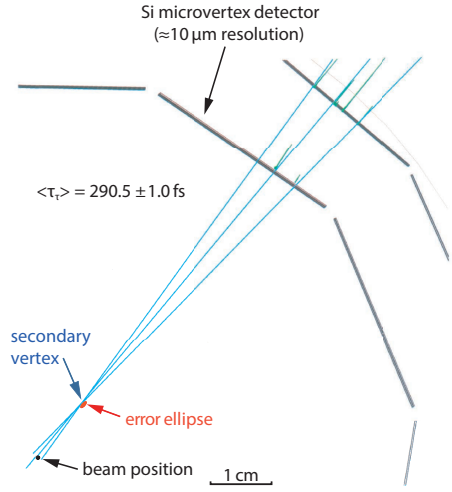
where βc is the particle's velocity and γ is the Lorentz factor for the transformation from the τ rest system to the laboratory system. In the electron–positron collider LEP at centre-of-mass energies around 91 GeV τ^+ and τ^- have been pair-produced in the reaction



(a) Reconstruction of a τ-pair event.



(b) Enlarged detail (20:1).



(c) Enlarged detail (200:1).

Fig. 8.1 Event display of the reaction $e^+e^- \to Z^0 \to \tau^+\tau^-$ in which one of the two τ leptons decays into three pions. Panels (b) and (c) show enlarged details demonstrating the precise measurement of track hits in the silicon microvertex detector and the recognition of a so-called 'secondary vertex' (OPAL detector at the e^+e^- collider LEP, source: CERN).

$$e^+e^- \to Z^0 \to \tau^+\tau^- .$$

For τ leptons with mass m, energy E and momentum p one obtains for $\gamma\beta$ in 'natural' units ($\hbar = c = 1$):

$$\gamma\beta = \frac{E}{m}\,\frac{p}{E} = \frac{p}{m} = \frac{45.5\,\text{GeV}}{1.78\,\text{GeV}} \simeq 25.6 \quad \Rightarrow \quad \gamma\beta c\tau \simeq 2.2\,\text{mm}\,.$$

Hence the typical distance between production and decay of the τ leptons lies in the millimetre range. Figure 8.1 shows an event of this type with an enlarged detail of the interaction region and the decay vertex. By employing silicon microstrip detectors,

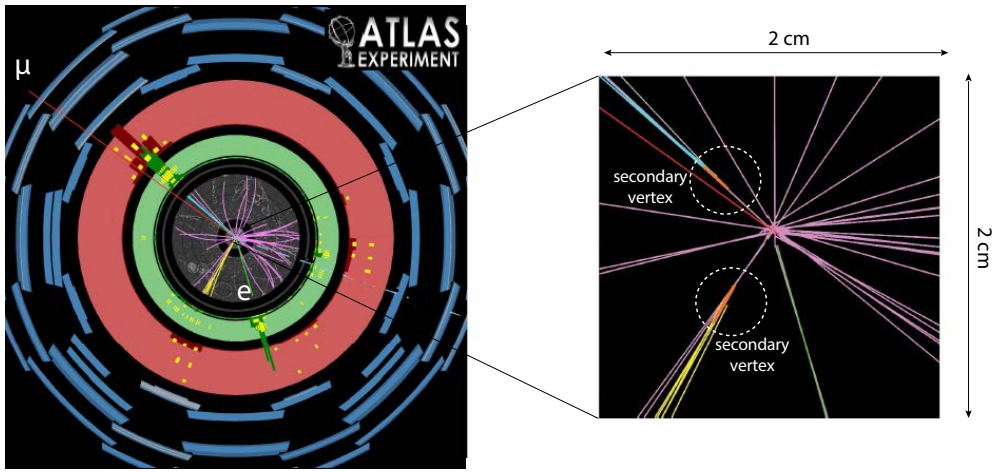


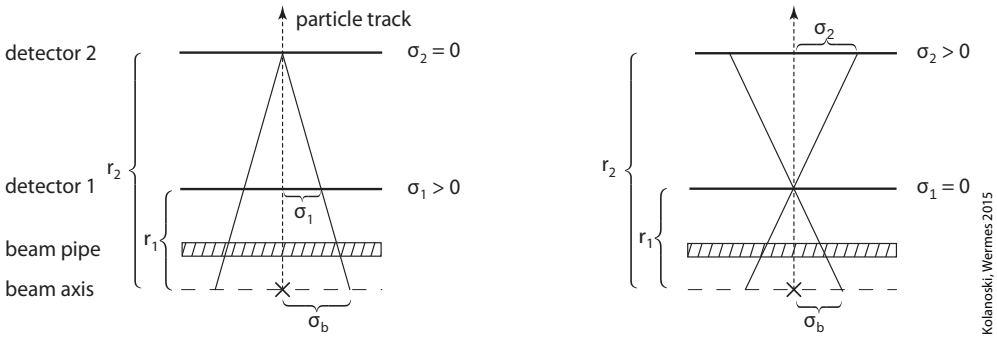
Fig. 8.2 Top–antitop quark pair production in a pp collison at the LHC (ATLAS experiment). Both top quarks decay into a charged lepton ($e^{\pm}$ or $\mu^{\pm}$, respectively) plus (anti)neutrino (invisible) and a b (or $\bar{b}$) quark having lifetimes of order picoseconds. In the enlarged detail one recognises the secondary vertices of the b-quark jets which emerge displaced from the collision point. Source: CERN.

particle tracks measured in large detectors with dimensions of several metres could be extrapolated to the interaction point with a precision in the range 10–100 μm.

Example 2. Similarly demanding is the recognition of secondary vertices in particle bundles, called *jets*, for example in the reaction $pp \to t\bar{t} + X \to (b\,\mu^+\nu)\ (\bar{b}\ e^-\bar{\nu}) + X$ shown in fig. 8.2, which is a computer reconstruction of a collision event in an LHC detector. In this case one wants to distinguish jets originating from a secondary vertex from those which originate from the primary vertex, which means from the collision point of the reaction. The abbreviations t and $\bar{t}$ as well as b and $\bar{b}$ in fig. 8.2 denote top and bottom quarks and their antiparticles, respectively. A top quark decays promptly into a bottom quark plus additional quarks which form hadrons in the final state. The bottom hadrons have lifetimes in the order of 1.5 ps and decay after a momentum dependent path length given by (8.1) which can reach distances of millimetres to centimetres for typical momenta at the LHC.

A simplified detector. In order to demonstrate which parameters are crucial for a *microvertex detector* consisting of layers of silicon detectors we consider a simple two-layer *microstrip detector*, installed outside of the beam pipe near the interaction point of a collider experiment (see fig. 8.3). The layers shall be cylindrically arranged around the beam direction at distances r_1, r_2 and have position resolutions σ_1, σ_2 in the plane perpendicular to the beam. The problem will be simplified by assuming planar detector modules (as in fig. 8.1) and considering straight tracks passing the layers perpendicularly.

First we assume that detector 2 is perfect ($\sigma_2 = 0$) and detector 1 has the resolution $\sigma_1 > 0$. Then the error σ_b of the impact parameter, that is, the distance of closest approach to the interaction point in the plane shown in fig. 8.3(a), is obtained as a scaled image of the position error according to



(a) Detector 2 assumed to be perfect. (b) Detector 1 assumed to be perfect.

Fig. 8.3 Simplified two-layer microstrip detector in planar geometry shown in the projection perpendicular to the microstrips which provide the resolution. The symbol × marks the interaction point and σ_b designates the measurement error due to the track extrapolation to the interaction region.

$$\frac{\sigma_b}{\sigma_1} = \frac{r_2}{r_2 - r_1} \,.$$

Inversely, assuming detector 1 to be perfect (fig. 8.3(b)), we obtain:

$$\frac{\sigma_b}{\sigma_2} = \frac{r_1}{r_2 - r_1} \,.$$

By quadratically summing the two respective resolutions and by adding a term σ_{ms} due to multiple scattering in the material of the beam pipe and the first detector layer we obtain the total resolution as

$$\sigma_b^2 = \left(\frac{r_2}{r_2 - r_1} \, \sigma_1 \right)^2 + \left(\frac{r_1}{r_2 - r_1} \, \sigma_2 \right)^2 + \sigma_{\text{ms}}^2 \,. \tag{8.2}$$

For the effect of multiple scattering on the vertex resolution see section 3.4 with fig. 3.33 on page 69 and section 9.4.6. From (8.2) we draw the following conclusions:

- The coefficient of σ_1 is always larger than the one of σ_2. Therefore the resolution of detector 1 should be as good as possible.
- The 'lever arm' $r_2 - r_1$ should be as large as possible to make the image scaling factor small.
- r_1 should be small and therefore the first layer of a vertex detector should be as close to the interaction point as possible.
- In order to keep the effect of multiple scattering small the beam pipe and detector 1 should constitute as little radiation lengths as possible to the traversing particles, which means the material should have low atomic number and should be thin.

Semiconductor detectors in combination with ASIC chips (application specific integrated circuit, see section 17.6) allow these requirements to be satisfied with resolutions of a few micrometres. Compared to gaseous detectors (chapter 7) semiconductor detectors mainly differ in terms of the following characteristics:

- The density of solids is larger than that of gases by about three orders of magnitude. Therefore, at the same detector thickness a traversing charged particle deposits a correspondingly higher ionisation energy in the detector ($\frac{dE}{dx} \propto \rho$) which is available

for charge carrier generation. However, the probability for multiple scattering and for undesirable background reactions is also correspondingly larger since the radiation length is much smaller.

– In semiconductors, the energy deposited by ionising radiation releases electron–hole pairs, rather than electron–ion pairs as in gases, which corresponds to raising electrons to higher energy levels such that they and their 'hole' counterparts are freely movable. The average energy necessary for the generation of an electron–hole pair in a semiconductor is about five times smaller than for the generation of an electron–ion pair in a gas. Therefore more charge per deposited energy is released, resulting in a higher achievable resolution for the measurement of the deposited energy.
– The detectors need not be installed in a container, as in the case of gas as detector medium.
– By micro-structuring[1] the electrodes, a high granularity and an excellent position resolution can be achieved.
– However, large detector volumes, respectively large areas, are difficult and very cost intensive to be accomplished (see e.g.the tracking detectors of the LHC experiments [4, 298]).

Semiconductor detectors are deployed in a multitude of areas, apart from nuclear and particle physics, for example, for X-ray diagnostics in biology, medicine or material science as well as for autoradiography (e.g. fig. 8.71) and real-time dose monitoring in radiotherapy.

8.2 Semiconductor basics

In this section we summarise the basic properties of semiconductors as far as they are relevant for detectors. For more detailed and advanced presentations we allude to the respective literature, for example [929, 928, 881, 248].

8.2.1 Semiconductor materials for detectors

In terms of their electrical conductivity, all solids can be divided into conductors, semiconductors and insulators. The conductivity ranges of the three groups are shown in fig. 8.4 together with corresponding examples of some group representatives. The typical values of resistivities of semiconductors lie between 10^{-3} and $10^{8}\,\Omega\,\mathrm{cm}$. Table 8.1 gives an overview of the different semiconductor materials which are generally also of interest for semiconductor detectors. Besides silicon and germanium as pure element

Table 8.1 Element and compound semiconductors.

Element	Compounds			
	IV–IV	III–V	II–VI	IV–VI
(C) Si Ge	SiC SiGe	AlP, AlAs, AlSb, BN, GaAs, GaP, GaSb, InAs, InP, InSb	CdS, CdSe, CdTe, ZnS, ZnSe, ZnTe, HgS, HgSe, HgTe	PdS, PbTe

[1] Meanwhile micro-structuring is also employed in gaseous detectors; see section 7.9 on *micro pattern gas detectors*.

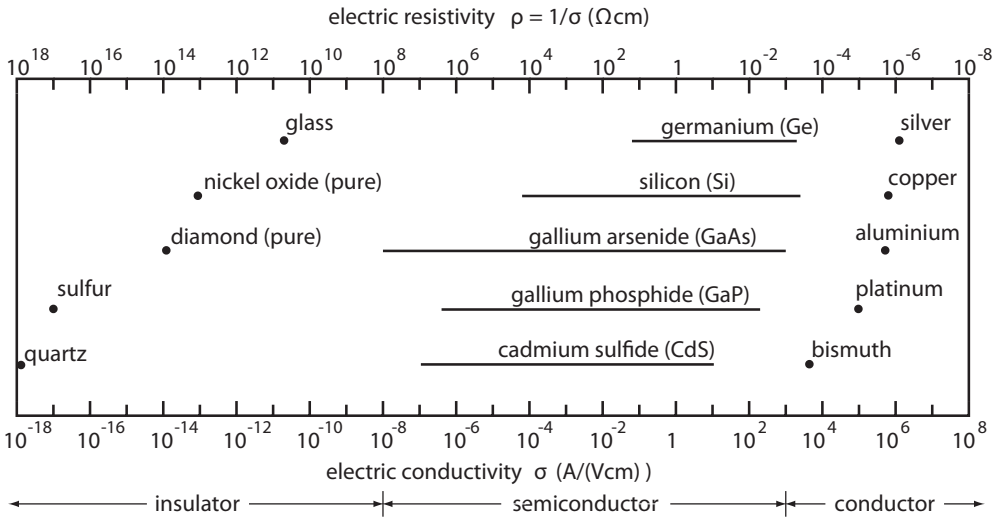


Fig. 8.4 Conductors, semiconductors and insulators (adapted from [926]).

semiconductors there are compound semiconductors. They are composed of elements of the third and fifth main group of the periodic system (so-called III–V semiconductors) or of elements of the second and sixth main group (II–VI semiconductors).

For the detection of particles or radiation (X-rays or gamma rays) the most important semiconductors are silicon (Si), germanium (Ge), gallium arsenide (GaAs) and cadmium telluride (CdTe). The relevant properties of these semiconductors, together with diamond, are listed in table 8.2. Silicon and germanium have long been used as detector materials and have been most intensively studied. More recently, GaAs and CdTe detectors have been developed, in particular because of their high atomic number with correspondingly good absorption properties for X-rays. Synthetic diamond, despite being classified as insulator, has also become of interest as a detector material because of its radiation hardness.

Silicon is by far the most widely used material for semiconductor detectors. Readout chips are nearly exclusively produced from silicon. After oxygen, silicon is the second most abundant element of the Earth's crust with a mass fraction of about 28%. However, it only occurs in compounds such as silicon dioxide (SiO_2, also called silica) or silicate minerals. Pure silicon is highly reactive and instantaneously oxidises in air.

Elementary silicon is produced from SiO_2 with a purity level of more than 95 % by reduction with base metals like aluminium or magnesium or, on the industrial level, with pure carbon in an arc furnace (carbothermal reduction). The silicon so produced is used for many industrial applications. However, for microelectronics and photovoltaics higher purities are required which can be reached through various processes. For example, in a first step of such a process elementary silicon reacts with hydrogen chloride to form trichlorosilane ($SiHCl_3$) which can be highly purified by repeated fractional distillation. In the next step the trichlorosilane is reduced in a hydrogen atmosphere on hot silicon rods.

The silicon rods can be further purified by the *float-zone process* (FZ). In this process a narrow zone of a vertical silicon rod is molten with the help of an induction

Table 8.2 Properties of silicon, germanium, gallium arsenide, cadmium telluride and diamond. D = diamond lattice, ZB = zinc blende lattice, temperature dependent quantities given at 300 K.

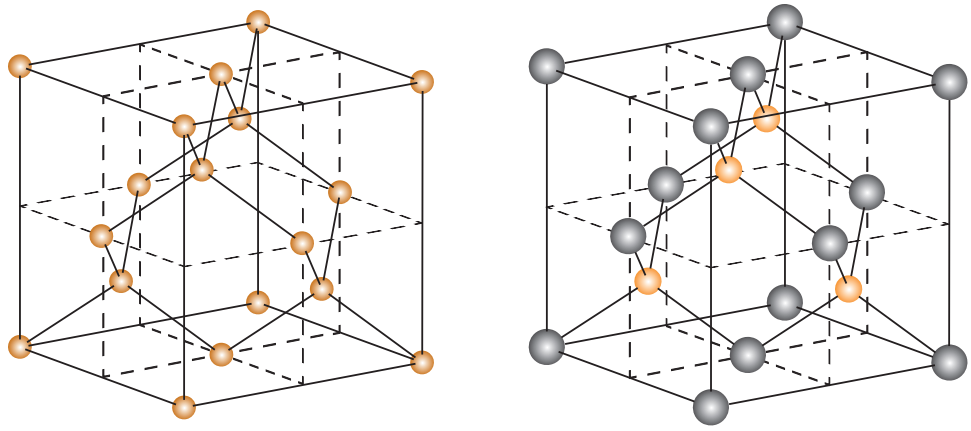
Property	Si	Ge	GaAs	CdTe	Diamond
atomic number (Z)	14	32	31/33	48/52	6
atom mass (u)	28.09	72.60	72.32	120.0	12.01
density ρ (g/cm^3)	2.328	5.327	5.32	5.85	3.51
crystal structure	D	D	ZB	ZB	D
lattice constant (Å)	5.431	5.646	5.653	6.48	3.57
semiconductor type	indirect	indirect	direct	direct	indirect
band gap E_G (eV)	1.12	0.66	1.424	1.44	5.5
intr. carrier density (cm^{-3})	1.01×10^{10}	2.4×10^{13}	2.1×10^{6}	10^{7}	≈ 0
resistivity (Ω cm)	2.3×10^{5}	47	10^{8}	10^{9}	$\approx 10^{16}$
dielectric constant (ϵ)	11.9	16	13.1	10.2	5.7
radiation length X_0 (cm)	9.36	2.30	2.29	1.52	12.15
average energy for (e/h) creation (eV)	3.65	2.96	4.2	4.43	13.1
thermal conductivity ($\frac{W}{cmK}$)	1.48	0.6	0.55	0.06	>18
mobility ($\frac{cm^2}{Vs}$)					
electrons μ_n	1450	3900	8500	1050	≈1800
holes μ_h	500	1800	400	90	≈2300
lifetime					
electrons τ_e	>100 µs	∼ms	1–10 ns	0.1–2 µs	≈100 ns
holes τ_h	>100 µs	∼ms	20 ns	0.1–1 µs	≈50 ns

coil. The rod is then moved through the coil and with this the molten zone is also moved up or down towards the ends. The process takes advantage of the smaller concentration in the growing crystal than in the melt for most impurities. The impurities thus prefer to remain in the melt and can then be swept out at the ends of the rod, leaving behind a crystallised purer silicon. If properly done the newly crystallising material can be obtained as a single crystal. Another way of growing mono-crystals employs the so-called *Czochralski process* (Cz [924]): a rod-mounted seed crystal is dipped into molten high-purity silicon and is slowly pulled upward under steady rotation. At cool down the silicon solidifies as a single crystal.

For application as detectors the final high-purity crystals (resistivity $\rho >$ kΩ cm) are cut into thin wafers of 200–300 µm thickness and 10 cm or 15 cm diameter. Wafers for electronic chips can have lower resistivity and are delivered in sizes up to 30 cm and thicknesses up to 800 µm. FZ silicon is preferred when ingots with very low oxygen content are required. FZ silicon is generally purer than Cz silicon. However, FZ ingots can rarely be grown to sizes larger than 15 cm in diameter. The vast majority of the commercially grown silicon is Cz silicon due to its better resistance to thermal stress, production speed, and low cost. It also features a high oxygen concentration, a property that is beneficial for radiation tolerance (see section 8.12.1).

8.2.2 Crystal lattice and energy bands

Crystal lattices. Silicon and germanium crystallise in a diamond lattice structure (fig. 8.5(a)), gallium arsenide and cadmium telluride and other compound semiconduc-



(a) Diamond lattice (Si, Ge). (b) Zinc blende (GaAs, CdTe).

Fig. 8.5 Cubic face-centred lattice structure of some important semiconductors. (a) Primitive cell of the diamond lattice with identical atoms. The carbon atoms sit at the corners, at the centres of the cube faces and at the centres of those four of the eight octants that are not direct neighbours (adapted from [986]). (b) In the zinc blende lattice of compounds the four centres of the non-connected octants are replaced by the other atom. Cubic lattices feature only one lattice parameter or *lattice constant*, which are 5.43 Å (Si), 5.66 Å (Ge), 5.65 Å (GaAs), 6.48 Å (CdTe), and 3.567 Å (diamond) (from [992]).

tors in the zinc blende structure (fig. 8.5(b)) or in a zinc sulfide structure (hexagonal). In contrast to metals, semiconductors do not form close-packing of equal spheres, hence the number of closest neighbours per atom is comparatively small with typical lattice spacings of about 5.5 Å. For the semiconductors addressed in this book they are 5.43 Å (Si), 5.66 Å (Ge), 5.65 Å (GaAs), 6.48 Å (CdTe) and 3.567 Å (diamond) (from [992]).

In the diamond lattice each silicon atom is covalently bonded to four nearest neighbours arranged at the corners of a regular tetrahedron [537]. A lattice cell can be imagined as two *face-centred cubic* (fcc) cells penetrating each other, one shifted by a quarter of the cube's diagonal against the other (see e.g. [120]). In this lattice all atoms are of the same type whereas in a zinc blende structure each of the two shifted cubes consists of the other atom type of the compound (fig. 8.5(b)). In GaAs, for example, the gallium atoms lie in the centre of a tetrahedron formed by the arsenic atoms and vice versa. Each atom has four nearest neighbours with which it forms covalent bonds. The packing density is only half that of a *body-centred cubic* lattice (bcc).

For some applications the cut direction through the crystal is important, for example, to obtain a surface which provides the densest packing of atoms. The crystal orientation is described by *Miller indices* which represent the orientation of lattice planes by a triplet of numbers (---) (see e.g. [926,120]). In cubic lattices, for instance, (100) corresponds to a plane parallel to the y–z plane, while (111) denotes a plane diagonal through the cube. Figure 8.6 shows specific crystal planes which are often taken as cut planes. The (111) plane is the plane with the highest density of atoms in an fcc lattice.

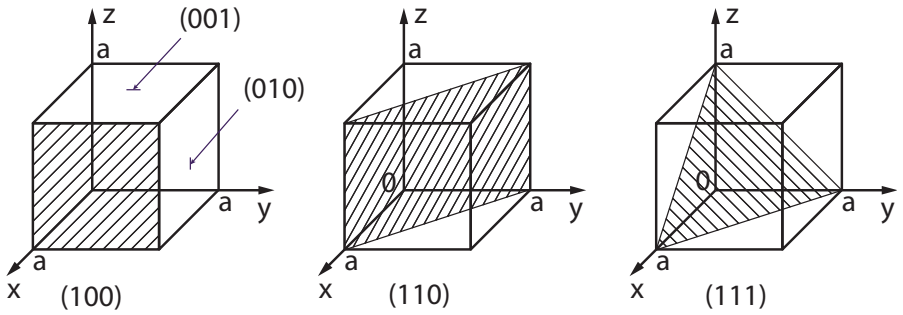


Fig. 8.6 Specific lattice planes in cubic crystals which for semiconductors are often the cutting planes. (---) are *Miller indices* denoting the crystal plane orientation (see text).

Energy bands. Owing to the dense periodic arrangement of the atoms in a solid state lattice the energy levels of individual atoms are split due to the influence of many neighbouring atoms in the vicinity. The energy levels of some level groups lie energetically so dense (meV) that one speaks of *energy bands* which are separated from each other by a *band gap* when the level groups have a certain distance in energy. The two highest lying energy bands, called the valence band (VB) and the conduction band (CB), govern the electrical conduction properties of semiconductors. Within a band the energy levels are so dense that transitions to unoccupied levels are easily possible if the band is not completely filled. In this way the conduction properties merely depend on the band occupation which in turn depends on the relative energetic position of the bands. Figure 8.7 displays a rough schematic overview of the energy-band structures of conductors, semiconductors and insulators.

In insulators the valence band electrons have very strong bonds between neighbouring atoms which do not break easily. The Pauli principle allows at most two electrons with opposite spins in an energy level. The valence electrons of neighbouring atoms are not free and hence do not contribute to conduction. All states of the valence band are occupied, whereas the conduction band states are empty. Corresponding to the strong interatomic bonds in insulators there exists a very large energy gap between

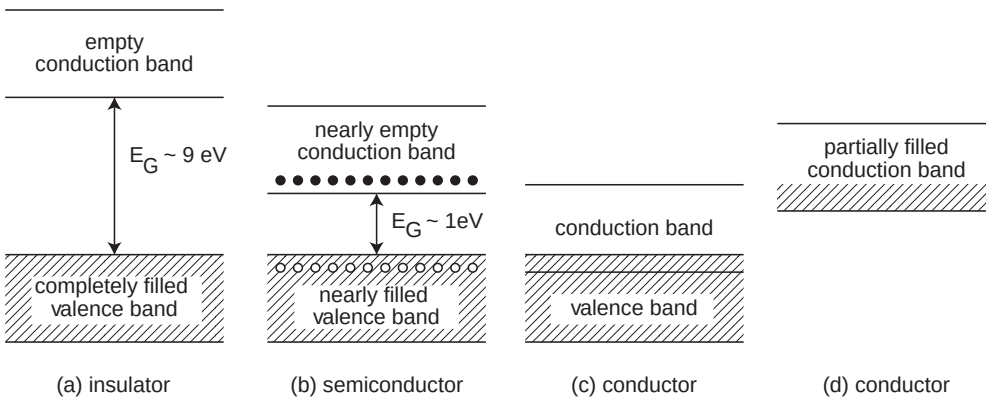


Fig. 8.7 Schematic energy-band structure of insulators (a), semiconductors (b) and conductors (c,d). E_G is the energy distance of the band gap.

the valence band and the conduction band (typically the gap is about 9 eV, fig. 8.7(a)). It is highly improbable that by thermal excitation an electron rises from the valence band to the conduction band. Current flow is impossible.

In semiconductors the bonds between neighbour atoms are less strong than in insulators leading to a smaller energy-band gap (1.12 eV in silicon, fig. 8.7(b)), such that it can be overcome by thermal excitations or by external electric fields. The bond breaks and a free electron and a free hole remain. Hence freely mobile electrons exist in the conduction band (and holes in the valence band) which in an electric field lead to current flow.

For conductors (metals) the conduction band is either partially filled (fig. 8.7(d)), or valence and conduction band overlap (fig. 8.7(c)) without a band gap in between. Transitions between levels inside the bands are easily possible and current conduction requires only minimal energy.

The energetic distance between the bands, the size of the band gap E_G, depends on the lattice spacing. Temperature or pressure therefore influence the gap size (see e.g. [929]). With increasing pressure and rising temperature, the gap size of semiconductors changes. For Si it is 1.12 eV at room temperature, at $T = 0\,\mathrm{K}$ it is 1.17 eV and at $T = 800\,\mathrm{K}$ the gap shrinks to 0.92 eV. For GaAs the respective values are 1.42 eV (300 K), 1.52 eV (0 K), and 1.19 eV (800 K). Pressure decreases the distance of atoms in the lattice and hence also the band gap by about 2 meV per bar. At very low temperatures most electrons (*all* electrons would only be the case at $T = 0\,\mathrm{K}$) of a semiconductor are in the valence band and do not contribute to conduction. At higher temperatures, however, some electrons are excited into the conduction band by thermal energy and the semiconductor becomes conductive. The electrons leave holes in the valence band which also contribute to the conductivity. They are interpreted as positive charge carriers.

Direct and indirect semiconductors. For a more comprehensive representation of the energy-band structure the energy–momentum (E–$\vec{k}$) relation of the electrons must be considered as a function of the crystal orientation. For *indirect* semiconductors, as for example for Si or Ge, the electrons in the troughs of the conduction band (in E-$\vec{k}$ space) have a different *crystal momentum* $\hbar k$ (for a definition see page 265) than the holes in the peaks of the valence band (fig. 8.8(a)). Therefore, for a transition between a VB maximum and a CB minimum a momentum transfer to the crystal lattice is necessary. For a *direct* semiconductor, like for example GaAs, this is not the case: VB maximum and CB minimum have the same k value. The generation of e/h pairs and their recombination can hence take place without momentum transfer to the lattice (fig. 8.8(b)). For direct semiconductors light generation by e/h recombination is therefore more efficient than for indirect semiconductors which makes GaAs a preferred base material for light emitters like LEDs. For a more profound discussion of the subject the reader is referred to the literature, e.g. [894, 926, 929].

8.2.3 Intrinsic semiconductors

In this section the conduction properties of semiconductors without external impurity addition (doping) are described, mostly with silicon as the example. Doped semiconductors are discussed in section 8.2.4.

As described in the previous section, electrons from the valence band can be lifted by thermal excitation into the conduction band, leaving holes in the valence band that can be filled by another VB electron. Both, electrons and holes, can move freely in

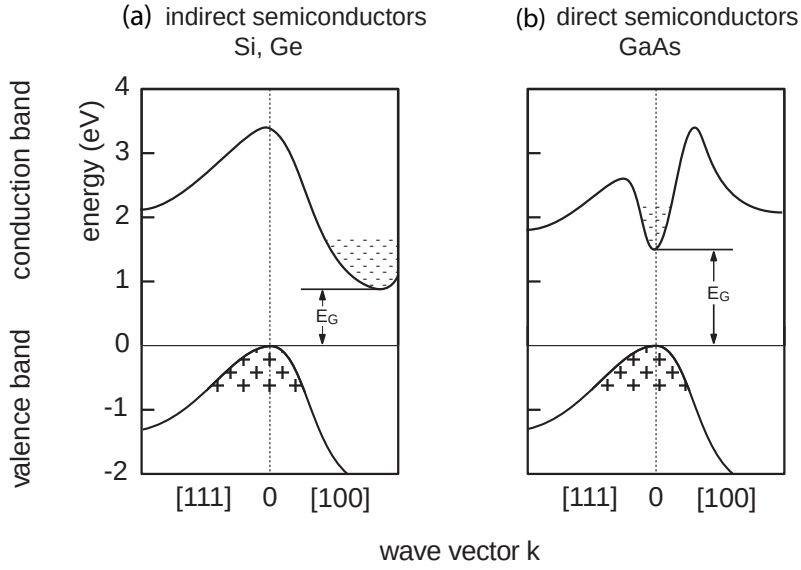


Fig. 8.8 Band structures of germanium, silicon (a) and gallium arsenide (b), respectively. The indices [k l m] designate directions within the crystal defined as $k\vec{a}_1 + l\vec{a}_2 + m\vec{a}_3$, where the $\vec{a}_i$ are base vectors of the (cubic) lattice cells. Each figure shows to the left or right, respectively, the $E(k)$ relation (in one dimension) for wave propagation in two different k directions as indicated by the indices. The origin is arbitrarily placed at the $E(k)$ position of the valence band maximum. Ge and Si are indirect semiconductors, GaAs is a direct semiconductor. For conduction phenomena in semiconductors predominantly the states near the band edges are important.

their respective band and can be treated as free particles which differ only by their 'effective' masses.

In a crystal the wave vector $\vec{k}$ of an electron or hole wave is not proportional to the electron momentum (not a momentum eigenstate of the Schrödinger equation). Regardless, $\hbar\vec{k}$ is a generalised momentum defined in a periodic potential and is called the *crystal momentum*. The *effective mass* emerges as a consequence of the dependence of the electron or hole energy E on the crystal momentum, that is $E = E(\vec{k})$. This dependence reflects the fact that electron or hole move in a periodic lattice potential. Considering first only a one-dimensional movement of an electron wave packet, its group velocity is:

$$v = \frac{d\omega}{dk} = \frac{1}{\hbar}\frac{dE}{dk}$$

and its acceleration

$$a = \dot{v} = \frac{1}{\hbar}\left(\frac{d^2E}{dk^2}\right)\frac{dk}{dt}\,. \tag{8.3}$$

In an electric field $\mathcal{E}$ the moving electron absorbs the energy

$$dE = e\mathcal{E}v dt = \frac{e\mathcal{E}}{\hbar}\frac{dE}{dk}dt\,. \tag{8.4}$$

With $dE = \frac{dE}{dk}dk$ and using (8.3) and (8.4) the acceleration a becomes

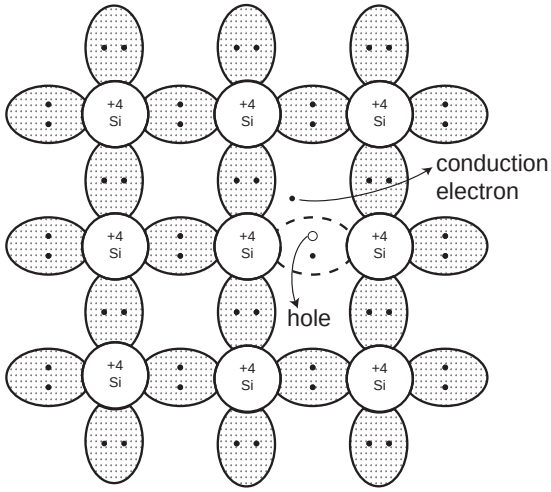


Fig. 8.9 Intrinsic conduction. An electron and a hole are generated thermally. They can move freely in the solid (adapted from [926]).

$$a \;=\; \underbrace{\frac{1}{\hbar^2}\frac{d^2E}{dk^2}}_{1/m^*} e\mathcal{E}\;. \tag{8.5}$$

This defines the effective mass m^*

$$\frac{1}{m^*_{ij}} = \frac{1}{\hbar^2}\frac{\partial^2 E(k)}{\partial k_i \partial k_j}\,, \tag{8.6}$$

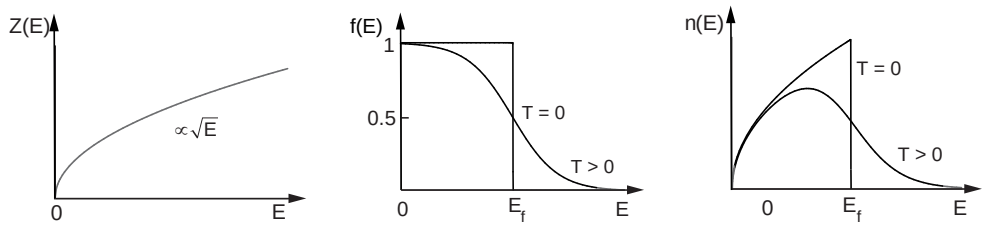
which depends on the band curvature (d^2E/dk^2) in momentum space. In full generality m^* has a tensor structure and depends on the direction of the electron or hole movement in the crystal. For germanium and silicon, two different values exist for the effective mass of electrons depending on their direction; for GaAs only one value exists. The values of m^* for electrons in Si at 4 K, longitudinally and transverse to the [100] crystal axis, are $m_n^{*,\mathrm{long}} = 0.92\,m_e$ and $m_n^{*,\mathrm{trans}} = 0.19\,m_e$ [491] where $m_e = 511\,\mathrm{keV/c^2}$ is the electron rest mass. Likewise one obtains several effective masses for holes, speaking of 'heavy' ($m_p^{*,\mathrm{heavy}} = 0.53\,m_e$) and 'light' ($m_p^{*,\mathrm{light}} = 0.16\,m_e$) holes [491].

8.2.3.1 Charge carrier concentration in thermal equilibrium

In order to compute the charge carrier concentration $n(E)$ in thermal equilibrium, one must know the density of states $Z(E)$ and their occupation probability $f(E)$:

$$n(E)\,dE \;=\; Z(E)\,f(E)\,dE\,. \tag{8.7}$$

The density of allowed states $Z(E)$ specifies the number of states (per volume and energy) that can be occupied by electrons or holes, respectively. For this we first contemplate how many momentum states there are between two shells of a sphere with radii p and $p + dp$ in momentum space. The included momentum volume is $4\pi p^2\,dp$. A phase-space unit cell that can be occupied by exactly one space-momentum state, has the volume h^3. Since electrons can be in two different spin states one obtains the momentum density, normalised to the unit space volume $V = 1$:



(a) Density of states. (b) Fermi–Dirac distribution. (c) Electron density.

Fig. 8.10 (a) Density of states, (b) Fermi distribution and (c) population density (for electrons) in solids.

$$Z(p)\,dp = 2\frac{4\pi p^2\,dp}{h^3}\,.$$

Employing the non-relativistic relation $E = p^2/(2m)$ we obtain

$$\frac{dE}{dp} = \frac{p}{m^*} \;\Rightarrow\; dp = \frac{m^*}{p}\,dE = \frac{m^*}{\sqrt{2m^*E}}\,dE$$

$$\Rightarrow\; 4\pi p^2\,dp = 4\pi(2m^*E)\frac{m^*}{\sqrt{2m^*E}}\,dE$$

and hence the density of energy states as

$$Z(E)\,dE = 4\pi\left(\frac{2m^*}{h^2}\right)^{3/2}\sqrt{E}\,dE\,. \tag{8.8}$$

The number of states per volume and energy unit that can be occupied by electrons, increases proportional to the square root of the energy of the state (see fig. 8.10(a)): $Z(E) \propto \sqrt{E}$.

Fermi–Dirac distribution. Being spin-1/2 particles the solid state electrons (and also the holes) are distributed over the quantum energy states of the band structure according to the Fermi–Dirac distribution (fig. 8.10(b)):

$$f_n(E) = \frac{1}{\exp\left(\frac{E-E_f}{kT}\right)+1} \tag{8.9}$$

with $k =$ Boltzmann constant, $T =$ temperature and $E_f =$ Fermi energy. For holes in the valence band the occupation probability is defined analogously:

$$f_p(E) = \frac{1}{\exp\left(\frac{E_f-E}{kT}\right)+1} \tag{8.10}$$

with the subscripts n and p denoting the distribution of negative and positive charge carriers (electrons or holes), respectively.

In a rigorous definition, the *Fermi energy* or *Fermi level* E_f is the chemical potential of electrons in a solid. For our purpose a definition employing the occupation probability is sufficient. For $T = 0\,\mathrm{K}$ the electrons completely occupy the energy bands

up to the Fermi level. States above E_f are not occupied. For $T > 0\,\mathrm{K}$, E_f is defined by the energy having an occupation probability of 50 % (fig. 8.10(b)). For intrinsic semiconductors, the *intrinsic Fermi level* E_f lies approximately (at $T = 0\,\mathrm{K}$ exactly) in the middle of the band gap.

The product of the density of states and the occupation probability yields the charge carrier density; see fig. 8.10(c).

In situations as for semiconductors where a 'forbidden' region without energy levels exists, we obtain for the electron and hole density of states in conduction and valence bands, respectively (index C denotes conduction band, V valence band):

$$Z(E)dE = 4\pi \left(\frac{2m_n^*}{h^2}\right)^{3/2} \sqrt{E - E_C} \quad \Theta(E - E_C)\, dE\,, \tag{8.11}$$

$$Z(E)dE = 4\pi \left(\frac{2m_p^*}{h^2}\right)^{3/2} \sqrt{E_V - E} \quad \Theta(E_V - E)\, dE\,. \tag{8.12}$$

The step function Θ accounts for the abrupt transitions at the band edges to the gap. By multiplication of $Z(E)\,dE$ with the respective Fermi–Dirac distribution (8.9) or (8.10) one obtains $n(E)\,dE$ and $p(E)\,dE$ including the effect of the band gap, as shown in fig. 8.11.

We can approximate the expression for $f_n(E)$ assuming that E_f is approximately in the middle of the band gap and hence $(E - E_f) \gg kT$:

$$f_n(E) = \frac{1}{\mathrm{e}^{\frac{E-E_f}{kT}} + 1} \approx \mathrm{e}^{-\frac{E-E_f}{kT}} \qquad \text{(Boltzmann distribution function)} \tag{8.13}$$

In addition, the integration over the conduction band energies can be performed up to ∞, a valid approximation since f_n decreases exponentially. Considering that $f_n(E) + f_p(E) = 1$ one can adopt the same approximations for holes.

Integration of the carrier densities over the respective band energies yields the number densities n for electrons and p for holes in their respective bands:

$$n = 2\left(\frac{m_n^*\, kT}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(-\frac{E_C - E_f}{kT}\right) = N_C \exp\left(-\frac{E_C - E_f}{kT}\right), \tag{8.14}$$

$$p = 2\left(\frac{m_p^*\, kT}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(-\frac{E_f - E_V}{kT}\right) = N_V \exp\left(-\frac{E_f - E_V}{kT}\right), \tag{8.15}$$

where N_C and N_A define the effective densities of states in conduction and valence bands:

$$\begin{aligned} N_C &= 2\left(\frac{m_n^*\, kT}{2\pi\hbar^2}\right)^{\frac{3}{2}} \approx 3.05 \times 10^{19}\,\mathrm{cm}^{-3}\,, \\ N_V &= 2\left(\frac{m_p^*\, kT}{2\pi\hbar^2}\right)^{\frac{3}{2}} \approx 2.55 \times 10^{19}\,\mathrm{cm}^{-3}\,. \end{aligned} \tag{8.16}$$

For the numerical values in (8.16) silicon at 300 K has been assumed. The effective masses thereby are computed averages over the various effective masses that occur for electrons and holes as explained previously (page 266) depending on the number and shapes of the contributing band minima and maxima. The numbers in (8.16)

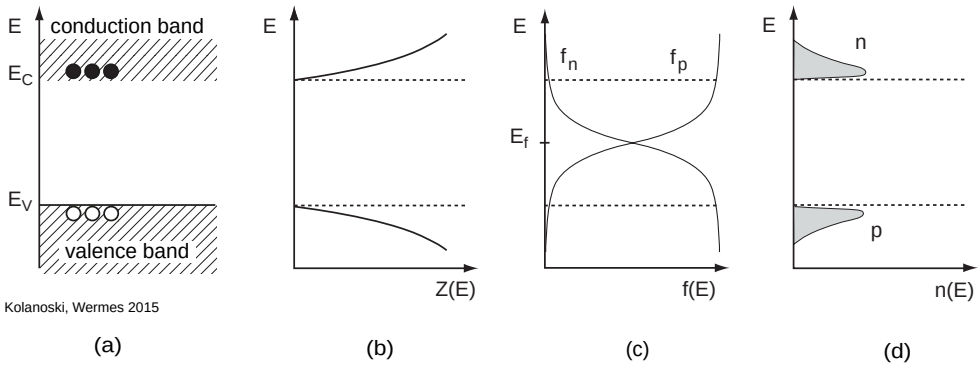


Fig. 8.11 Energy-band model for undoped silicon (schematic): (a) band diagram, (b) density of states, (c) occupation probabilities, (d) charge carrier densities in conduction and valence band, respectively.

are obtained with $m_n^* = 1.14\, m_e$ and $m_p^* = 1.01\ m_e$ which are averages from several computations yielding similar results (see [654], page 94). Figure 8.11 displays the discussed situation for intrinsic semiconductors.

The mass–action law, known from chemistry (see e.g. [990]), holds. In thermal equilibrium, generation and recombination processes of electrons in the conduction band and of holes in the valence band are in balance, that is, $n = p = n_i$ or

$$n \cdot p = n_i^2 = const\,. \tag{8.17}$$

Equation (8.17) also holds when one charge carrier type dominates the other, caused for example by doping (see section 8.2.4), since in thermal equilibrium the generation-recombination rates are proportional to the product of the carrier densities.

With (8.14) and (8.15) we obtain from (8.17):

$$\begin{aligned} n_i^2 &= n\,p = N_C\,N_V\,\exp\left(-\frac{E_C - E_V}{kT}\right) \\ &= N_C\,N_V\,\exp\left(-\frac{E_G}{kT}\right)\,. \end{aligned} \tag{8.18}$$

Note the strong, approximately exponential temperature dependence of the intrinsic carrier concentration:

$$n_i \propto T^{3/2}\exp(-E_G/2kT), \tag{8.19}$$

even somewhat increased by the temperature dependence of $N_C \times N_V$ ($\propto T^3$). Using numerical values from (8.16) at 300 K one obtains for silicon an *intrinsic charge carrier density*[2] of (see also table 8.2):

$$n_i \approx 1.01 \times 10^{10}\,\mathrm{cm}^{-3}\,. \tag{8.20}$$

n_i depends on the width of the band gap $E_G = E_C - E_V$ and on the temperature, but not on the Fermi energy E_f.

[2] In the literature, for a long time $n_i = 1.45 \times 10^{10}\,\mathrm{cm}^{-3}$ was listed, a value which according to newer calculations [474, 906, 654] is no longer tenable, in agreement with measurements of the resistivity of intrinsic silicon [907].

Since in an intrinsic semiconductor the concentration of holes is equal to the concentration of electrons ($n = p$), we have with (8.14) and (8.15):

$$(m_n^*)^{3/2} \exp\left(-\frac{E_C - E_f}{kT}\right) = (m_p^*)^{3/2} \exp\left(-\frac{E_f - E_V}{kT}\right)$$

$$\Rightarrow E_f = \frac{E_C + E_V}{2} + \frac{3kT}{4} \ln\left(\frac{m_p^*}{m_n^*}\right). \tag{8.21}$$

For $T = 0\,\mathrm{K}$ the Fermi energy hence is exactly central in the band gap. A small deviation from the centre occurs for finite temperature due to the different effective masses of electrons and holes.

Conductivity in intrinsic silicon. The conductivity of intrinsic silicon is given by

$$\sigma_i = n_i\, e\, (\mu_e + \mu_h) \simeq 2.8 \times 10^{-4}\, (\Omega\,\mathrm{m})^{-1} \tag{8.22}$$

where the intrinsic carrier concentration $n_i \approx 1.01 \times 10^{10}/\mathrm{cm}^3$ and the mobility values $\mu_{e,h}$ given in table 8.2 have been used. For comparison, the conductivity of copper is $\sigma_{\mathrm{Cu}} = 10^8\, (\Omega\,\mathrm{m})^{-1}$.

8.2.4 Doping: extrinsic semiconductors

By introducing impurity atoms in a semiconductor one can selectively change its conduction properties. For example, by placing pentavalent elements (P, As, Sb), called *donors*, in a tetravalent semiconductor (Si, Ge) an excess of conduction electrons occurs compared to the holes (n doping, fig. 8.12(a)). Doping with trivalent atoms (B, Al, Ga), called *acceptors*, leads to an excess of holes (p doping, fig. 8.12(b)). Doped semiconductors are also called extrinsic semiconductors in contrast to undoped intrinsic semiconductors. As the impurities are neutral atoms the doped semiconductor remains electrically neutral. Merely, the available number of charge carrier types (electrons or holes) increases or decreases, respectively, and thus changes the conduction properties of the semiconductor.

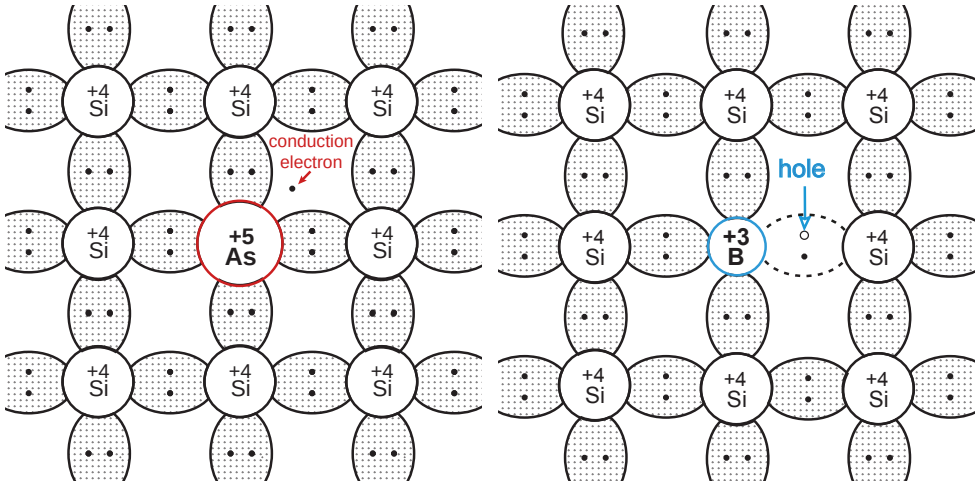
The mass-action law $n_i^2 = np$, introduced in the previous section (eq. (8.17)), also holds for doped semiconductors. Density increase of one charge carrier type decreases the other, since an excess for example of electrons facilitates recombination and thus hole removal, to a level below the intrinsic density.

Space-charge neutrality. If all impurities are uniformly distributed in the semiconductor the net charge density for every volume element of the semiconductor is zero. This means that adding up all positive and negative charges, assuming that donors and acceptors are fully ionised, one obtains for the charge density

$$\rho = e\,(n - p + N_D - N_A) = 0 \quad \Rightarrow \quad n - p = N_D - N_A\,. \tag{8.23}$$

Combining (8.23) with the equilibrium relationship (8.17) we obtain the *majority carrier* densities, the electron density in an n-type semiconductor ($N_D \gg n_i \gg N_A$) and the hole density in a p-type semiconductor ($N_A \gg n_i \gg N_D$) as:

$$n_n = \frac{1}{2}\left(N_D - N_A + \sqrt{4n_i^2 + (N_D - N_A)^2}\right) \approx |N_D - N_A| \approx N_D\,, \tag{8.24}$$



(a) n doped (b) p doped

Fig. 8.12 Schematic bonding representation in n- and in p-doped semiconductors (adapted from [926]).

$$p_p = \frac{1}{2}\left(N_A - N_D + \sqrt{4n_i^2 + (N_D - N_A)^2}\right) \approx |N_A - N_D| \approx N_A\,,$$

and for the *minority carriers* (holes and electrons, respectively):

$$p_n \approx \frac{n_i^2}{N_D - N_A} \approx \frac{n_i^2}{N_D}\,, \qquad n_p \approx \frac{n_i^2}{N_A - N_D} \approx \frac{n_i^2}{N_A}\,. \tag{8.25}$$

A numerical example illustrates typical relationships in silicon. We assume that a pure silicon crystal is doped with $N_D = 10^{16}\,\mathrm{cm}^{-3}$ As atoms. The respective carrier densities then are

$$n \;\approx\; N_D \approx 10^{16}\,\mathrm{cm}^{-3}\,, \qquad p \;\approx\; \frac{n_i^2}{N_D} \approx \frac{10^{20}}{10^{16}} = 10^4\,\mathrm{cm}^{-3}\,. \tag{8.26}$$

Energy bands in doped semiconductors. In the energy-band model the level of the fifth valence electron (donor level, E_D) is located close below the conduction band (fig. 8.13(a)). This creates new states to be accounted for by the density of states distribution (fig. 8.13(b)). The difference in energies between donor level and bottom edge of the conduction band are in the order of some 10^{-2} eV, such that at room temperature most donors are ionised. This raises the Fermi level from its intrinsic value E_f to a value E_F for the case with doping (fig. 8.13(c)). The population densities change in favour of the electrons in the conduction band (fig. 8.13(d)).

For n doping we obtain

$$\begin{aligned} n \approx N_D \;&=\; N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_C \exp\left(-\frac{E_C - E_f}{kT}\right) \exp\left(-\frac{E_f - E_F}{kT}\right) \\ &=\; n_i \exp\left(\frac{E_F - E_f}{kT}\right)\,, \end{aligned} \tag{8.27}$$

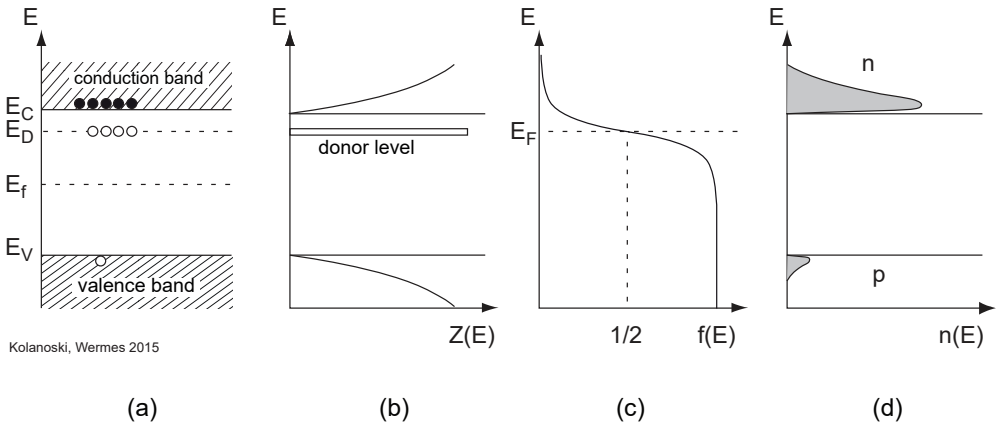


Fig. 8.13 Band-model situation in silicon with n doping (schematic): (a) energy-band structure with additional donor level, (b) densities of states in conduction and valence band, respectively, as well as donor level(s), (c) Fermi–Dirac occupation probability for electrons respecting the presence of the donor level, (d) population densities of electrons in the conduction band and holes in the valence band, respectively. For p doping the relationships reverse: E_D is replaced by E_A near the valence band, E_F moves below E_f, and the hole density becomes larger than the electron density.

with n_i given by (8.14). For p doping we have:

$$p \approx N_A \;=\; n_i \exp\left(\frac{E_f - E_F}{kT}\right) . \tag{8.28}$$

The distance between the Fermi levels E_F and E_f hence is a measure for the change from intrinsic to extrinsic material.

With the numerical values of example (8.26) we obtain for the relevant energy distances (for $T = 300\,\mathrm{K}$):

$$\begin{aligned} E_F - E_f &= kT \ln \frac{N_D}{n_i} = 0.348\,\mathrm{eV}\,, \\ E_C - E_F &= kT \ln \frac{N_C}{N_D} = 0.209\,\mathrm{eV}\,, \\ E_F - E_V &= 1.12\,\mathrm{eV} - 0.209\,\mathrm{eV} = 0.911\,\mathrm{eV}\,. \end{aligned} \tag{8.29}$$

For p doping the corresponding replacements in (8.29) must be made: $N_D \to N_A$, N_C, $E_C \to E_V$, and the signs of the differences are reversed. If both, donors and acceptors are present and/or for high doping concentrations the relationships are more complicated and the approximations made in (8.14) and (8.15) are no longer valid. Depending on donor/acceptor concentrations the Fermi level E_F can also be located above the donor level E_D (or below the acceptor level E_A). It can even lie inside the conduction or valence bands. At $T = 0$ and n doping, E_F is by definition between E_D and E_C.

Other than by selective doping also other impurities cause additional energy levels that can act as donors or acceptors, depending on their location within the band gap. The probability that an impurity atom becomes a doping state is not the same for

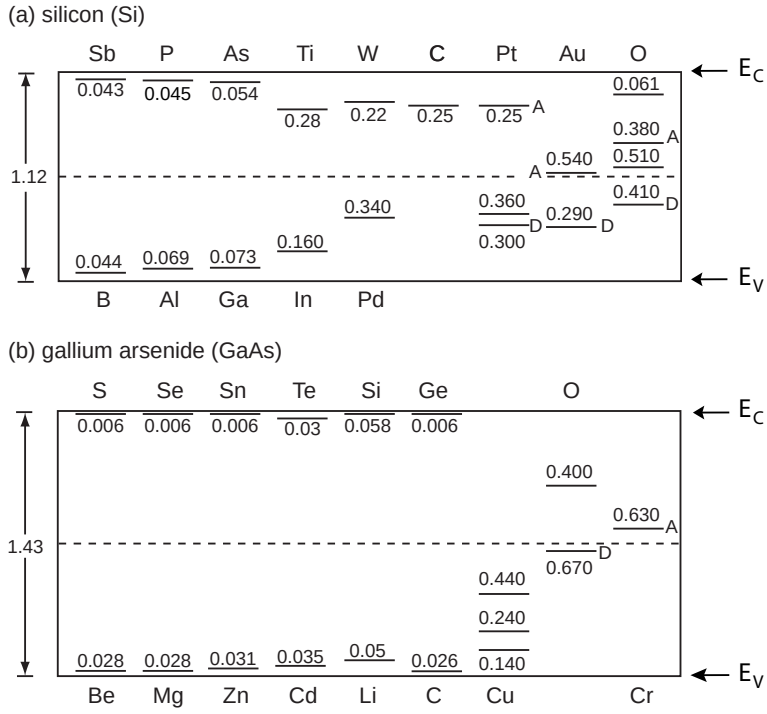


Fig. 8.14 Ionisation energies (energies needed for transition into CB, respectively VB) of different impurities in Si (a) and in GaAs (b). All data in eV. A,D additions indicate the acceptor/donor nature for levels where this is not obvious from their position in the bandgap (adapted from [929]).

donors and acceptors (see [929]). Levels in the middle of the gap (deep levels) can act as generation-recombination centres increasing the so-called detector *leakage current* which corresponds to the current of a reverse-biased diode (see also section 8.12.1). Figure 8.14 shows the energetic positions inside the band gap of different elements as impurities in silicon and in gallium arsenide. The two states called EL2 near the middle of the gap in GaAs are due to situations when a Ga atom takes the place of an As atom and vice versa. Since they lie near the gap centre they play an important role in the operation of GaAs detectors (see section 8.13.2). Figure 8.15 shows the concentration and temperature dependence of the positions of the extrinsic Fermi level E_F as the difference to the intrinsic level: $E_F - E_f$. With decreasing concentration of impurities the difference shrinks, also with increasing temperature. Especially the former plays an important role when choosing doping profiles.

8.3 Junctions

A semiconductor detector is a special case of a diode which is known as an electronic circuit element. To understand this type of detector, in particular the physics of boundaries and interfaces is important, in fact not only the semiconductor–semiconductor interfaces but also the metal–semiconductor- and the metal–insulator–semiconductor interfaces. For the large majority of semiconductor detectors in particle physics the

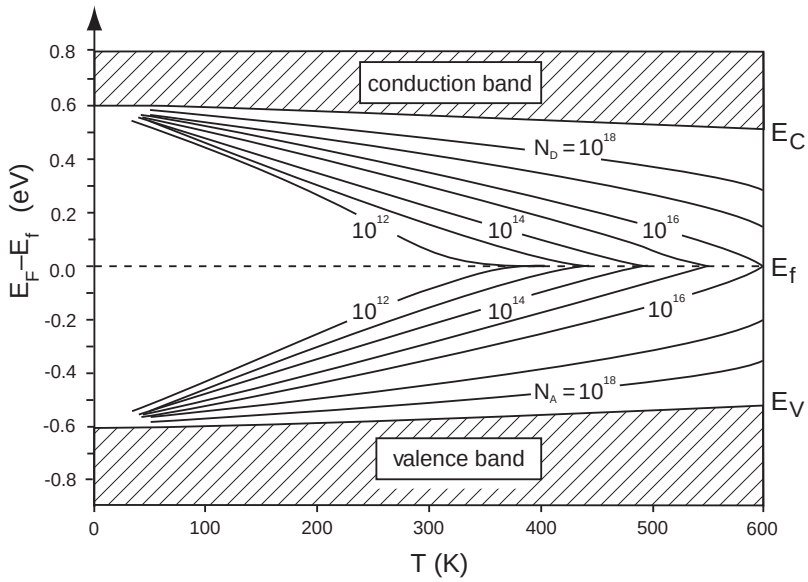


Fig. 8.15 Dependence of the difference between intrinsic (E_f) and extrinsic (E_F) Fermi energies on temperature and on impurity concentration (adapted from [486]).

base material is silicon. Their performance is predominantly determined by the boundary properties between p- and n-doped silicon (pn boundary). Metal–semiconductor boundaries occur at connections to the outside (metal contacts). Metal–oxide–semiconductor interfaces (MOS), which have revolutionised microelectronics (MOS transistors), are important for special detector types and designs (for example CCDs) and are indispensable for microelectronics implementations in the readout of such detectors by so-called *application specific integrated circuits* (ASICs, see section 17.6). Unavoidably, however, MOS layers are also formed at contacts, since silicon easily oxidises in air. In the following sections we discuss in some more detail semiconductor–semiconductor, metal–semiconductor and metal–oxide–semiconductor interfaces.

8.3.1 The pn junction as a detector

When bringing p- and n-doped semiconductor materials in contact a so-called pn boundary is formed. In the energy-band model this is illustrated in a simplified way in fig. 8.16, assuming an abrupt transition. In the p-doped part of the crystal, holes are the dominant charge carriers (majority carriers) and the Fermi level is near the valence band edge. In the n-doped part the relationships are reversed: the majority carriers here are electrons and the Fermi level is near the conduction band edge. The strong concentration gradients of the two charge carrier types at the boundary lead to a diffusion current I_{diff} (see section 4.7.2), that is, electrons of the n-doped part diffuse into the p-doped part of the crystal and holes from the p-doped part diffuse into the n-doped part. At the boundary, recombination of both carrier types occurs leading to a zone which is free of charge carriers, called the *depletion zone*.

While there are no free charge carriers in the depletion region, the atomic cores remain ionised after e–h recombination has taken place and the region is no longer neutral but features a *space charge*. The p layer has a negative, the n layer a positive

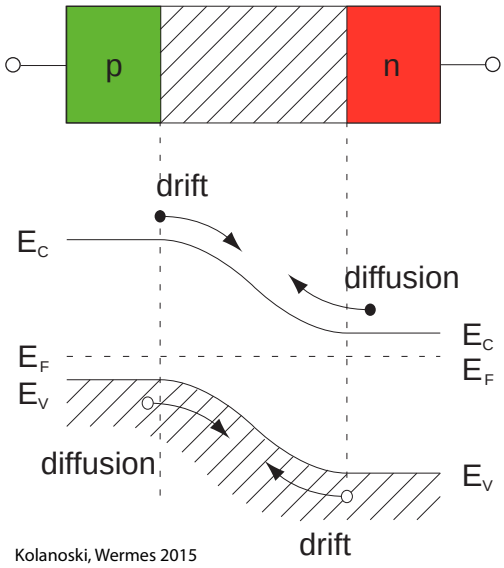


Fig. 8.16 Drift and diffusion currents for a pn boundary.

space-charge density $\rho(x)$. Hence the depleted region is also called *space-charge region*. Due to the opposite space charge in the p and in the n layer, an intrinsic electrical field is formed causing a drift current I_{drift} in the opposite direction to the diffusion current I_{diff}. The created electrical potential causes the energy levels to bend across the boundary, whereas the Fermi level, being a thermodynamic quantity defined via an occupation probability, remains constant (see discussion on page 278).

A one-dimensional picture is sufficient to discuss the main features. Figure 8.17 shows the shapes of charge density, electric field and potential for an idealised abrupt transition. Due to the (assumed) constant space charge regions the electric field rises linearly with a maximum $\vec{\mathcal{E}}_{max}$ right at the boundary between the p and n layers.

The resulting *built-in voltage* or *diffusion potential* V_{bi} for silicon has a value of about 0.6 V, for GaAs about 1.2 V. Without external voltage drift current and diffusion current are in equilibrium.

In equilibrium the space-charge region only depends on the doping of the semiconductor parts. The charge density $\rho(x)$ is given by:

$$\rho(x) = \begin{cases} -eN_A & \text{for} \quad -x_p < x < 0\,, \\ +eN_D & \text{for} \quad 0 < x < x_n\,. \end{cases} \tag{8.30}$$

From this the electric field $\mathcal{E}(x)$ and the potential difference V_{bi} can be calculated. Since the semiconductor is electrically neutral in the regions outside the space-charge zone, such that in these regions there is no field, we demand as boundary conditions $\mathcal{E}(-x_p) = \mathcal{E}(+x_n) = 0$, where x_p and x_n are the depths of the space-charge zones in the p-doped and n-doped parts of the semiconductor, respectively.

To guarantee the overall neutrality of the semiconductor the number of charge carriers in both parts of the space-charge region must be equal. The *neutrality condition* holds:

$$N_A\, x_p = N_D\, x_n\,. \tag{8.31}$$

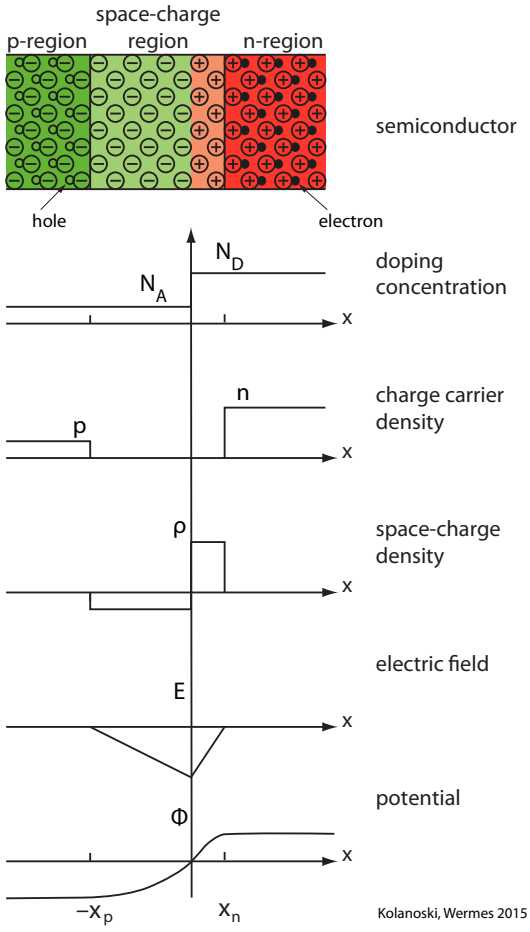


Fig. 8.17 Doping and space charge densities, electric field strength and electric potential at an abrupt pn transition.

The space-charge region extends farther into the more weakly doped part of the semiconductor than into the more strongly doped part.

Employing the one-dimensional Maxwell equation for the $\mathcal{E}$-field and with (8.30) as well as (8.31) we obtain with the boundary conditions $\mathcal{E}(-x_p) = \mathcal{E}(+x_n) = 0$:

$$\frac{d\mathcal{E}}{dx} = \frac{1}{\epsilon\epsilon_0}\rho(x) \tag{8.32}$$

$$\Rightarrow \mathcal{E}(x) = \begin{cases} \frac{-eN_A}{\epsilon\epsilon_0}(x + x_p)\,, & -x_p < x < 0\,, \\ \frac{+eN_D}{\epsilon\epsilon_0}(x - x_n)\,, & 0 < x < x_n\,. \end{cases} \tag{8.33}$$

The field maximum is at $x = 0$:

$$\mathcal{E}_{\max} = -\frac{eN_A}{\epsilon\epsilon_0}x_p = \frac{eN_D}{\epsilon\epsilon_0}(-x_n)\,. \tag{8.34}$$

The voltage drop over the depletion zone, the built-in voltage V_{bi}, is hence directly obtained from the carrier densities as the difference of the potentials in the n and p regions outside of the space-charge region: $V_{bi} = \phi_p - \phi_n$, where $\phi_p = \phi(-x_p)$ and

$\phi_n = \phi(x_n)$ are the potentials at the boundaries of the space-charge region, which at the same time are the integration limits for the computation of the potential:

$$\phi(x) = \begin{cases} \phi_p + \frac{eN_A}{2\epsilon\epsilon_0}(x + x_p)^2 , & -x_p < x < 0 , \\ \phi_n - \frac{eN_D}{2\epsilon\epsilon_0}(x - x_n)^2 , & 0 < x < x_n . \end{cases} \tag{8.35}$$

The potential is quadratic over the range of the space-charge zone and constant outside. The built-in voltage V_{bi} is related to the difference between the intrinsic (E_f) and extrinsic (E_F) Fermi levels in the respective n and p regions, because in equilibrium E_F is constant over the junction, while E_f follows the band bending (see fig. 8.16):

$$E_f - E_F^p = -e\phi_p = kT \ln \frac{N_A}{n_i} \qquad \text{and} \qquad E_F^n - E_f = e\phi_n = kT \ln \frac{N_D}{n_i} \tag{8.36}$$

and can hence be expressed as:

$$V_{bi} = \phi_n - \phi_p = \frac{kT}{e} \ln \frac{N_A N_D}{n_i^2} \overset{Si}{\approx} 0.4 - 0.8\,\mathrm{V} , \tag{8.37}$$

where the range is typical for silicon doping concentrations.

Depending on the depletion depths x_p and x_n the voltage V_{bi} can be writen as:

$$\begin{aligned} V_{bi} &= -\int_{-x_p}^{x_n} \mathcal{E}(x)dx = \frac{e}{2\epsilon\epsilon_0}(N_A x_p^2 + N_D x_n^2) \\ &= \frac{e}{2\epsilon\epsilon_0} x_p^2 \frac{N_A}{N_D}(N_A + N_D) . \end{aligned} \tag{8.38}$$

In the last step of the above derivation the neutrality condition (8.31) has been used whereby x_p can be replaced by x_n. From (8.38) one obtains the depths of both depletion zones into their respective parts of the boundary region:

$$x_p = \sqrt{\frac{2\epsilon\epsilon_0}{e} V_{bi} \frac{N_D}{N_A\,(N_D + N_A)}} , \tag{8.39}$$

$$x_n = \sqrt{\frac{2\epsilon\epsilon_0}{e} V_{bi} \frac{N_A}{N_D\,(N_D + N_A)}} , \tag{8.40}$$

$$\frac{x_p}{x_n} = \sqrt{\frac{N_D}{N_A\,(N_D + N_A)} \frac{N_D\,(N_D + N_A)}{N_A}} = \frac{N_D}{N_A} . \tag{8.41}$$

which is (8.31). The carrier-free depletion zone is larger in the more weakly doped part of the boundary.

Typical for silicon detectors are doping concentrations and related depletion depths in the order of:

$$\begin{aligned} N_A &= 10^{19}\,\mathrm{cm}^{-3} \qquad N_D = 2.3 \times 10^{12}\,\mathrm{cm}^{-3} \\ \Rightarrow x_p &\approx \sqrt{\frac{2\epsilon\epsilon_0}{e} V_{bi} \frac{N_D}{N_A^2}} = 4 \times 10^{-6}\,\mu\mathrm{m} \\ \Rightarrow x_n &= x_p \frac{N_A}{N_D} \approx 20\,\mu\mathrm{m} . \end{aligned}$$

One can neglect the contribution of x_p to the depletion region, thus obtaining the depletion-zone width d:

$$N_A \gg N_D \;\Rightarrow\; x_p \ll x_n\,, \qquad \Rightarrow \qquad d \approx x_n \approx \sqrt{\frac{2\epsilon\epsilon_0}{e} V_{bi} \frac{1}{N_D}}\,. \tag{8.42}$$

Fermi level over a boundary. An important principle which can be generally employed when considering energy levels in boundary structures, is that in thermal equilibrium the Fermi level is constant over the region of the boundary. This means here particularly the situation without externally applied voltage.

The proof is delivered by considering the current densities in semiconductors following eq. (4.131) on page 124. Thermal equilibrium means that there exists no net current flow, that is, for example for holes (for electrons analogously):

$$j_p = j_{p,\text{drift}} + j_{p,\text{diff}} = +e\mu_p p\,|\vec{\mathcal{E}}| - eD_p \frac{dp}{dx} = 0\,, \tag{8.43}$$

where p, μ_p and D_p are hole density, mobility and diffusion coefficient, respectively, and $\vec{\mathcal{E}}$ is the electric field. With (8.28) one obtains for the derivative:

$$\frac{dp}{dx} \;=\; \frac{p}{kT}\left(\frac{dE_f}{dx} - \frac{dE_F}{dx}\right)\,. \tag{8.44}$$

The electric field $|\vec{\mathcal{E}}|$ in (8.43) is the potential gradient over the boundary or, equivalently, of the energy levels (divided by e) E_f, E_C, E_V which bend in the same way. Using E_f we can then write:

$$|\vec{\mathcal{E}}| = \frac{1}{e}\frac{dE_f}{dx}\,. \tag{8.45}$$

The hole density $p(x)$ is, according to (8.44), determined by the distance of the Fermi energy E_F from the intrinsic Fermi level E_f and changes over the region of the boundary. Using the Einstein relation (eq. (4.137) on page 125) we obtain from (8.43), (8.44) and (8.45):

$$j_p \;=\; \mu_p p\left(\frac{dE_f}{dx} - \frac{dE_f}{dx} + \frac{dE_F}{dx}\right) = 0 \;\Rightarrow\; \frac{dE_F}{dx} = 0\,.$$

In thermal equilibrium the Fermi level E_F is constant over a boundary structure.

Carrier injection and continuity equation. The equations for the current densities, (4.131) in chapter 4 and (8.43) in the previous section, hold for stationary equilibrium conditions. To describe carrier injection or extraction, that is, generation or recombination of charge carriers, the continuity equations can be exploited. The change in carrier concentrations is the difference of the generation and recombination rates plus the net current which flows in and out of the considered region:

$$\begin{aligned} \frac{\partial n}{\partial t} &= G_n - R_n + \frac{1}{e}\vec{\nabla}\vec{j}_n\,, \\ \frac{\partial p}{\partial t} &= G_p - R_p + \frac{1}{e}\vec{\nabla}\vec{j}_p\,, \end{aligned} \tag{8.46}$$

where $G_{n,p}$ and $R_{n,p}$ are the generation and recombination rates in the volume, respectively, due to causes like external bias, energy deposition by particles or by light, or by generation-recombination centres in the band gap, created for example by lattice defects. For the one-dimensional case considered here the equations can explicitly be written as follows:

$$\begin{aligned} \frac{\partial n}{\partial t} &= G_n - R_n + n\mu_n \frac{\partial \mathcal{E}}{\partial x} + \mu_n \mathcal{E} \frac{\partial n}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2}\,, \\ \frac{\partial p}{\partial t} &= G_p - R_p + p\mu_p \frac{\partial \mathcal{E}}{\partial x} + \mu_p \mathcal{E} \frac{\partial p}{\partial x} + D_p \frac{\partial^2 p}{\partial x^2} \end{aligned} \tag{8.47}$$

with notations as in (8.43). A solution of these equations is usually found numerically by computer programs.[3]

8.3.2 The pn junction under external voltage

Application of an external voltage V_{ext} between the n and p sides of the junction changes the width of the depletion region depending on the size and polarity of the applied voltage (fig. 8.18). The system is no longer in thermal equilibrium and (8.17) changes to $np > n_i^2$ or $np < n_i^2$. The Fermi level is no longer constant over the junction.[4] If the applied voltage V_{ext} at the p side is positive relative to the n side ($V_{ext} > 0$, *forward bias*), the electrostatic potential over the depletion zone (V_{bi}) reduces by V_{ext} and the drift current decreases in comparison to the diffusion current. Relative to the equilibrium situation more electrons diffuse from the n side to the p side and more holes from the p side to the n side, where they become minority carriers (minority carrier injection). The depletion region becomes thinner and the energy-band bending is weaker. By applying *reverse bias* with V_{ext} having the same direction as V_{bi} ($V_{ext} < 0$, negative voltage at the p side or positive at the n side relative to the respective other side) the electrostatic potential is raised relative to the equilibrium state, thus counteracting the diffusion current which therefore becomes smaller. The depletion zone gets wider and the energy-band bending becomes stronger. In equations (8.39), (8.40) and (8.42) V_{bi} must be replaced by $V_{bi} - V_{ext}$ where V_{ext} is to be taken negative for reverse bias. Figure 8.19 shows the carrier concentrations over a pn junction, which we will calculate in the following for the cases of forward (fig. 8.19(a)) and and reverse bias (fig. 8.19(b)). In this we make use of the following assumptions [926]:

(1) the boundaries between the space-charge region and the external neutral regions are abrupt,
(2) the carrier densities at the boundaries of the space-charge region are determined by the electrostatic potential $V_{bi} - V_{ext}$,
(3) the injected minority carriers have a much smaller density than the majority carriers and the latter hence remain almost unchanged,
(4) no recombination or generation currents occur in the space-charge region and the electron and hole currents are therefore constant.

[3] An example for a computer program to solve the continuity equations (8.46) is the program TCAD [925].

[4] To treat non-equilibrium conditions with the same formalism as introduced for equilibrium conditions, quasi-Fermi levels can be introduced, which are different for electrons and holes. see e.g. [929].

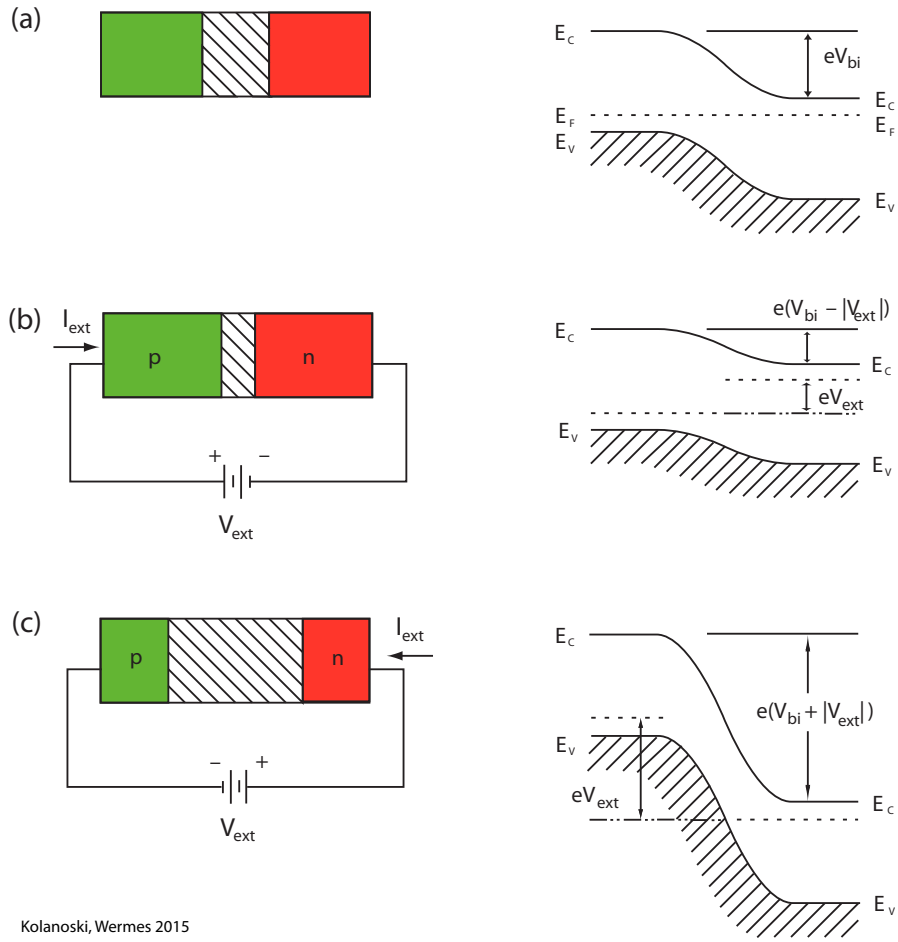


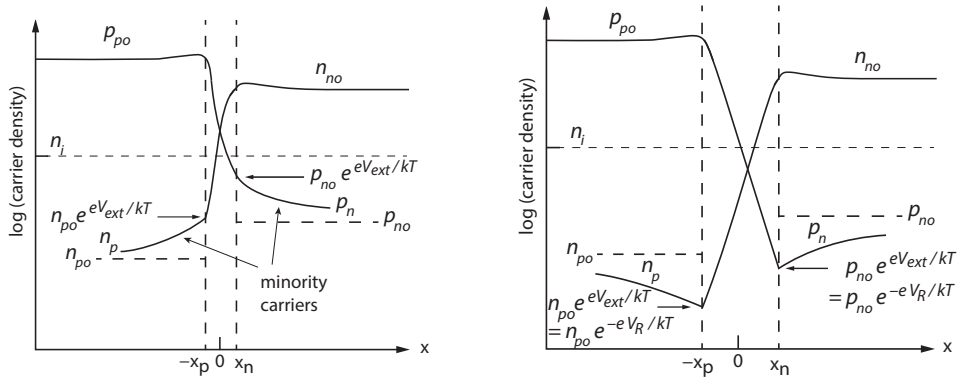
Fig. 8.18 Diode under different voltage conditions: (a) without external voltage, (b) with forward voltage applied, (c) with reverse voltage applied (operation condition for semiconductor detectors).

With this we can compute the carrier densities first at the boundaries of the space-charge region and then with the help of the Boltzmann transport equation (see chapter 4) also in the space-charge region [926].

In the neutral regions outside the space-charge region we can approximately assume equilibrium conditions. The concentrations of the majority carriers in these regions are given by the doping concentrations N_A and N_D, respectively. Following (8.37) V_{bi} for a pn junction can hence can be expressed as

$$V_{bi} = \frac{kT}{e}\ln\frac{N_A N_D}{n_i^2} \approx \frac{kT}{e}\ln\frac{p_{po}n_{no}}{n_i^2} = \frac{kT}{e}\ln\frac{n_{no}}{n_{po}} = \frac{kT}{e}\ln\frac{p_{po}}{p_{no}}\,, \tag{8.48}$$

where (8.17) has been used. Here $n_{no} \approx N_D$ and $p_{po} \approx N_A$ are the majority carrier concentrations (electrons in the n region, holes in the p region) and the index o indicates the relationships in thermal equilibrium (without external bias). The minority carrier densities in equilibrium correspondingly are n_{po} and p_{no}. Solving for the



(a) Forward bias. (b) Reverse bias.

Fig. 8.19 Charge carrier concentration over a pn junction structure for (a) forward and (b) reverse biasing (operation mode of semiconductor detectors) of a pn boundary in logarithmic representation (adapted following Shockley [892, 929]). V_R in (b) denotes the reverse voltage with reverse sign compared to V_{ext} for forward biasing.

majority carrier densities we obtain:

$$n_{no} = n_{po} \exp(eV_{bi}/kT)\,, \tag{8.49}$$
$$p_{po} = p_{no} \exp(eV_{bi}/kT)\,. \tag{8.50}$$

These equations establish the relationship between the carrier concentrations at the ends of the space-charge region and V_{bi}. With assumption (2) on page 279 we can also assume that the same relation holds, if the electrostatic potential is increased or decreased by V_{ext}. We then obtain the following non-equilibrium equations at the ends of the depletion region:

$$n_n\,(x = x_n) = n_p \exp(e(V_{bi} - V_{ext})/kT) \approx n_{no}\,, \tag{8.51}$$
$$p_p\,(x = -x_p) = p_n \exp(e(V_{bi} - V_{ext})/kT) \approx p_{po}\,. \tag{8.52}$$

Here assumption (4) of page 279 has been used which states that the minority carrier injection caused by V_{ext} leaves the majority carrier density almost unchanged. Insertion of (8.49) and solving for n_p yields for the electron concentration at the boundary of the space-charge region on the p side ($x = -x_p$) and analogously for the hole concentration p_n at the n-side end ($x = x_n$):

$$n_p(x = -x_p) = n_{po} \exp(eV_{ext}/kT)\,, \tag{8.53}$$
$$p_n(x = x_n) = p_{no} \exp(eV_{ext}/kT)\,. \tag{8.54}$$

As a consequence of the exponential dependences in (8.49)-(8.54), the majority carrier concentrations drop by many orders of magnitude over the space-charge region (note the logarithmic representation in fig. 8.19). In reverse bias (fig. 8.19(b)) their concentrations at the end of the space-charge region lie well below the minority carrier equilibrium concentrations and can be negligibly small in these regions for high reverse bias voltages.

The further progression of the densities into the neutral outer zones is determined by the carrier diffusion lengths (characteristic distance until recombination) L_n and

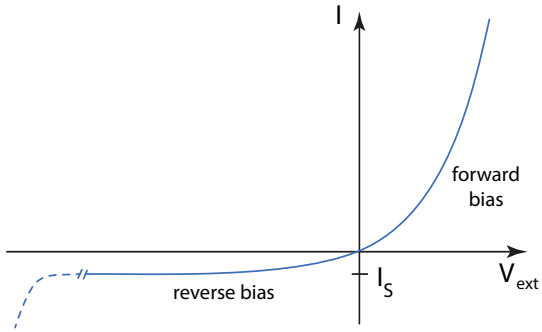


Fig. 8.20 Ideal pn-diode characteristic according to the Shockley equation (8.57). The dashed line characterises the breakthrough at high reverse bias voltages, which is not included in (8.57).

L_p and can be computed by means of the Boltzmann transport equations (see also chapter 4) [929, 872]:

$$n_p(x \leq -x_p) \;=\; n_{po} + n_{po}\,\left(\exp\left(eV_{ext}/kT\right) - 1\right)\,\exp\left(\frac{x + x_p}{L_n}\right)\,, \tag{8.55}$$

$$p_n(x \geq x_n) \;=\; p_{no} + p_{no}\,\left(\exp\left(eV_{ext}/kT\right) - 1\right)\,\exp\left(-\frac{x - x_n}{L_p}\right)\,. \tag{8.56}$$

Within the depletion zone the carrier concentration decreases from its majority carrier value p_p (n_n) on one side of the zone down to the minority carrier value p_n (n_p) on the other side. The functional progression of the concentration into the depletion zone is determined in very good approximation by the concentration values at the end of the depletion region, that is, at the transition to the neutral regions. This means they are determined by the doping concentrations and the external bias voltage. Doping concentrations typical for semiconductor detectors are given on page 277 in section 8.3.1.

Current-to-voltage (I–V) characteristic. Under the specified idealised assumptions no current is generated in the depletion zone; it originates exclusively from the outside regions without space charge [926]. The total current is constant over the semiconductor and is governed by the minority carrier densities and their diffusion characteristics, for example for holes given by the current density j_p in (8.43). The current shows an exponential dependence on the externally applied bias voltage, which has its origin in the exponential dependences of the carrier concentrations:

$$I = I_S\left(\mathrm{e}^{eV_{ext}/kT} - 1\right)\,, \tag{8.57}$$

with the (reverse) saturation current

$$I_S \;=\; eA\left(\frac{D_n n_{po}}{L_n} + \frac{D_p p_{no}}{L_p}\right) \;\approx\; eA\left(\frac{D_n n_i^2}{L_n N_A} + \frac{D_p n_i^2}{L_p N_D}\right)\,, \tag{8.58}$$

where A is the cross-sectional area and $D_{n,p}$ the diffusion constants for electrons and holes, respectively. The meanings of the other quantities have already been given above. The second term follows from (8.17) with $N_A \approx p_{po}$ and $N_D \approx n_{no}$. Equation (8.57) is known as Shockley equation [892] (characteristic of an ideal diode) and drawn in fig. 8.20.

Semiconductor detectors are operated in reverse bias (positive potential at the n-doped side of the diode) in order to obtain a depletion zone as large as possible. The depletion zone is the detection volume to detect particles (see section 8.4). The bias voltage V_{ext} then has the same sign as V_{bi}. Energy deposited in detector generates electron–hole pairs. In an undepleted[5] silicon detector these would quickly recombine and would not generate a signal on the electrodes of the detector. In a detector depleted of charge carriers, however, the probability for recombination is low. The generated e/h pairs drift in the E-field to the electrodes without losses and generate a current/charge signal (see chapter 5).

When applying an external voltage V_{ext} to silicon, (8.42) changes to

$$d_{n,p} \approx x_{n,p} \approx \sqrt{\frac{2\epsilon\epsilon_0}{e}\frac{1}{N_{D,A}}\left(V_{bi}+V_{ext}\right)} \;\approx\; 3.6\times10^{7}\sqrt{\frac{V_{ext}/\mathrm{V}}{N_{D,A}/\mathrm{cm}^{-3}}}\;\mu\mathrm{m}\,, \tag{8.59}$$

where $x_{n,p}$ and $N_{D,A}$ are depleted regions and doping concentrations for n or p type, respectively. Expressed in terms of the wafer resistivity $\rho \simeq (e\,\mu_{e,h}\,N_{D,A})^{-1}$ one obtains

$$d_n \approx 0.55\sqrt{\frac{\rho}{\Omega\,\mathrm{cm}}\frac{V_{ext}}{\mathrm{V}}}\;\mu\mathrm{m}\,, \qquad\qquad d_p \approx 0.32\sqrt{\frac{\rho}{\Omega\,\mathrm{cm}}\frac{V_{ext}}{\mathrm{V}}}\;\mu\mathrm{m}\,, \tag{8.60}$$

where the numerical difference for p and n doping is due to the different mobilities of electrons and holes. Usually the detector grade of silicon sensors or wafers is specified by quoting the resistivity.

In order to deplete the volume of a typical silicon detector with thickness $d = 300\,\mu\mathrm{m}$ and a given n-substrate doping of $2.3\times10^{12}\ \mathrm{cm}^{-3}$ ($\rho \approx 2\,\mathrm{k}\Omega\,\mathrm{cm}$) a bias voltage of $V_{ext} \approx 160\,\mathrm{V}$ must be applied.

Diode capacitance. Due to the charge carrier free zone an area diode can be regarded as a plate capacitor filled with a dielectric. The capacitance to the backside is

$$\frac{C}{A} = \frac{\epsilon\,\epsilon_0}{d}\,.$$

Here A is the area and d the width of the depletion zone. The relative permittivity ϵ for silicon has the value 11.9 (table 8.2). For a planar, 300 µm thick fully depleted Si area diode the capacitance is

$$\frac{C}{A} \simeq 35\,\frac{\mathrm{pF}}{\mathrm{cm}^2}\,,$$

assuming that the capacitance to the backside dominates all other capacitance contributions. Since according to (8.59) the carrier-free zone d depends on the external bias voltage, it is possible to determine the *full depletion voltage*, that is, the voltage necessary to deplete the entire sensor volume, by measuring the capacitance. Figure 8.21 shows the dependence of the capacitance C on the external bias voltage. When the capacitance assumes a constant value, upon bias increase the full depletion voltage is reached. The depletion zone cannot grow larger than the thickness of the detector.

[5] Detector materials like diamond and also GaAs and CdTe can be operated without employing depletion due to their much smaller intrinsic carrier densities (see table 8.2).

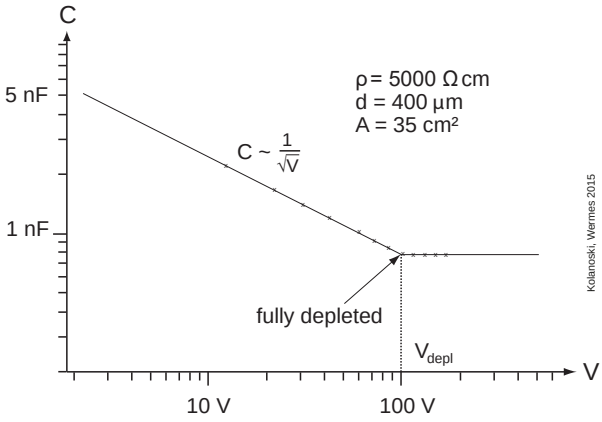


Fig. 8.21 Determination of the full depletion voltage from a measurement of the diode capacitance as a function of the applied voltage.

Leakage current. For a reverse biased diode or a semiconductor detector the so-called leakage current I_L is observed, which can have volume as well as surface contributions originating from different physical sources. It exists even in ideal diodes originating from the movement of minority carriers through the boundary (by diffusion), for example of electrons from the p region to the n region (I_S in (8.57) and (8.58)).

For detectors the most important cause of the (volume) leakage current is thermal generation of electron–hole pairs in the depletion region (see also section 8.12). Its size is strongly influenced by the presence of impurities that can act as generation/recombination centres in the band gap. The size of this generation current is proportional to the depleted volume $A \times d$ underneath an electrode, where A is the electrode area and d the depleted depth ($d \propto \sqrt{V_{ext}}$, see (8.59)):

$$I_L^{gen} = eAd\frac{n_i}{\tau_g}\,, \tag{8.61}$$

where τ_g denotes the charge carrier generation lifetime, defined as the time constant needed for the generated charge carriers to return to the equilibrium situation. Because n_i is strongly temperature dependent (8.19) with $n_i \propto T^{3/2}\exp(-E_G/2kT)$, I_L also strongly depends on temperature. Assuming that the main cause of current generation comes from impurity or defect levels (traps) in the band gap and using a single trap ansatz [677] with equal trapping cross sections for electrons and holes (more precisely $v_{th}^p\sigma_p \approx v_{th}^n\sigma_n$, v_{th} = thermal velocity), the dependence of the generation lifetime on the trap level position is approximately $\tau_g \propto \cosh(\Delta E_t/kT)$, where $\Delta E_t = E_t - E_f$ is the energy distance of the trap level to the intrinsic Fermi level E_f (see also eq. (8.109) in section 8.12.1). This dependence is minimal for $\Delta E_t = 0$ and shows that leakage current generation by defects is most effective if (deep-level) defects are located near E_f, that is, near the middle of the band gap according to (8.21). In this case, the temperature dependence of τ_g is given only by that of the thermal velocities (given in eq. (4.88) on page 106), that is, $\tau_g \propto T^{-1/2}$, yielding an approximate $T^2\exp(-E_G/2kT)$ dependence for the leakage current. Since the temperature dependence of I_L is very important for many applications, usually a parametrisation is employed leading to:

$$I_L^{gen} \propto T^2 \exp\left(-\frac{E_a}{2kT}\right)\,, \tag{8.62}$$

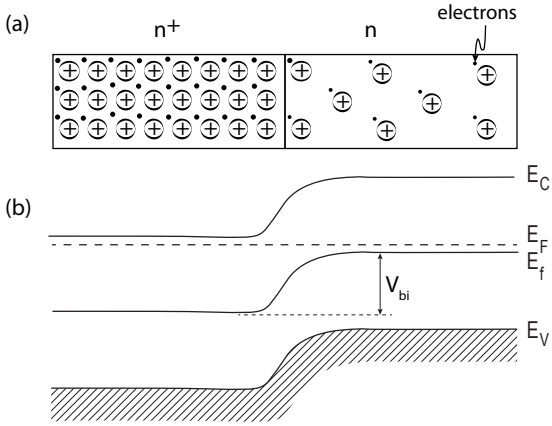


Fig. 8.22 An n^+n boundary. (a) Contact of two differently n-doped semiconductor layers, (b)representation in the energy-band model (adapted from [677] with kind permission of Springer Science+Business Media).

where the parameter E_a is called the *activation energy* or *effective energy* having a best-fit value of $E_a = 1.21\,\text{eV}$ [474, 308].

The typical order of magnitude of the leakage current in a silicon detector at room temperature lies in the range of nA/cm^{-2} [654], but it can reach values of $\mu\text{A/cm}^{-2}$ when the detector is exposed to high radiation fluxes. Measurement of the leakage current therefore is an important method to characterise radiation damage (see section 8.12).

Surface contributions to the leakage current mostly have their origin in the fabrication or the handling of the detector, like damage of the surface or deposits of materials. One can distinguish both contributions (surface and volume) in principle by their different characteristic dependences on the bias voltage. The volume contribution is proportional to the volume under the electrode, that is, proportional to $d \propto \sqrt{V_{\text{ext}}}$, whereas the surface current contribution is rather more linear with the applied voltage. The latter, however, also depends on the geometry of the surface structures. In most cases the volume contribution to the leakage current is dominant.

8.3.3 The n^+n or p^+p boundary

Boundaries of the same doping type but with different doping concentrations, (n^+n or p^+p),[6] also create a potential gradient similar to a pn structure. As fig. 8.22 shows for an abrupt n^+n structure, a space-charge region is also created in this case if due to the concentration difference electrons diffuse from the n^+ region into the n region. The n side hence becomes negative whereas the n^+ side gets positive. The generated electric field stops the diffusion process. In the energy-band model the Fermi level E_F is constant over the structure after thermal equilibrium is reached, the energy bands bending down from left to right. As for the pn boundary a diffusion voltage V_{bi} is created. The conduction characteristics of n^+n or p^+p structures are (nearly) ohmic.

8.3.4 The metal–semiconductor transition

At the metal contacts of a semiconductor detector a different type of interface is created, a metal–semiconductor transition. The behaviour of this interface depends

[6] In the literature it is customary to denote strong (weak) or very strong (very weak) doping concentrations with n^+ (n^-) and n^{++} (n^{--}), relative to the n doping of the substrate; correspondingly for p doping.

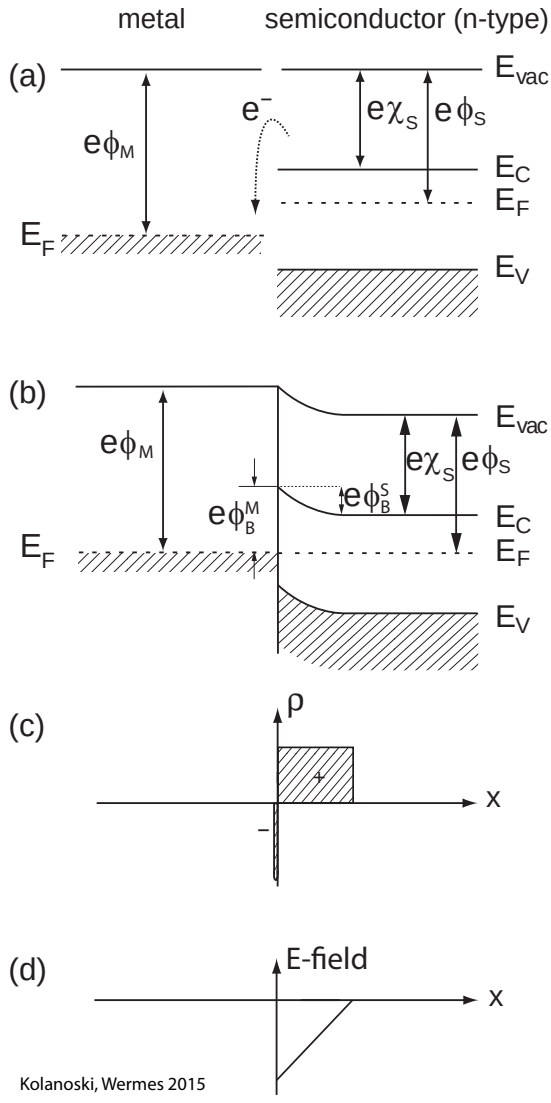


Fig. 8.23 Metal–semiconductor contact with $e\phi_M > e\phi_S$ (Schottky contact) without external voltage applied: (a) metal and n-doped semiconductor before contacting, each being separately in thermal equilibrium, (b) contact in thermal equilibrium after electron transport (with resulting energy-band bending), (c) charge distribution, (d) electric field.

on the type of metal and on the doping type and strength of the semiconductor. The following characteristic properties of metals are essential to understand metal–semiconductor contacts:

– Different to semiconductors the conduction band in metals is partly filled with electrons (see fig. 8.7). Hence the Fermi level sits inside the conduction band of the metal.
– In metals the number of conduction electrons is very large in comparison (in the order of $10^{23}\,\mathrm{cm}^{-3}$). In a perfect conductor no electric field can develop ($\vec{E} = 0$) and there is no potential difference between the ends of the conductor. A description of the metal surface of the boundary is therefore entirely determined by the surface charges.

Figure 8.23 illustrates the energy relationships for a metal–semiconductor transition assuming an n-type semiconductor. The *work function* represents the energy needed to lift an electron from the Fermi energy into the *continuum*, also called *vacuum*: $e\phi = E_{vac} - E_F$. For metals ($e\phi_M$) the work function depends on the metal type, for semiconductor($e\phi_S$) on type and level of the doping. Another quantity, important for the semiconductor side of the interface, is the *electron affinity* $e\chi_S$. The electron affinity describes the difference in energy of a free electron to the energy it has when it has entered the lattice environment. It hence is defined by the energy difference between the bottom edge of the conduction band and the continuum (vaccum): $E_{vac} - E_C$. Note that $e\chi_S$ is independent of the doping concentration and is smaller than $e\phi_S$.

The Schottky contact. We consider first the case that the metal work function is larger than the semiconductor work function ($e\phi_M > e\phi_S$) and the Fermi level is inside the band gap, that is, $e\chi_S < e\phi_S$ (not too high doping). This case is illustrated in fig. 8.23 for the transition of a metal to n-type silicon. The relationships for p-type silicon can be found in the relevant literature (e.g. in [929]), in particular the Fermi level relationships are reversed. Figure 8.23(a) shows the boundary surfaces before coming in contact, each in thermal equilibrium. When put in contact electrons of the semiconductor flow into the metal due to the existing energy difference, until thermal equilibrium is reached again and the Fermi levels in metal and semiconductor are equal ($E_F^M = E_F^S$). The bands bend up in the semiconductor towards the boundary (fig. 8.23(b)). In the metal a negative surface charge builds up. In the semiconductor the loss of electrons results in a positive space charge, free of charge carriers (fig. 8.23(c)). Figure 8.23(d) shows the resulting electric field. The relationships recall of those in a diode (pn transition) with a barrier at the boundary. The difference $e(\phi_M - \chi_S) =: e\phi_B^M$ between metal work function and electron affinity is called the *Schottky barrier* and is given by two material parameters. It is this barrier height which metal electrons need to overcome to reach the semiconductor. The height of $e\phi_B^M$ is unchanged by an external bias voltage (see e.g. [928], p. 164). On the semiconductor side, however, the barrier height for electrons $e(\phi_M - \phi_S) =: e\phi_B^S$ is given by the height of the band bending. The height of this barrier is determined by the built-in voltage $V_{bi} = \phi_B^S$, which is quite analogous to the pn transition and can be increased or decreased by the applied external bias voltage. Such a rectifying contact is called the *Schottky contact*.

In thermal equilibrium the two currents formed by electrons from the metal into the semiconductor and from the semiconductor into the metal are balanced. With an applied external bias voltage V_{ext} the potential difference at the interface is either reduced (forward bias) or increased (backward bias). The current from the metal into the semiconductor is unchanged in comparison to the equilibrium case since the barrier height is unchanged. The current flow from the semiconductor into the metal, however, changes exponentially similar to the current in a pn boundary (compare with eq. (8.57)):

$$I = I_S \left(\mathrm{e}^{eV_{ext}/kT} - 1 \right) \tag{8.63}$$

with the saturation (reverse) current

$$I_S = A_R^* \, T^2 \mathrm{e}^{-e\phi_B^M/kT} \,,$$

where A_R^* is the effective Richardson constant with values $A_R^* = 110\,\mathrm{A\,K^{-2}cm^{-2}}$ for

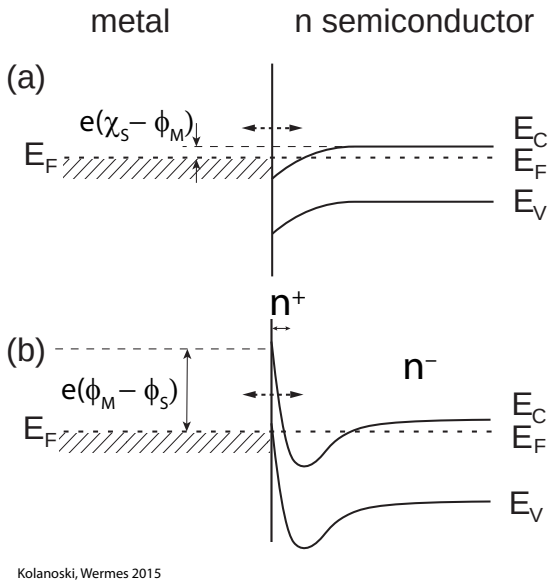


Fig. 8.24 Ohmic contact: a) vanishing contact resistance, b) narrow tunnel barrier arising from high doping ($> 10^{19}\,\mathrm{cm}^{-3}$) of the semiconductor near the interface (see also [926]).

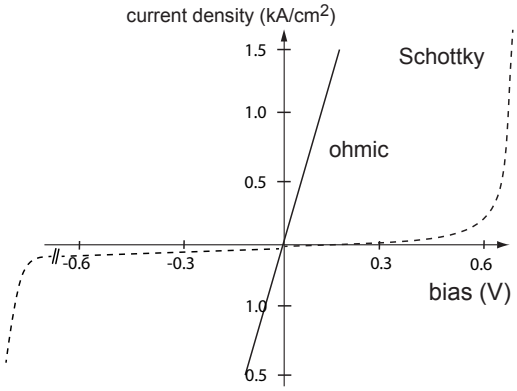


Fig. 8.25 Typical current–voltage characteristics for a rectifying metal–semiconductor contact (Schottky: GaAs with doping concentration of $10^{15}/\mathrm{cm}^3$, dashed line) and an ohmic contact ($10^4\ \Omega\,\mathrm{cm}^2$, solid line).

n-type silicon and $A_R^* = 132\,\mathrm{A\,K^{-2}cm^{-2}}$ for p-type silicon, respectively [926]. The external voltage is positive for forward bias and negative for reverse bias, respectively. Contrary to the pn boundary, where the characteristic is mainly determined by minority carriers, in a metal–semiconductor contact the majority carriers are responsible for the current–voltage characteristic (8.63).

The ohmic contact. In the case that the metal work function is smaller than the semiconductor work function, $e\phi_M < e\phi_S$, the interface becomes an *ohmic contact* defined by a vanishingly small contact resistance. This is because no or only vanishingly small barriers exist for electrons (fig. 8.24(a)). Electrons move from the metal into the n-type semiconductor, there generating a negative net charge. The bands bend up, as seen from the interface. No space-charge region builds up and the potential barrier for electron flow to the metal is small. Only small external voltages are needed to ensure electron movement over the boundary in both directions. Figure 8.24(a) shows this case after thermal equilibrium has set in ($E_F^M = E_F^S$).

Often it is difficult to find a suitable metal which does not produce a barrier when coming in contact with a doped semiconductor and at the same time has good contacting properties (sticks to the semiconductor, can be contacted with other metals (wire bonds), etc.). Therefore, another method is usually employed to establish an ohmic contact for silicon detectors for metals with $\phi_M > \phi_S$. By means of very high n^+ doping ($\gtrsim 10^{19}\,\mathrm{cm}^{-3}$) at the contact surface (about 10 nm deep) the Fermi level E_F is pulled up almost to the conduction band edge. After contacting with the metal ϕ_M is hence smaller than ϕ_S. At the metal-facing edge of the n-type semiconductor a spatially very thin barrier is formed which can be tunnelled by electrons upon application of a voltage (fig. 8.24 (b)), thus creating an almost linear current–voltage characteristic. The further course of the band structure follows the one of an n^+n structure, as described in section 8.3.3.

In fig. 8.25 the current–voltage characteristics for both cases are shown. The Schottky transition yields a rectifying (diode) contact, the ohmic contact is drawn here as a linear[7] characteristic, as for a resistor.

8.3.5 The MOS transition

A MOS transition (MOS = *metal–oxide–semiconductor*) is a double interface structure made of three media: metal, oxide, and semiconductor [991]. The MOS structure plays a very important role in microelectronics because it is the most widely used structure in field-effect transistors (FET). For detailed and extensive descriptions we refer the reader to the manifold literature, for example [896, 929, 248] or [486].

By far the largest part of chip electronics, including the readout of detectors, see section 17.6, is based on MOSFETs which are usually employed as a combination of NMOS and PMOS transistors realised on the same substrate (CMOS electronics), thus allowing one to build complex circuits based on complementary logic (see e.g. [985]). Today CMOS technology and CMOS logic (see also section 17.7.4) represent the most used technology for integrated circuits and the most often used logic family. In order to accommodate both transistor types in the same substrate one transistor type is embedded in a dedicated, differently doped area, a so-called 'well'. For example, in a p-type substrate PMOS transistors are embedded in n-wells. An essential advantage of the CMOS technology is that circuits can be made without any resistors and that digital circuitry can be designed such that the transistors only draw current during switching such that the power used to a large extent only depends on the switching frequency. For deeper reading we refer to the manifold respective literature given above.

Strongly connected are sensor and readout chip in silicon strip and particularly in silicon pixel detectors[8] (sections 8.6.2, 8.7 and 8.10). Other semiconductor detectors contain MOS structures, too, for example MOS-CCDs (section 8.9). DEPFET pixel detectors (section 8.10.2) contain an individual MOSFET in every pixel to convert the signal charge into an amplified current.

A perspective as well as a cross-sectional illustration of a MOS structure is shown in fig. 8.26. Nowadays, for transistors the metal is almost exclusively replaced by strongly doped polysilicon (n^{++} or p^{++}) since polycrystalline silicon sustains higher

[7] The current–voltage characteristics of an ohmic metal–semiconductor contact may not be strictly linear. It is regarded as 'ohmic' if the voltage drop is small compared to the voltage drop across the 'active' (semiconductor) region of the device [818].

[8] 'Pixel' is a short term for 'picture element'. Pixel detectors contain two-dimensional sensor units. In the language used they are understood to be much smaller than 'pads' in corresponding detectors.

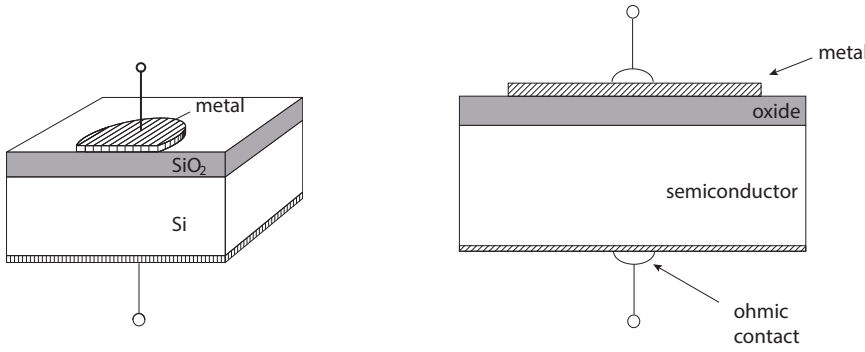


Fig. 8.26 Perspective view (a) and cross section (b) of a MOS structure.

temperatures without reacting with the oxide. The physics of the MOS structure is, however, unchanged and so is the name (metal). Metal contacts to semiconductor detectors – if not directly contacted to the semiconductor – can also be applied via an oxide layer and then typically are Al–SiO_2–Si MOS structures.

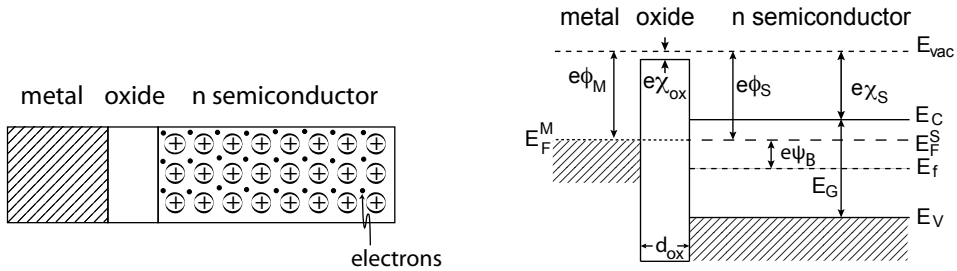
Different to the so far discussed interfaces semiconductor–semiconductor and metal–semiconductor, we now have an insulator in between the metal and the semiconductor. To a very good approximation no DC current flows through this insulator when an external bias voltage is applied. We can thus consider metal and semiconductor separately in thermal equilibrium. Under the assumption of an abrupt transition, the energy-band model provides a good description of the phenomena that depend on the applied bias V_{ext}. To illustrate this we consider in figs. 8.27 to 8.32 MOS structures with an n-doped semiconductor (corresponding to the situation for a PMOS transistor), following our choice so far in this chapter. For the description with p-doped semiconductor material, often used in the literature (situation for NMOS transistors) the same considerations hold with reversed signs (see e.g. [928]).

We consider first an ideal MOS structure defined by the (non-realistic) assumptions that the work functions of metal and semiconductor are equal, $e\phi_M = e\phi_S$, and that besides charges in the semiconductor, other charge only exists on the metal surface; this means in particular that there are no (fixed) charges in the oxide or at the interface boundaries [929]. This also means that without external bias the structure is field-free and no energy-band bending appears; the bands are 'flat' (fig. 8.27(b)). From the figure we extract:

$$0 \;=\; e\phi_M - e\phi_S = e\phi_M - \left(e\chi_S + \frac{E_G}{2} - e\psi_B \right) ,$$

where it has been assumed that the intrinsic Fermi level of the semiconductor E_f lies in the middle of the band gap (see eq. (8.21)). The notation is as in section 8.3.4: χ_S = electron affinity, E_G = band gap energy, $e\psi_B = E_F^S - E_f$ is the difference between extrinsic and intrinsic Fermi level in the semiconductor bulk.

A real MOS structure differs from the ideal structure of fig. 8.27 in that the difference of the work functions does not vanish and that generally there are charges present in the oxide or at the interfaces. The latter is particularly relevant when radiation damage is an issue (see section 8.12). Here we first neglect oxide charges. The



(a) Structure. (b) Energy-band representation.

Fig. 8.27 Ideal MOS structure in which there is no electric field (adapted from [929] and [677]).

size of the work function in the semiconductor $e\phi_S$ depends on the doping type and concentration. For silicon it lies typically between 4.1 eV and 5.2 eV. For aluminium, a typical contact metal for silicon detectors, the work function in metal $e\phi_M$ is, with 4.1 eV, most often smaller than the semiconductor work function (fig. 8.28(a)):

$$e\phi_{MS} = e\phi_M - e\phi_S < 0\,. \tag{8.64}$$

The voltage $V = \phi_{MS}$ corresponds to the diffusion voltage V_{bi} in (8.37) for the pn boundary. In the oxide the electron affinity is substantially smaller than the metal work function and also than the affinity in the semiconductor; in SiO_2 it is $e\chi_{ox} = 0.9$ eV.

When contacting the layers the Fermi levels in metal E_F^M and semiconductor E_F^S become equal (see also page 278) and the vacuum levels over the different regions of the structure remain connected. Note that here a constant Fermi level is achieved by rearrangement of charges in the layers without effective charge transport in the form of a current, as is the case for a pn junction. The energy bands (i.e. the potentials) on the one side shift relative to the other (fig. 8.28(b)). Since there is no charge transport through the oxide possible from one side to the other, the energy-band bending is due to a charge carrier redistribution (accumulation or depletion of carriers). The potential (and the energy bands) linearly progresses in the oxide, since the field in the oxide is constant; in the semiconductor the dependence is quadratic up to the region without space charge (that is, without charge accumulation or depletion) in which the energy bands are flat (constant). In order to re-establish the flat-band state a voltage must be applied externally (for a MOS transistor this is the gate voltage) of size

$$V_{ext} = V_{FB} = -\phi_{MS}\,, \tag{8.65}$$

accordingly called a *flat-band voltage*. The system thereby leaves thermal equilibrium ($E_F^M \neq E_F^S$). Attention must be paid regarding the sign of the potentials and voltages: we apply V_{ext} on the metal side relative to ground (semiconductor side). ϕ_{MS} is the potential difference between metal and semiconductor which is compensated by $V_{ext} = V_{FB}$ (between metal contact and opposite electrode) to establish the flat-band state. V_{FB} is positive for n-bulk silicon and negative for p-bulk.

From these relationships we distinguish the following states of the MOS transition depending on the applied voltage between metal and semiconductor:

Flat-band condition $V_{ext} = V_{FB}$ (fig. 8.29):

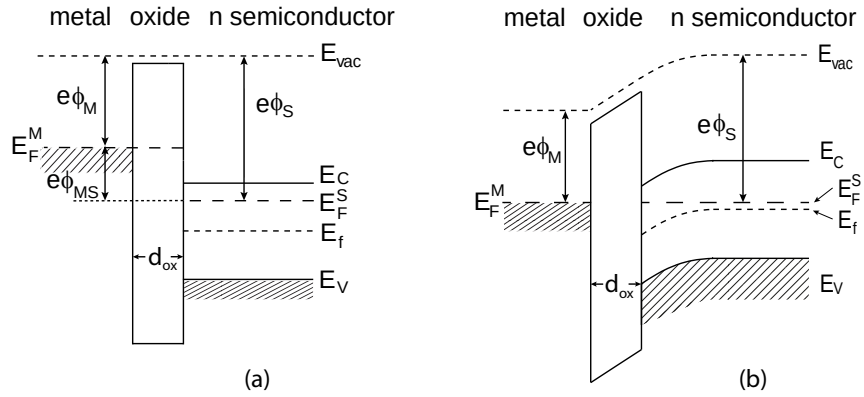
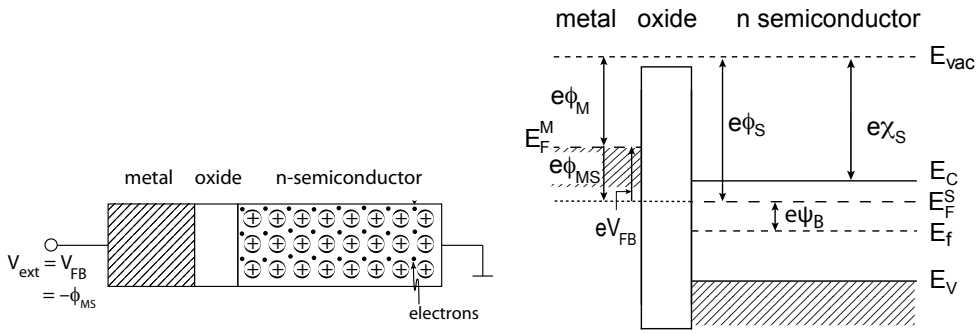


Fig. 8.28 Real MOS structure without external bias. (a) Energy-band structure before contact; (b) MOS structure after contact in thermal equilibrium showing the band bending (adapted from [929] and [677]).

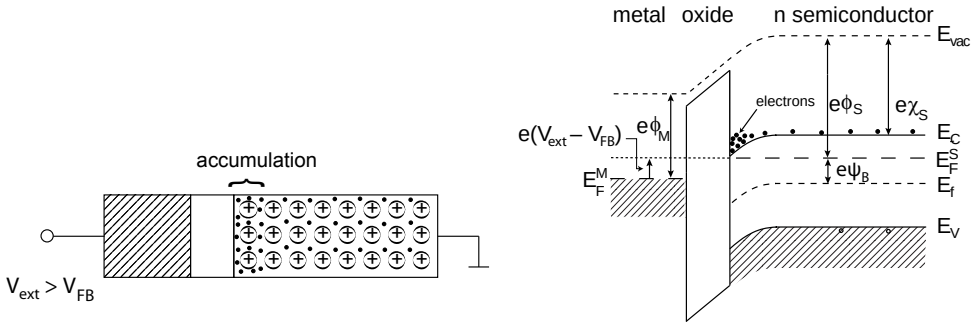


(a) Structure. (b) Energy-band representation.

Fig. 8.29 MOS structure after applying an external bias voltage $V_{ext} = V_{FB} = -\phi_{MS}$: flat-band condition (adapted from [929] and [677]).

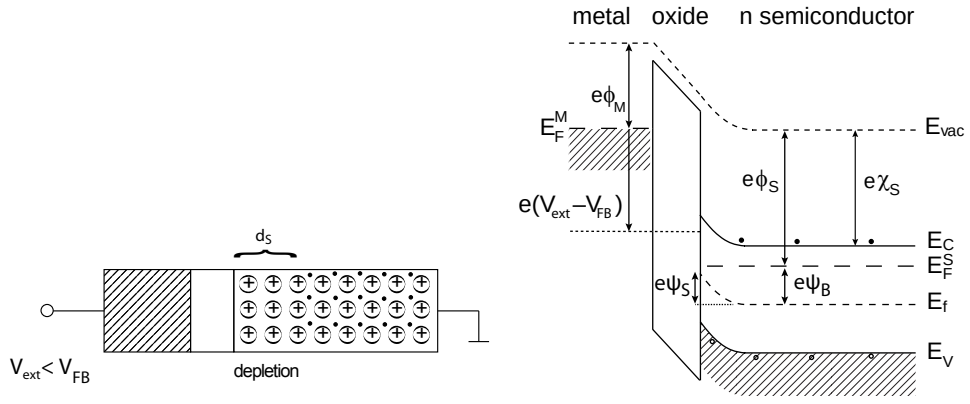
No energy-band bending exists. The free charge carriers of the semiconductor are uniformly distributed over the crystal up to the interface.

Accumulation $V_{ext} > V_{FB}$ (fig. 8.30):
The positive potential (relative to the equilibrium) on the metal side also moves the potential at the oxide–semiconductor boundary in the positive direction and pulls the mobile electrons of the semiconductor towards the insulator. An even stronger (than already present) accumulation of electrons at the oxide–semiconductor boundary occurs. As a consequence the energy bands bend up, as seen from the interface towards the semiconductor. Since no current flows through the structure there is also no potential drop in the semiconductor and the energy levels farther away from the interface remain flat. At the boundary of the semiconductor to the oxide the distance of the conduction band to the Fermi energy becomes smaller. Due to the exponential



(a) Structure. (b) Energy-band representation.

Fig. 8.30 MOS structure: accumulation. Electrons are pulled towards the interface by the positive external voltage (adapted from [929] and [677]).



(a) Structure. (b) Energy-band representation.

Fig. 8.31 MOS structure: depletion. The negative external voltage pushes the electrons in the n semiconductor away from the interface (adapted from [929] and [677]).

dependence of the electron concentration (8.27), $n_n = n_i \exp((E_f - E_F)/kT)$, the electron accumulation layer is so thin that it can be approximated by a surface-charge density:

$$Q_{acc} = -\epsilon_{ox}\epsilon_0 \frac{V_{ext} - V_{FB}}{d_{ox}} = -C_{ox}\,(V_{ext} - V_{FB})\,, \tag{8.66}$$

where

$$C_{ox} = \epsilon_{ox}\epsilon_0/d_{ox} \tag{8.67}$$

is the oxide capacitance per area.

Depletion $V_{ext} < V_{FB}$ (fig. 8.31):
The negative potential (relative to the equilibrium) pushes the electrons away from the insulator. A depletion zone develops at the interface between insulator and semiconductor. The energy bands bend down, as seen from the interface. The energy distance between the valance band and Fermi level E_F^S decreases at the interface. To understand the electrical relationships in the structure the following considera-

tions are useful. In a one-dimensional approximation, surface charge, electric field $\mathcal{E}_S$ (positive direction from the metal to the semiconductor) and potential ψ_S at the semiconductor boundary can be expressed relatively simply as a function of the depletion depth d_S, if one assumes that the potential vanishes deep in the semiconductor bulk and no oxide charges are present [677]:

$$Q_S = eN_D d_S\,, \qquad \mathcal{E}_S = -\frac{eN_D}{\epsilon_S\epsilon_0} d_S\,, \qquad \psi_S = \frac{eN_D}{2\epsilon_S\epsilon_0} d_S^2\,. \tag{8.68}$$

The potential ψ_S characterises the strength of the band bending (fig. 8.31(b)). The field in the oxide $\mathcal{E}_{ox}$ is constant and according to Gauss' law directly proportional to the field at the semiconductor surface:

$$\mathcal{E}_{ox} = \frac{\epsilon_S}{\epsilon_{ox}}\mathcal{E}_S = -\frac{eN_D}{\epsilon_{ox}\epsilon_0} d_S\,. \tag{8.69}$$

The difference between external voltage and flat-band voltage therefore amounts to:

$$V_{ext} - V_{FB} = -\psi_S + \mathcal{E}_{ox} d_{ox} = -\frac{eN_D}{\epsilon_0} d_S \left(\frac{d_S}{2\epsilon_S} + \frac{d_{ox}}{\epsilon_{ox}}\right)\,. \tag{8.70}$$

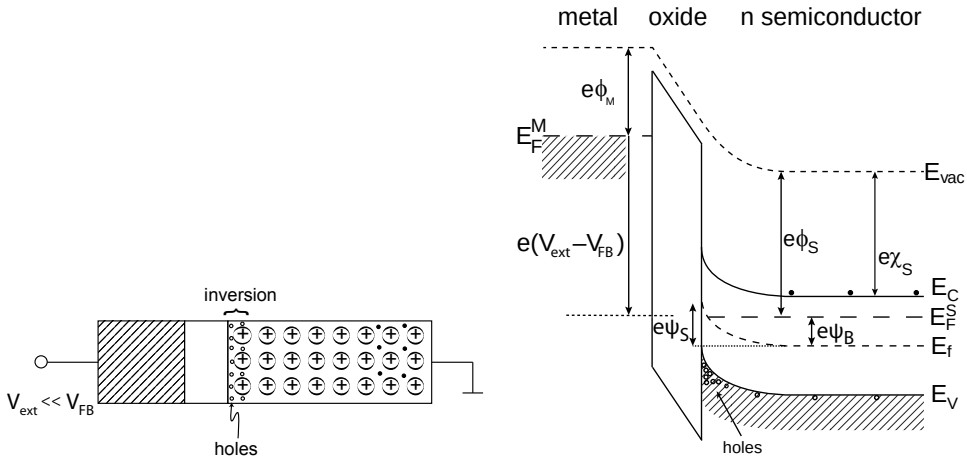
The thickness d_S of the depletion layer in the semiconductor hence depends on the applied voltage relative to the flat-band voltage and on the electrical parameters of the oxide:

$$d_S = \sqrt{\frac{2\epsilon_S\epsilon_0}{eN_D}(V_{FB} - V_{ext}) + \left(\frac{\epsilon_S}{\epsilon_{ox}} d_{ox}\right)^2} - \frac{\epsilon_S}{\epsilon_{ox}} d_{ox}\,. \tag{8.71}$$

Inversion $V_{ext} \ll V_{FB}$ (fig. 8.32):
Choosing an external bias voltage which is substantially smaller than the flat-band voltage, the band bending will increase even more strongly until at the insulator–semiconductor interface the intrinsic Fermi level E_f is higher than the extrinsic Fermi level E_F^S in the middle of the semiconductor bulk, that is, $e\psi_S > e\psi_B$. The effective conductor type at the Si–SiO_2 interface changes from n to p conducting, the carriers being thermally generated holes. The minority carrier concentration (here: holes in the n region) at the interface (denoted by (0)) now increases exponentially:

$$p_n^S = n_i \exp\left(\frac{E_f(0) - E_F^S(0)}{kT}\right)\,. \tag{8.72}$$

Thermally emerging minority carriers (holes) are sucked from the semiconductor bulk to the insulator layer. The charge carrier layer immediately at the interface boundary is inverted (minority carriers instead of majority carriers) and screens the adjacent depletion zone in the semiconductor from the external voltage. Inversion is the state of the MOS structure which is essential for field-effect transistors [988]. The current in the transistor channel flows between the contacts called *source* and *drain* and is steered by the *gate* contact. The gate voltage is our external voltage V_{ext}. In inversion the drain current is made of minority carriers (i.e. holes for n substrate). One speaks of 'weak' inversion if $e\psi_S \gtrsim e\psi_B$, which means that at the semiconductor interface to the oxide $E_f \gtrsim E_F^S$ holds. Then the hole concentration at this point is as large or larger than the intrinsic electron concentration n_i. 'Strong'



(a) Structure. (b) Energy-band representation.

Fig. 8.32 MOS structure: inversion. The external voltage is large such that the electrons are pushed far away from the oxidee into the semiconductor. Thermally produced holes can therefore accumulate at the interface towards the oxide (adapted from [929] and [677]).

inversion sets in if the hole concentration of the inversion layer p_n^S becomes larger than the electron concentration in the semiconductor bulk, which is governed by the bulk doping concentration ($n_n^B \approx N_D$). This is the case if

$$\begin{aligned} p_n^S &= n_i \exp\left(\frac{e\psi_S - e\psi_B}{kT}\right) \gtrsim n_n^B \approx N_D = n_i \exp\left(\frac{e\psi_B}{kT}\right) \\ &\Rightarrow \psi_S - \psi_B \gtrsim \psi_B \\ &\Rightarrow \psi_S \gtrsim 2\psi_B \end{aligned} \tag{8.73}$$

holds. The depletion depth then is almost maximal since—due to the screening effect of the inversion layer—further lowering of V_{ext} rather causes an increase of the layer charge Q_{inv}. With (8.68) and the assumptions made there, it is

$$d_{max} = \sqrt{\frac{2\epsilon_S\epsilon_0(2\psi_B)}{eN_D}}\,, \tag{8.74}$$

having typical values in the range of 10 nm. The electric field at the semiconductor–oxide interface correspondingly is:

$$\mathcal{E}_S = -\frac{eN_D}{\epsilon_S\epsilon_0} d_{max} = -\sqrt{\frac{2eN_D(2\psi_B)}{\epsilon_S\epsilon_0}}\,. \tag{8.75}$$

In the oxide the field is constant and can, as in (8.69), be obtained from $\mathcal{E}_S$ by multiplying with the ratio of the permittivities of oxide and semiconductor, and taking into account the inversion layer at the oxide–semiconductor boundary in addition:

$$\mathcal{E}_{ox} = -\frac{1}{\epsilon_{ox}\epsilon_0}\left(eN_D\, d_{max} + Q_{inv}\right)\,. \tag{8.76}$$

The applied potential V_{ext} at the metal contact hence drops over the MOS structure to the semiconductor surface as follows (with $\phi_{MS} = -V_{FB}$ from (8.64)):

$$\begin{aligned} V_{ext} &= \phi_{MS} - 2\psi_B + \mathcal{E}_{ox} d_{ox} \\ &= \phi_{MS} - 2\psi_B - \frac{d_{ox}}{\epsilon_{ox}\epsilon_0} \left(eN_D\, d_{max} + Q_{inv}\right) \qquad (8.77) \\ &= \underbrace{\phi_{MS} - 2\psi_B - \frac{\sqrt{2e\ \epsilon_S \epsilon_0 N_D (2\psi_B)}}{C_{ox}}}_{V_T} - \frac{Q_{inv}}{C_{ox}}\,, \qquad (8.78) \end{aligned}$$

where in the last step d_{max} in (8.74) and C_{ox} in (8.67) have been used. Equation (8.77) defines the threshold voltage V_T as the external voltage necessary to reach strong inversion. Raising $|V_{ext}|$ the charge in the inversion layer increases according to

$$Q_{inv} = (V_T - V_{ext})\ C_{ox}\ . \qquad (8.79)$$

Oxide charges. The above picture of the MOS structure is changed by the existence of charges fixed in the oxide layer. Since the current in the channel of MOS transistors apart from the gate voltage also depends on charges residing at the interfaces or in the MOS structure, the existence of additional charges can influence the transistor's behaviour. By various effects, in particular due to the influence of external radiation, different physical types of charge appearances can be created (see also section 8.12.3):

Fixed oxide charges: These are positive Si ions directly at the Si–SiO_2 interface, created during the oxidation process. Through reactions with Si–O and Si–Si complexes they become stationary.

Fixed SiO_x defects: These lattice defects can also be created at the interface during oxidation or by radiation damage during detector or chip operation in the entire oxide layer. They act as charge traps.

Fixed interface charges: Such defects can be created by lattice defects as for example unsaturated bonds at the interface.

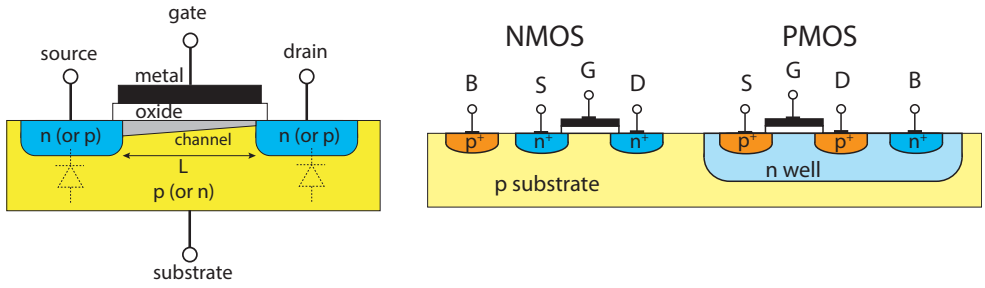
Mobile charges: These are predominantly Na^+ or K^+ ions caused by impurities in the oxide layer or at the Si–SiO_2 boundary.

These parasitic charges together are denoted *oxide charges* in the following. They generate an electrical field which alters the flat-band voltage:

$$V_{FB} = -\phi_{MS} - \frac{\Sigma Q_i^{ox}}{C_{ox}}\,, \qquad (8.80)$$

with C_{ox} from (8.66) being the oxide capacitance per area with thickness d_{ox}.

It is important to know the charge density in the oxide and the interface charge density in order to be able to describe the response of a semiconductor detector or a readout chip, for example after damage by ionising radiation (described in section 8.12.3). While the charge density in the oxide can be determined from the change of the flat-band voltage after irradiation, the interface charges can be measured by a special diode which is controlled by an external MOS gate (*gate controlled diode*, GCD) [165].



(a) MOSFET. (b) CMOS.

Fig. 8.33 (a) Basic MOSFET structure. The diodes (dashed) indicate the pn boundaries between terminal and substrate. (b) CMOS realisation of NMOS and PMOS transistors on the same substrate (here p-type).

MOS transistor. The MOS structure is the basis of the MOS field-effect transistor or MOSFET. There is a variety of field-effect transistor types which follow the same principle, namely that a transistor current is controlled by an electric field (a voltage) applied externally. For example, in a *junction field-effect transistor* (JFET) the MOS's metal–oxid junction is a pn or np junction instead. MOSFETs, however, are by far the most common transistors for digital circuits and MOS transistor ICs are produced in thousands of wafer factories around the world. A discussion of their characteristics, properties and operation goes far beyond the scope of this book and we refer to the dedicated literature, for example [929, 639, 472] and many others. Here we intend to only very briefly sketch the MOSFET structure and explain some nomenclature which will be used further in this text.

Figure 8.33(a) shows the rudimentary structure of a MOSFET. The MOS structure discussed above in this section is called the *gate*. Depending on the gate voltage, it can provide accumulation, depletion and inversion (see section 8.3.5) at the boundary between oxide and silicon (called *substrate* or *bulk*). In addition two highly doped regions are implanted, called *source* and *drain*. They form pn or np diodes with the oppositely doped substrate. Usually MOSFETs are operated 'in inversion'; this means if the substrate is p doped (n doped) the charge carriers under the gate are electrons (holes) and vice versa. With proper biasing, drain at a 'higher' potential for the respective charge carrier polarity than source, a current between source and drain is generated whose strength can be controlled by the voltage applied to the gate. The current flows in a geometrically confined area underneath the gate, narrowing somewhat towards the drain terminal due to a gate–substrate potential drop when going from drain to source. The current is directly proportional to the area given by the gate length L (fig. 8.33(a)) and its width W. MOS transistors (MOST) are called NMOS, if the substrate is p doped and source and drain terminals are n doped. They are called PMOS if the terminals are p doped and the substrate is n doped.

Obviously in a p-doped substrate only NMOS transistors can be realised, unless a large n-doped area, called an *n well* is implanted which can host the implantations that are needed for a PMOS transistor. Such a construction is shown in fig. 8.33(b), where an NMOST and a PMOST are placed next to each other. The technology realising such constructions is called *complementary* MOS or CMOS technology. CMOS realisations of electronic circuits have advantages compared to those realised with PMOS or NMOS transistors only. The most important advantage is the fact that typical circuit elements,

like for example the one in fig. 17.4(a) on page 721 (an inverter) with an NMOS and a PMOS transistor on top of each other, only draw current during the (short) time when the input level changes from high to low or vice versa. For a DC level at the input either the PMOST or the NMOST is switched off and the circuit does not draw current.

8.4 Particle detection by semiconductor detectors

As already alluded to in the introduction to this chapter, semiconductor detectors have become the most important type for tracking detectors in high energy physics experiments, especially when high rate and radiation levels are faced such as at the LHC. Dedicated literature is available, for example the books by Hartmann [506] or by Rossi et al. on pixels [837]. Here and in sections 8.5–8.10 we describe important fundamentals and operation characteristics of silicon detectors used for tracking and explain essential features and differences of the various detector types. As of today tracking detectors and also large imaging detectors are based on silicon as detection material. Developments of other sensing media such as GaAs, CdTe and artificial diamond for smaller imagers or trackers are ongoing. The main features of these alternative semiconductors are described in section 8.13. Radiation effects in silicon are dealt with in section 8.12.

When a charged particle loses energy by ionisation or a photon is absorbed in a semiconductor detector (see chapter 3) part of the released energy is used to generate electron–hole pairs. The charge carrier pairs are separated in the electric field applied to the semiconductor and drift inside the bulk toward the electrodes on which the movement induces an induction signal. The bulk must be sufficiently free or depleted from other charge carriers. The electronic signal is determined in size and shape by the generated of e/h pairs, the velocity of their drift movement, and the electrode geometry. This is explained in detail in chapter 5. The drift velocity depends on the carrier mobility μ and the magnitude of the electric field in the semiconductor according to the Drude model [366]:

$$v_D = \mu(E)\, E\,,$$

where the mobility itself is field dependent such that at high electric fields the drift velocity saturates (see section 4.7.1 on page 122).

In a silicon detector (pn diode), for full depletion (without overvoltage, see section 5.4.1), the electric field linearly drops to zero from one side to the other. The average field is just half that of the maximum field right at the boundary. In overvoltage operation a constant amount is added to the linear functional course (see fig. 5.10 on page 148).

Spatially sensitive semiconductor detectors (microstrip or pixel detectors) used in high energy physics are usually very thin (typically 200–300 µm). Typical velocities in such detectors are in the order of 50 µm/ns. The maximum time that the charge carriers need to traverse the space charge region hence typically is rather short:

$$t = \frac{d}{v_D} \approx \frac{200\text{–}300\,\mu\text{m}}{50\,\mu\text{m/ns}} \approx 4\text{–}6\,\text{ns}\,.$$

To compute the average charge signal (total signal, integrated over time) obtained when a high energy charged particle traverses a 300 µm thick silicon detector, we need to know the particle's average energy loss and the energy w_i needed on average to

generate an electron–hole pair (see table 8.2). This energy is larger than the band gap E_G since part of the deposited energy is used for other processes but e/h creation, namely for lattice (phonon) excitations.

The energy loss of ionising particles is distributed according to the Landau distribution (fig. 14.9 on page 554) which is asymmetric and therefore the *most probable value* (mpv) differs from the mean energy loss $\langle \frac{dE}{dx} \rangle$. For a minimum-ionising particle in a 300 µm thick silicon detector the latter amounts to $\langle \frac{dE}{dx} \rangle \approx 0.39$ keV/µm = 117 keV/300 µm, the most probable value is about 84 keV/300 µm. The average energy w_i to create an electron–hole pair is temperature dependent [483]. A commonly used recent value for w_i at 300 K is 3.65 eV. Hence we obtain for the total charge of a typical signal

$$\left\langle \frac{dE}{dx} \right\rangle \frac{300\,\mu\text{m}}{w_i} = \frac{117 \times 10^3\,\text{eV}}{3.65\,\text{eV}} \approx 32\,000\,e/h \mathrel{\widehat{=}} 5\,\text{fC} \qquad \text{(average)}$$

$$\text{mpv} \qquad \approx 23\,000\,e/h \mathrel{\widehat{=}} 3.7\,\text{fC} \qquad \text{(most probable value)}\,.$$

The time dependence of the current signal for a depleted silicon diode with space charge is given by (5.84) in section 5.4:

$$i_S(t) = \frac{e}{d}\left(\frac{a}{b} - x_0\right)\left(\frac{1}{\tau_e}\mathrm{e}^{-t/\tau_e}\,\Theta\left(T^- - t\right) + \frac{1}{\tau_h}\mathrm{e}^{t/\tau_h}\,\Theta\left(T^+ - t\right)\right), \qquad (8.81)$$

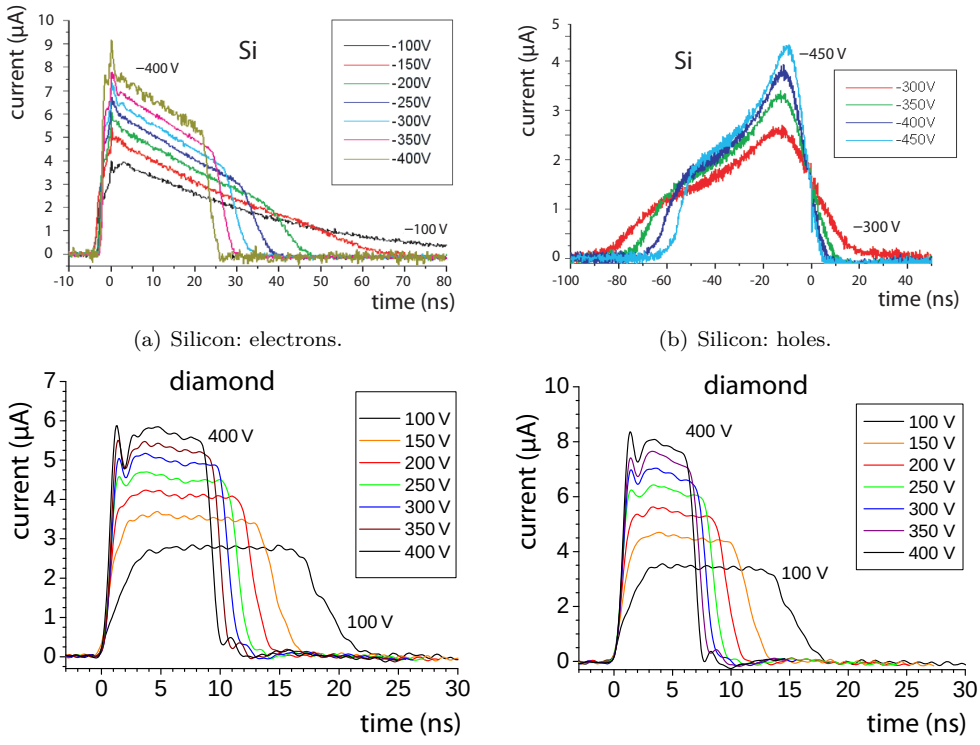
where Θ denotes the step function, x_0 the point of origin of the charge-carrier pair, a, b are parameters describing the electric field, defined in section 5.4, equation (5.72), and $\tau_{e,h}$ are characteristic charge-collection times for electrons and holes, respectively. The corresponding charge signal results after time integration (eq. (5.85) in section 5.4) having the form before arrival of the charges

$$Q_S(t) = -e\,\frac{a - b\,x_0}{b\,d}\left(\mathrm{e}^{\,t/\tau_h}\Theta\left(T^+ - t\right) - \mathrm{e}^{-t/\tau_e}\Theta\left(T^- - t\right)\right). \qquad (8.82)$$

Here T^+ and T^- are the respective arrival times for electrons and holes. For a particle track one must integrate over all current, respectively, charge contributions along the track until their arrival at the electrodes (see section 5.4.2).

Figures 8.34(a) and (b) show the measured current signal as a function of time in a silicon detector in response to impinging α particles on the side opposite to the collection electrode, with set-ups chosen for (a) electron and (b) hole collection [409]. Alpha particles only penetrate a few microns into the detector and hence create a very local and large charge deposition on the injection side. This way one can selectively choose that only electrons (when irradiating from the p side) or holes (when irradiating from the n side), respectively, contribute to the signal generation. The characteristic functional form of the signal shape according to (8.81) can be seen. The exponential shape of the current amplitude decrease (increase) for electrons (holes) with time originates from the linear functional form of the electric field in the silicon bulk due to the constant space charge in the depleted depth of the detector (compare with eq. (5.84) in chapter 5). From the shape of the curves drift velocity and mobility of the charge carriers can be determined (see also table 8.2).

For comparison we show in figs. 8.34(c) and (d) the current signals measured in a single-crystal diamond detector (see section 8.13.4) which does not have a space-charge



(a) Silicon: electrons.

(b) Silicon: holes.

(c) Diamond: electrons.

(d) Diamond: holes.

Fig. 8.34 Measured current signals in a fully depleted silicon detector (1 mm thick) and a single-crystal diamond detector (500 μm) for different detector voltages [409]: (a,b) silicon, (c,d) diamond. The signals are generated by α particles entering on the side opposite to the collecting electrode. α particles, for example from a ^{241}Am source (see table A.1), only have a small range in silicon or diamond, such that the charge deposit occurs rather locally.

region, contrary to silicon, since diamond does not require doping to deplete the bulk from free charge carriers. The current amplitude is constant in time as a consequence of the constant electric field inside the diamond bulk. In the diamond substrate used the hole mobility is about 20% larger than the electron mobility.

Position measurement. In strip or pixel geometries, structured electrodes allow for precise space-point determination of the particle entrance. If there is only digital information available (1 = hit, 0 = no hit) the obtainable resolution is mostly given by the electrode pitch p. The average quadratic deviation (the variance) from the true entrance point is for perpendicular particle entrance and single hit response:

$$\sigma_x^2 = \frac{1}{p}\int_{-p/2}^{-p/2} x^2 dx = \frac{p^2}{12}\,. \tag{8.83}$$

A detailed discussion of the spatial resolution in detectors with segmented electrodes is given in appendix E. 'Binary' resolutions (i.e. without using pulse-height information) obtained with microstrip and pixel detectors with typical electrode pitches between 20 μm and several 100 μm reach about 10 μm.

For analog readout the charge signals measured on the readout electrodes can be

used for a more precise determination of the entrance point by weighted averaging of neighbouring strip or pixel signals. The algorithm performs best if the width of the charge cloud at the electrodes is of similar order of magnitude as the electrode pitch. For the typically 250–300 μm thin microstrip detectors used in particle physics the charge cloud reaches field-strength dependent extensions caused by diffusion in the order of 5–15 μm. Therefore the averaging mostly extends over two strips only:

$$x = \frac{S_1 x_1 + S_2 x_2}{S_1 + S_2}\,, \tag{8.84}$$

where $S_{1,2}$ are the signals (charges) of two neighbouring strips. By employing a very simple ansatz where the uncertainty on the signal parts $S_1 + S_2 = S$ is assumed to be given by the noise N and choosing—without loss of generality—the position at $x = 0$ with $x_1 = -x_2 = p/2$, we obtain the precision of the space-point reconstruction as:

$$\sigma_x = \left(\frac{N}{S}\right)\ p\ \sqrt{\beta}\,, \tag{8.85}$$

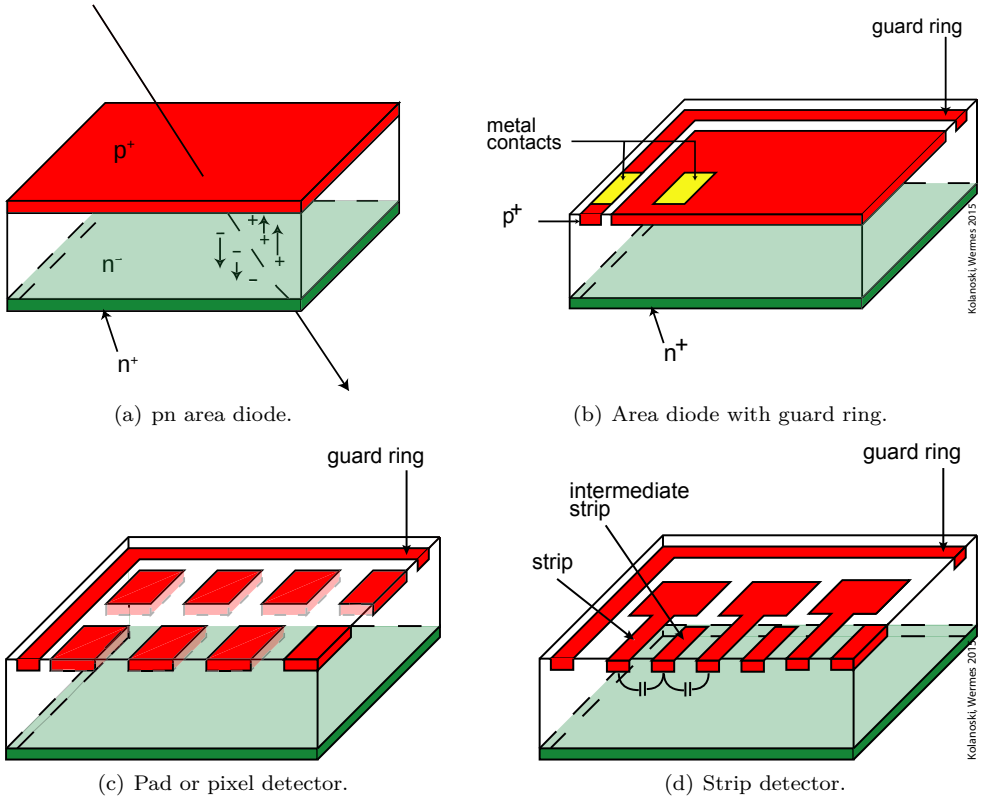
where SNR is the signal-to-noise ratio and β depends on the distribution of the total charge on both strips: $\beta = (S_1 - S_2)^2/S^2 = 1 - 2a_1(1 - a_1)$ with a_1 being the signal fraction S_1/S on strip 1. For (8.85) only uncorrelated noise of the individual readout channels has been considered and correlated noise contributions (common mode noise) have not been taken into account. A more explicit discussion of achievable resolutions with structured electrodes including noise considerations can be found in Appendix E and in [676]. Further effects influencing the achievable spatial resolution are discriminator thresholds and gain fluctuations in readout channels.

In the next section we show that by a method called *capacitive charge division* one can interpolate over readout strip distances which are much larger than the width of the charge cloud.

8.5 Single-sided silicon detectors

We distinguish and subdivide in the following according to single-sided or double-sided structuring and processing of detectors, that is, whether in the fabrication the silicon wafers require micro-structuring technology steps to be applied only on one or on both wafer sides. The former is much simpler and more cost effective which is an important aspect for building large area detectors. Double-sided processing requires special precautions to protect the face that is already processed. On the other hand, wafer processing from both sides allows structuring both detector faces, thus enabling a larger variability in the conceptual design of particle detectors. In particular, the same material amount of a detector layer can be exploited for measuring two (orthogonal) coordinates at the same time.

For silicon detector electrodes, selective doping of local volumes with impurity atoms, usually near the surface of the silicon bulk, is mostly done by ion implantation, rather than by the diffusion technique which is largely used in microelectronics. Ion implantation is bombardment with ions whose energy determines the depth of the implantation. Fabrication of silicon sensors is described for example in [677] or [506].



(a) pn area diode.

(b) Area diode with guard ring.

(c) Pad or pixel detector.

(d) Strip detector.

Fig. 8.35 Design concepts of single-sidedly processed pn detectors. The labels n^+ and n^-, respectively, denote strong/weak n doping, correspondingly for p doping (see also footnote on 285). Author: P. Fischer.

8.5.1 Geometries

In the following we present the most important design configurations of single-sidedly structured semiconductor detectors.

pn area diode: The simplest detector shape is the area diode which can be fabricated in sizes of a few cm^2. Typically it consists of a 300 µm thick (or thicker) silicon area, doped as shown in fig. 8.35(a). The active area usually is surrounded by a (*guard ring*) which can sink surface (leakage) currents at the edges of the diode, thus improving the noise behaviour (fig. 8.35(b)).

Pad and pixel detectors: One can subdivide the semiconductor at the surface in smaller implant areas (*pads*). If the pad sides reach dimensions of order 100 µm one speaks about pixel detectors (fig. 8.35(c)). Pixel detectors in HEP experiments have rectangular (for example 50 µm × (250)400 µm in ATLAS) or more square shaped pixels (for example 100 µm × 150 µm in CMS). Also here a guard ring protects the active detector area.

Strip detectors: If one subdivides the sensor area into strips with a typical pitch of 50–100 µm one speaks of microstrip detectors. These detectors deliver only one space coordinate of a traversing particle but are much simpler to read out than pixel

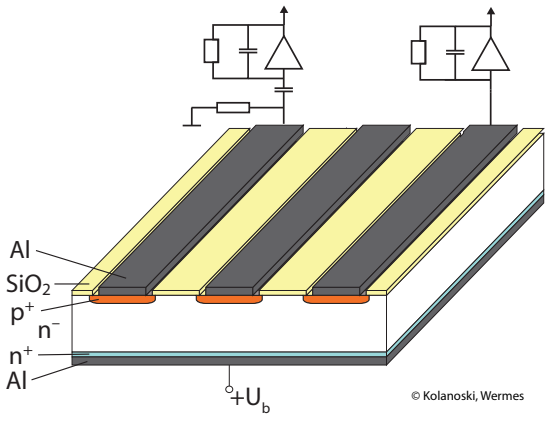


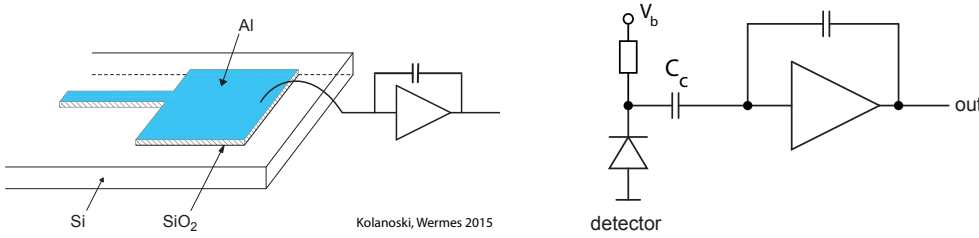
Fig. 8.36 Direct (DC, right) and capacitively coupled (AC, left) readout of a strip detector.

detectors: only N readout channels are needed compared to N^2 for pixel detectors and the connection to the chip electronics can be placed at the edge of the strip detector.

8.5.2 DC and AC coupling

Figure 8.36 shows the cross section of a microstrip detector which has a directly coupled (DC) readout on the right and a capacitively coupled (AC) readout connection on the left. The substrate silicon is weakly n doped (n^-), the strip implants on the top surface are strongly p doped (p^+). The backside contact is an ohmic semiconductor–metal coupling by means of an n^+ implant layer between the bulk silicon and the metal contact (see section 8.3.4). Above the p^+ implants there are Al strips separated by an oxide layer. For DC readout the preamplifier must take the thermally generated detector leakage current I_L and—depending on its magnitude—must compensate it by generating a current with opposite sign (leakage current compensation). For AC readout I_L is purged via the power line (here ground line); only the AC signal reaches the amplifier through the capacitance.

Principle disadvantages of a DC readout depending on the magnitude and variation of the leakage current are baseline or pedestal shifts and variations, a reduction of the range of output values (*dynamic range*) and potentially saturation effects in the electronics. AC coupling via a coupling capacitor avoids these disadvantages. For micro-structured strip detectors discrete capacitors cannot be used since the dimensions involved are too large. A suitable method is to use a thin grown SiO_2 layer



(a) Structure (bias resistor not shown). (b) Equivalent circuit.

Fig. 8.37 AC coupling in microstrip detectors.

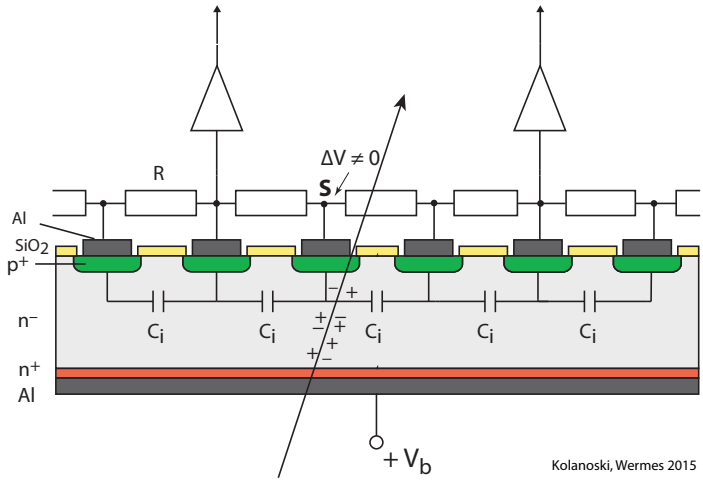


Fig. 8.38 Principle of capacitive charge division in microstrip detector readout.

between the electrode implant and the contact metal which constitutes a capacitor, which can be connected as shown in fig. 8.37 to the amplifier. The disadvantage of such a realisation of an AC coupling is the danger of so-called 'pin holes' which can lead to shorts and hence put the amplifier on DC potential. This danger grows with the oxide area.

8.5.3 Capacitive charge division

Since strip electrodes of a microstrip detector are coupled by their relative capacitance (interstrip capacitance C_i) a traversing particle generates signals not only on the 'hit' strip (i.e. the strip with maximum induction signal) but also on neighbour strips, as shown in fig. 8.38. Capacitive charge division readout exploits this interstrip coupling to reduce the number of readout channels without substantial resolution degrading. By means of the resistors the potential is kept constant in time on the electrodes. When a strip S is 'hit' as shown in the example of fig. 8.38 and Q is the collected charge on this electrode, a short potential change occurs at the strips position. This potential change appears at the amplifier inputs to the left and to the right of S by capacitive coupling $Q_L = \frac{2}{3}Q$ and $Q_R = \frac{1}{3}Q$ in the ratio of the series capacitances C_i to each side. Here only the interstrip capacitances to the nearest neighbours habe been taken into account and other capacitances (e.g. to the backside) have been neglected. Intermediate strips are not read out. From the charge division information measured on two adjacent readout strips one can interpolate the position of the particle entrance point. The (high ohmic) resistors between strips shown in fig. 8.38 are needed in order to provide the DC potential for the intermediate strips. If they were left floating, they would float up to a potential such that they would not collect any part of the signal charge, thus destroying the concept.

8.5.4 Biasing

In single-sided silicon detectors with structured electrodes only on the unstructured side a single contact is sufficient to provide bias potential. The readout electrodes, however, must each be put on a defined potential (for example ground potential). This

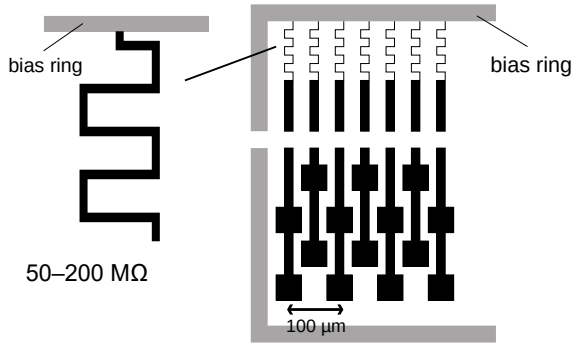


Fig. 8.39 Meander-type polysilicon resistors for biasing a silicon detectors (*polysilicon biasing).*

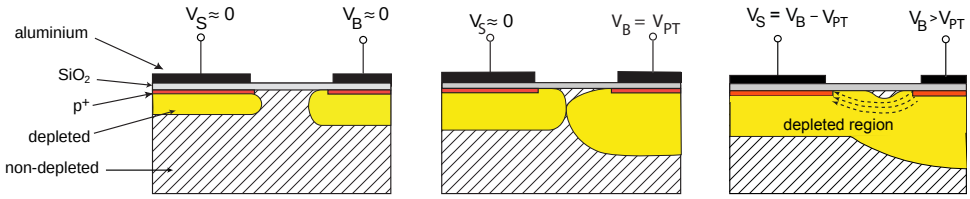
can in certain circumstances be achieved via the connected amplifier chip (virtual ground). Often, however, other techniques are needed, for example to allow for testing of sensor elements of the detectors before connection to the readout electronics is done (see also section 8.5.5). For double-sidedly structured detectors (section 8.6) explicit biasing of the microelectrodes is essential.

We discuss here some general biasing techniques for depletion of the sensor substrate in detectors with structured electrodes. The biasing potential must be applied through a high-ohmic connection in order to limit the current flow through the detector and to keep the parallel (thermal) noise contribution from this resistor low (see section 17.10.3). Different methods are used to realise high-ohmic biasing, of which we discuss the most common ones below.

Polysilicon resistors. Polycrystalline silicon grown on top of the SiO_2 has conductor properties with sheet resistances (see footnote on page 197) of up to 100 kΩ/□. The power line is applied on the Si crystal as a meander structure to make it as long as possible (fig. 8.39). Bias resistances of 50–200 MΩ can be reached this way

A disadvantage of this method is that because of the limited space available on the sensors it is not easy to obtain large resistances (for strip detectors this is still easier than for pixel detectors) and that the obtained resistances can vary from wafer to wafer. Furthermore the polysilicon application requires an additional processing step.

Punch-through biasing. The depletion zone under a (highly doped) implant also extends into the lateral direction. This is exploited in the *punch-through biasing* technique [594] shown in fig. 8.40. Contact between strip electrode (S, on the left in fig. 8.40) and bias strip or ring (on the right) is achieved by connecting them through the depleted region which grows under the bias implant with increasing voltage. Without bias voltage only the intrinsic depletion zones build up under the p^+ implants. When increasing the voltage at the bias implant the depletion zone grows into the region of the electrode strips. It touches the depletion zone of the strips at the voltage $V_B = V_{PT}$ (punch-through voltage). If the bias voltage supersedes V_{PT} the two depletion zones grow into each other and injected charge carriers can move freely in this zone. Conduction contact is obtained by *thermionic emission* from the punch-through contact. The strip's potential follows that of the bias strip less the voltage that drops over the depleted distance depending on the leakage current in this area. With further increase of the bias voltage the depletion zone also increases into the entire substrate volume up to complete depletion of the detector. The main advantage of this biasing technique over resistive biasing is the fact that much less space is needed. The technique can



(a) Detector without applied bias.

(b) Reverse bias voltage as large as *punch-through* voltage.

(c) Reverse bias voltage larger than *punch-through* voltage.

Fig. 8.40 Biasing of strip or pixel electrodes of a silicon detector by the *punch-through* biasing technique. The electrodes (e.g. strips) of the detector are put on the desired potential by applying a voltage V_B larger than V_{PT} to the punch-through contact (for example realised by another electrode strip on the outside (see text). V_S is the strip potential.

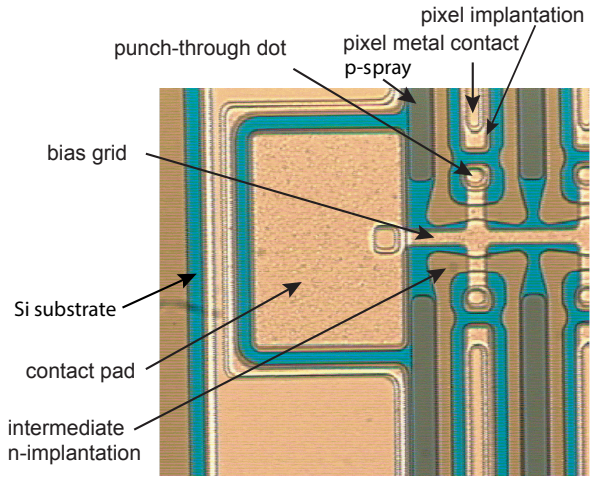


Fig. 8.41 Photo of a pixel sensor [3] (top view of an edge region). Four pixels are (partially) visible, consisting of a pixel implant, the metal contact and the punch-through-dot area. In between pixels an intermediate n-implantation and p-spray regions are visible. Bias voltage is applied by means of a *bias grid* by which the individual pixel implants receive a defined potential by employing the *punch-through* biasing technique.

also be implemented without additional processing steps.

8.5.5 Biasing of pixel detectors

For pixel detectors (see section 8.7) the question arises how a large number of pixels (of order 10^4–10^5 per module), individually isolated from each other, can be put on a definite potential. For *hybrid pixel detectors*, in which a *sensor* (e.g. a pn diode or a metallised diamond) is structured on one side only by pixels and an equally structured readout chip are different entities (see section 8.7), the bias potential is applied through a single contact from the back side of the sensor. The pixel electrodes on the structured side are bonded by microbumps to the readout chip employing the flip-chipping technology (see fig. 8.53) which establishes the connection [837] and thus obtains a defined potential. However, before the readout chip is bonded, this potential is in general not yet defined which renders testing and characterisation of the pixel sensor before assembly difficult if not impossible. By means of a *bias grid* (fig. 8.41) running in between all pixels and connecting the pixel implants by punch-through biasing, a defined potential can be applied on all, hence allowing a controlled biasing of the pixel sensor.

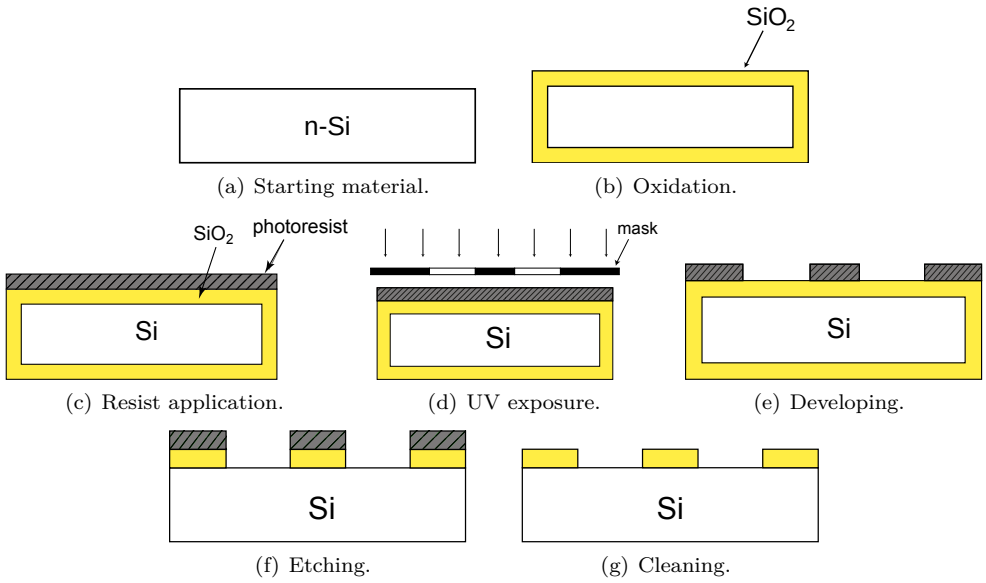
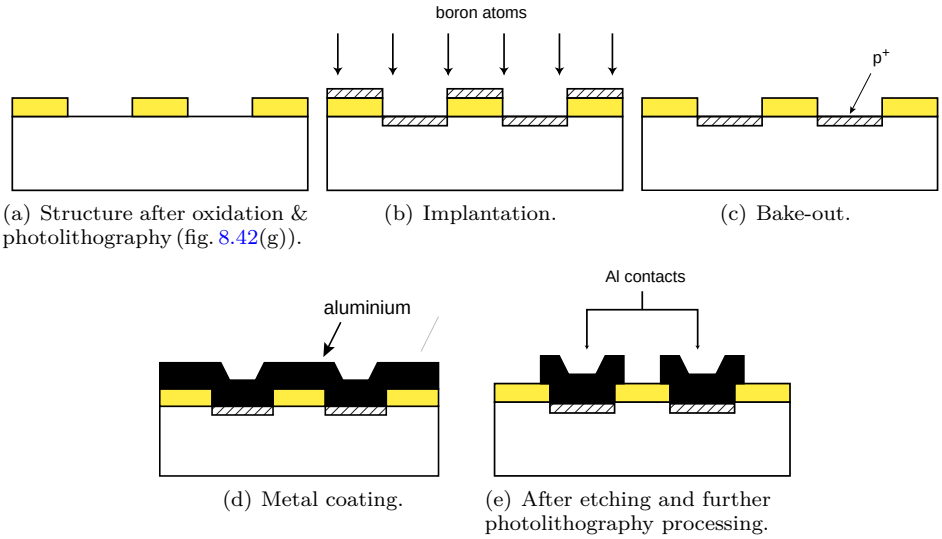


Fig. 8.42 Process steps of photolithography.

8.5.6 Structuring techniques

The microstructures needed for biasing, guard rings, signal readout and microchip fabrications on silicon wafers employ photolithography techniques. For details of this important technology we refer to the vast literature available, for example [796]. The basic processing steps for structuring semiconductor wafers by photolithography and etching techniques are rudimentarily illustrated in figs. 8.42 and 8.43 for p electrodes on n-doped substrate. The main processes employing typical process parameters are:

- The originally weakly doped silicon wafer (fig. 8.42(a)) is oxidised in a furnace at 1035 °C. An oxide layer of uniform thickness is created over the entire wafer (fig. 8.42(b)).
- In the oxide layer windows are opened by means of photolithography, through which doping should take place (fig. 8.42):
  - On the oxide layer UV-sensitive photo resist is applied (fig. 8.42(c)).
  - Through a mask with imprinted openings that have the shape of the structures to be created, the photoresist is exposed to UV light (fig. 8.42(d)).
  - During developing the exposed areas of the photoresist are removed, the oxide reappears (fig. 8.42(e)).
  - By etching the oxide layer is removed everywhere where it is not protected by the remaining photoresist (fig. 8.42(f)).
  - Thereafter the photoresist is removed again. Remaining is the wafer with the negative of the mask as an oxide layer (fig. 8.42(g)).
- The wafer now gets p^+-doped by ion implantation (fig. 8.43(b)) by bombarding with $5|, \times 10^{14}\,\mathrm{cm}^{-2}$ boron ions, accelerated to 15 keV (50 nm absorption depth, typical values).
- The material is heated to 600 °C for 30 min (fig. 8.43(c)) during which the boron atoms deposited on the oxide layer evaporate. In addition the lattice defects, which



(a) Structure after oxidation & photolithography (fig. 8.42(g)).
(b) Implantation.
(c) Bake-out.
(d) Metal coating.
(e) After etching and further photolithography processing.

Fig. 8.43 Processing steps in semiconductor structuring (here: p electrodes on n-substrate wafer).

were created by the bombardment, anneal.
– To fabricate the contacts an aluminium layer is applied by vapour deposition (fig. 8.43(d)).
– In a further photolithography process the aluminium layer is structured (fig. 8.43(e)).

Technology steps of this kind are applied among others during fabrication of CMOS readout chips (see also CMOS technology in section 17.7.4), of silicon sensor electrodes (strips, pixels, pads) and in the application of power or signal lines for detector biasing or signal readout. The most cost intensive steps are those involving UV masks.

8.5.7 Two-dimensional position information

To obtain two-dimensional (2D) position information with strip detectors one can introduce one or several crossed strip layers. A disadvantage here is that the material thickness doubles, giving rise to increased multiple scattering and secondary interactions. Double-sided strip detectors providing two coordinates with one sensor layer are discussed in the next section.

Often it is beneficial to arrange the strips of two layers not at 90° but at smaller angles. Reasons can be, for example, that all readout chips should best be arranged to one side and placed such that they are outside of the active detection area of other sensors. Furthermore, if the tracking detector is placed inside a magnetic field parallel to the beam direction (a typical arrangement for collider experiments) one can accept a somewhat worse resolution in the direction of the magnetic field (in most cases the beam direction) than in the plane perpendicular to it in which the transverse momenta are measured. In such a case one can design strip detectors with significantly smaller crossing angles (*stereo angles*). The SCT detector of the ATLAS experiment [4] is a tracking detector with four double-layers each consisting of two single-sided microstrip detectors twisted by a stereo angle of only 40 mrad with respect to each other. The construction of an SCT module with two twisted layers is shown in fig. 8.44. The small

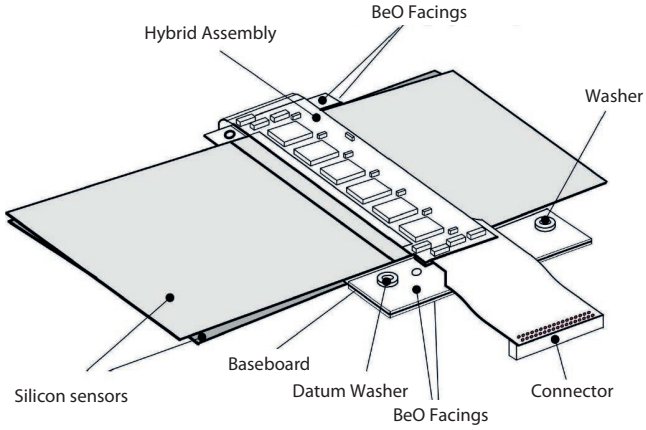
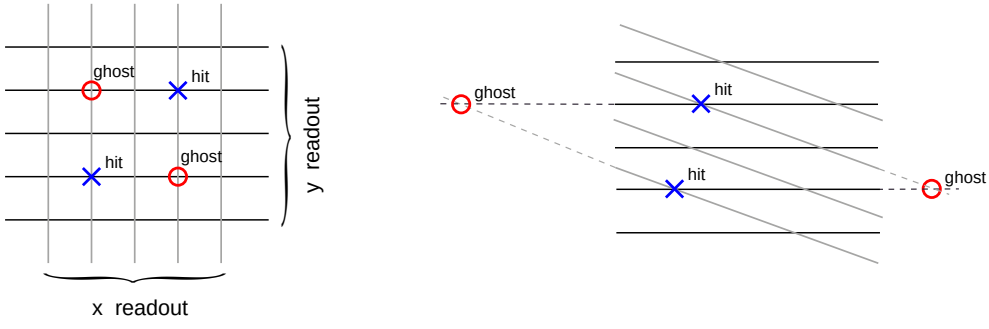


Fig. 8.44 Technical drawing of a module of the ATLAS microstrip detector SCT. Two single-sidedly structured microstrip detectors are twisted against each other by a stereo angle of 40 mrad. Particles enter the module from the top (from [4], source CERN).



(a) Orthogonal strips. (b) Small angle crossing.

Fig. 8.45 Ambiguities (ghosts) in the assignment of simultaneous hits: (a) orthogonal strips, (b) crossing under smaller angles. In (b) the ghost density is smaller.

stereo angle allows the two faces to be almost exactly on top of each other such that several modules can be aligned as a 'ladder' .

For two and even more so for several simultaneous hits, ambiguities occur in detectors with crossed strip layers, hence complicating a unique assignment of the true hit position. The ambiguities lead to effective hit occupancies per area which are much larger than for true 2D structures (pixels). Figure 8.45 illustrates the appearance of false hits (ghost hits) for an orthogonal arrangement of the strips (fig. 8.45(a)) and one with smaller stereo angles (fig. 8.45(b)). It can be seen that smaller stereo angles decrease the density of wrong hit assignments, thus simplifying the pattern recognition in track finding. See also similar stereo layouts in multiwire chambers, for example in fig. 7.42 of chapter 7.

Special track triggers (see chapter 18) profit enormously from limiting the number of readout channels. At low particle rates offline pattern recognition can still cope with the problem of multiple hit assignments. At high particle multiplicities, however, typically in cases when the hit probability of a readout electrode per readout cycle (occupancy) reaches the order of 1% of the data rate, the readout data recording can reach a limit and pattern recognition becomes more difficult or, in extreme cases, impossible. Solutions include shorter strips or generally smaller electrodes (i.e. pixels) to

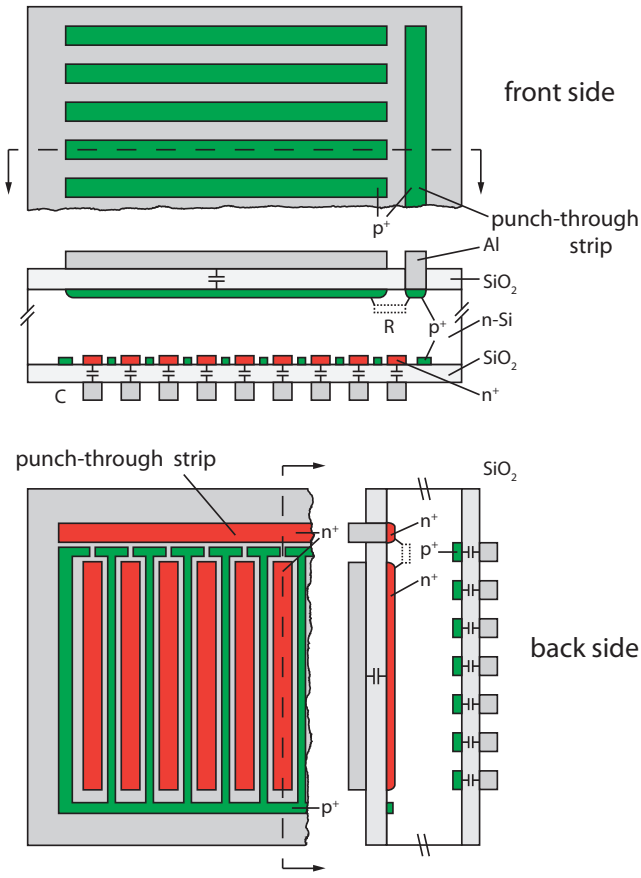


Fig. 8.46 Layout of a double-sided microstrip detector with *punch-through biasing* by means of dedicated biasing strips on both sides of the detector. The bulk is n-doped. p implants are shown in green, n implants in red. The 'front side' (diode side) is seen at the upper part of the figure in top and side view. The bottom part of the figure shows the 'back side' (non-diode side) of the detector (top and side view). The n strips are interleaved by p strips to block the electron accumulation layer (adapted from [677] with kind permission of Springer Science+Business Media).

reduce the occupancy per electrode. While this does not increase the detector material, the number of readout channels and therefore the complexity of the electronics vastly grows (see section 8.7).

8.6 Double-sided microstrip detectors

Double-sided microstrip detectors provide both coordinates of a particle passage from the same charge deposit in the detector layer. In addition the hits on both sides are correlated since they originate from the same charge deposition. This can be used to reduce ambiguities even further. Figure 8.46 shows the layout of a double-sided microstrip detector biased by dedicated *punch-through biasing* strips on both detector sides. The fabrication costs for double-sided detectors are significantly higher than for single-sided detectors because the wafers need dedicated measures for handling protection when both sides are sensitive. Besides, additional processing steps (mask steps) are necessary to prevent shorts between the strip implants on the non-diode side of the sensor (explained in the following section).

8.6.1 Special demands for double-sided detectors

Isolation of the n side. For the previously discussed single-sided strip detectors the p strips (on n substrate) are isolated from each other by the space-charge region

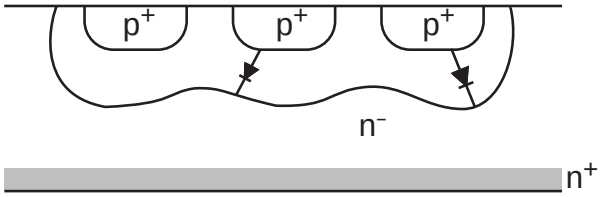


Fig. 8.47 On the 'diode side' of the detector (here p^+ in n) strip isolation is provided by the space charge region at the pn boundary.

created between strip and substrate after applying a bias voltage (see fig. 8.47). For clarity, we remain concerned here and in the following with p-in-n detectors having p-strip implants in an n substrate. For double-sided readout the question arises which implant type to use on the opposite side. With p^+ strips on both sides one would effectively operate two diodes 'against each other', resulting in a different electric field configuration with a minimum in the centre of the device.[9] Instead, by keeping the p^+-n-n^+ structure, the n side may be structured into strongly doped n^+ strips. The n^+-in-n interfaces, however, do not create a space-charge region and are hence not isolated from each other. As shown in fig. 8.48(a), under the silicon dioxide layer an electron accumulation layer develops which shortens the n^+ strips, as was discussed in section 8.3.5 on page 292. Electrons are accumulated by positive bias on the n strips and/or by positive charges, which may exist at the Si–SiO_2 interface, especially after irradiation (see also section 8.12.3). This problem does not exist on the opposite side for p strips in n bulk as these pn regions form carrier-free space-charge regions.

There are several methods to isolate the n strips against each other by interrupting the electron accumulation layer. They are illustrated in fig. 8.48:

p spray (fig. 8.48(b)) is a method in which the wafer is large-area p^--doped before the n^+ strip implant doping is applied. The accumulation of electrons is then compensated by the p doping. The p-doping concentration must be carefully tuned. If too low, rest electrons will stay at the Si–SiO_2 boundary that can lead to shorts; too strong doping causes very high field strengths at the p^-n^+ transitions. A good compromise is achieved if one chooses the doping concentration in the middle between the strips somewhat higher than at the edges (see [837] and references therein). The *p spray method* is also more cost effective than the two other methods described in the following.

p stop (fig. 8.48(c)) denotes a method in which a p^+ strip is implanted in between the n^+ strips. The p strip forms a diode together with the n substrate. The e^--accumulation layer is thereby interrupted. For the fabrication several mask steps are needed.

field plate (fig. 8.48(d)) is a method where a MOS structure is created between the n^+ strips. By applying a suitable voltage an electric field can be created in the substrate which pushes the accumulation layer away into the substrate.

Measurement of the full depletion voltage. For double-sided detectors with n^+-in-n strips on one side one can employ an elegant method to find the *full-depletion voltage* at which the detector is depleted over the entire volume (a method involving a measurement of the capacitance has been given in section 8.3.2 on page 283). By measuring the current or the resistivity between two neighbouring n^+ strips as a

[9] Such a configuration is, however, suited and used for silicon drift chambers, see section 8.8.

function of the applied bias voltage V_b, as shown in fig. 8.49, one observes a current flowing as long as the depletion region has not yet reached the n side, because there are still free carriers in the substrate near the n side. When full depletion is reached, however, a sudden drop in current occurs because the depletion region reaching the n side cuts that current off. This method allows dynamic adjustment of the required depletion bias during operation and can, for instance, also account for a compensation of charges caused by radiation damage.

8.6.2 Silicon strip detectors as vertex detectors

Silicon microstrip detectors are used in high energy physics experiments predominantly as so-called vertex detectors which enable measurements of the decay points (vertices) of long-lived particles, which have travel distances until their decay of some hundred micrometres to some millimetres. Prominent examples are b-quarks or τ-leptons with lifetimes in the range of picoseconds (see also section 8.1). Figure 8.50 shows vertex detector examples for a fixed-target experiment and for a collider experiment.

Large tracking detectors (trackers) based entirely or mainly on semiconductor detectors were first implemented in the LHC experiments ATLAS and CMS. Due to the large particle rates in proton–proton collisions conventional gaseous detectors can only be installed at significantly larger distances from the collision point. Silicon-based strip and pixel detectors can tolerate higher rates and are thus used for charged particle tracking and in particular for identification of secondary decay vertices (vertexing). Especially pixel detectors fulfil the latter task. The CMS experiment at the LHC has

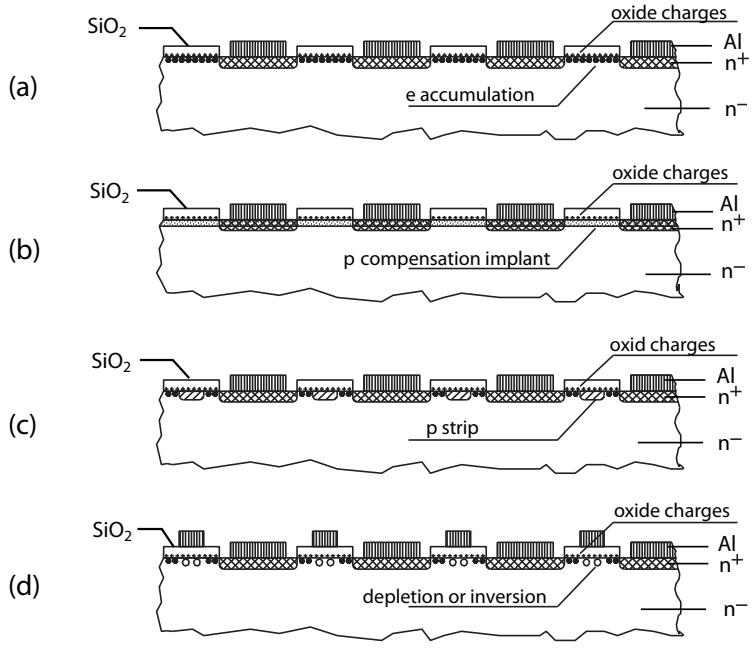
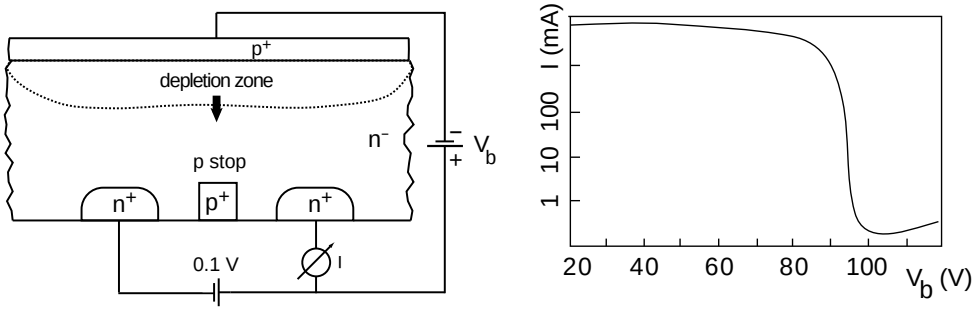


Fig. 8.48 Isolation problem and cures for n^+-in-n strips: (a) without compensation: electron accumulation; (b) p-spray isolation; (c) p-stop isolation; (d) field-plate isolation; MOS structure (adapted from [677], with kind permission of Springer Science+Business Media). Explanations given in the text.

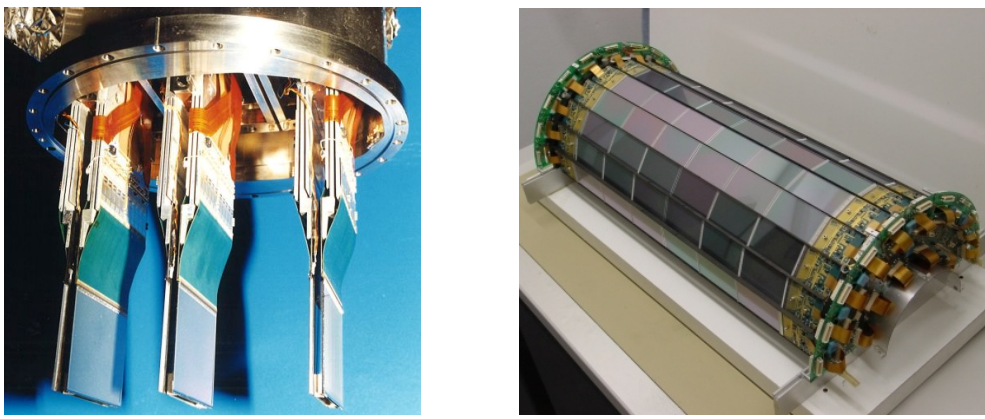


(a) Geometry of a double-sided strip detector. (b) I–V curve between neighbouring strips.

Fig. 8.49 Determination of the full depletion voltage: (a) Extension of the depletion region into the bulk with increasing bias until it reaches the electrode side; (b) current between neighbouring strips as a function of the bias voltage (adapted from [677] with kind permission of Springer Science+Business Media).

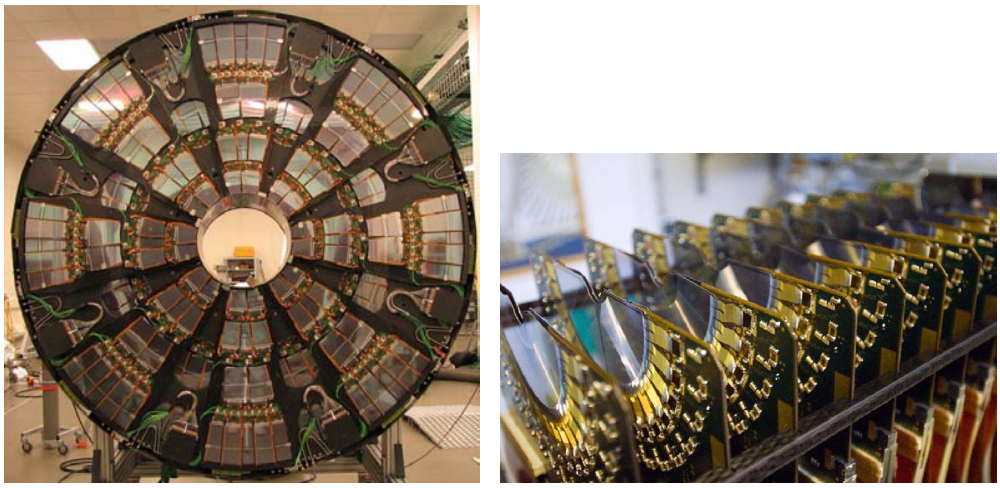
a strip tracker covering a total area of almost 200 m^2 [298]. Figure 8.51(a) shows the arrangement of the detector modules in the endcaps (TEC) of the strip detector. The LHCb experiment is specialised to detect long-lived bottom hadrons producing secondary vertices at a distance of a few millimetres from the production point. For this a dedicated microstrip vertex detector (VELO) has been installed which approaches the beam down to 7 mm [87, 760] (fig. 8.51(b)).

The different sizes of various silicon detectors used in collider experiments is illustrated in fig. 8.52. In the beginnings they were only employed as vertex detectors, in later times—especially at the LHC—silicon strip and pixel detectors covered large



(a) Fixed-target experiment (HERA-B). (b) Collider experiment (H1).

Fig. 8.50 Examples of microstrip detectors used as vertex detectors: (a) Si-strip detectors in the HERA-B experiment for precise identification of secondary vertices [161]. The detectors were placed only 10 mm away from the p beam which in the figure is running horizontally near the detectors. DESY/HERA-B Collaboration (photo: Iris Abt). (b) Double-layered Si vertex detector of the H1 experiment at HERA whose innermost layer was operated 5.75 cm away from the beam line [789]. Source: DESY/H1 Collaboration.



(a) CMS tracker endcap (TEC). (b) VELO detector (LHCb).

Fig. 8.51 Microstrip detectors in LHC experiments: (a) The endcap (TEC) of the CMS microstrip tracker has nine layers subdivided in trapezoidal sectors (petals) containing overlapping strip modules on front- and backside [298] (source: CERN/CMS Collaboration); (b) Si-strip detectors in the *Vertex Locator* (VELO) of the LHCb experiment [760, 87] for precise measurement of secondary vertices (source: CERN/LHCb Collaboration).

parts of the tracking volume also for track reconstruction and momentum measurement.

8.7 Hybrid pixel detectors

Pixel detectors [837] allow for true three-dimensional space-point determination even at high particle fluxes. At the LHC particle fluxes of the order of 10^{11} particles per second traverse the tracker volume. For a (design) luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ about 1200 particles traverse the tracking detector for every beam crossing occurring every 25 ns. The occupation density (*occupancy*) in the detectors near the interaction point is correspondingly high and for the reconstruction of the collision products a number of independent detector layers are needed. Up to radial distances of about 15 cm pixel detectors therefore are the instrument of choice with the pixels having typical dimensions of order $100\,\mu\mathrm{m} \times 100\,\mu\mathrm{m}$ in square or rectangular shape (e.g. $50\,\mu\mathrm{m} \times 250\,\mu\mathrm{m}$ [19]).

So-called *hybrid* pixel detectors are made of two parts (in contrast to monolithic pixels, section 8.10) which exactly match: the pixel sensor, a silicon diode structured in pixel cells as in fig. 8.35(c), and one or several readout chips with the same cell pattern. Sensor and chip are connected in every pixel by a conducting microconnection (*bump bond*). Figure 8.53(a) shows the principle layout of an individual pixel cell, fig. 8.53(b) that of a matrix with many cells.

The entire pixel detector is constructed in a modular arrangement. Figure 8.54 shows the layout of a module (an ATLAS module [3] as a typical example). The module consists of the silicon sensor in which traversing particles generate charges and 16 readout chips bonded to the sensor which amplify and discriminate the signal. Upon a hit the module generates chip number as well as row and column address of the hit plus a time mark (bunch-crossing number). The chips store this information until

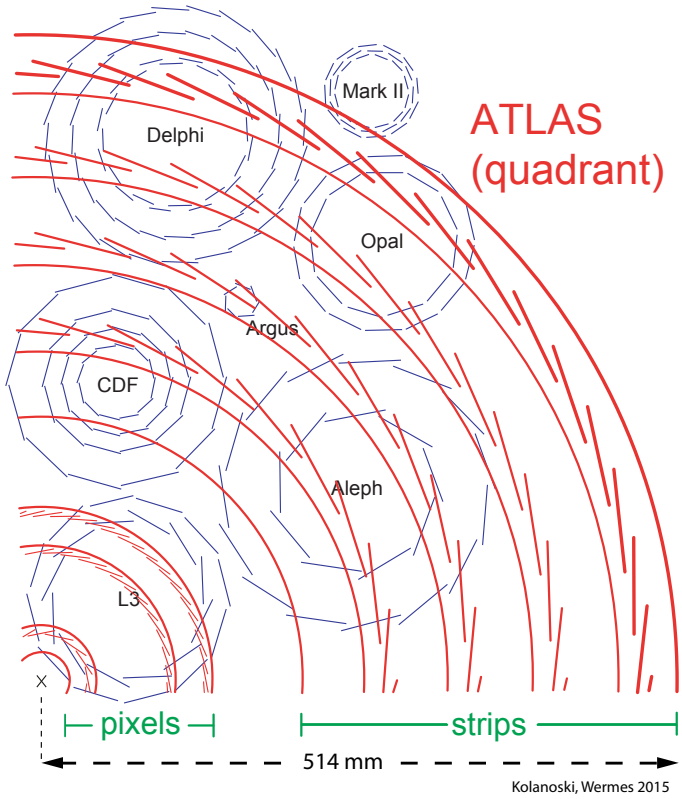
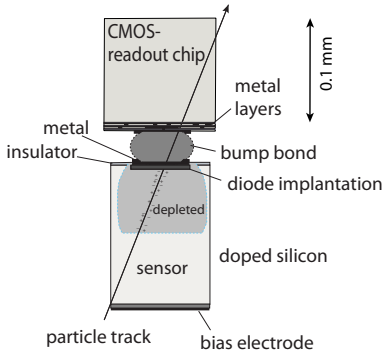
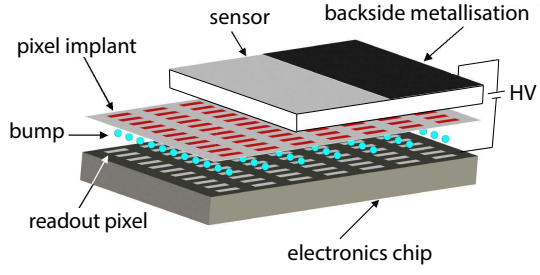


Fig. 8.52 Size comparison of various silicon detectors that have been installed in present and past collider experiments. The various smaller (vertex) detectors (all strip detectors) are compared with the size of one quarter of the ATLAS tracker (in red) consisting of pixels and strips. Not shown here is the semiconductor tracker of the CMS experiment [754, 298] consisting of silicon microstrip and pixel detectors with an outer radius of 1.25 m.



(a) Hybrid pixel cell.



(b) Pixel matrix.

Fig. 8.53 Hybride pixel detector: (a) Layout of an individual pixel cell composed of sensor and electronics cell, (b) hybrid pixel matrix. Sensor and electronics chip are subdivided into pixels of the same size, which are connected by *bump* contacts. The pixel implants in (b) are drawn separately from the Si sensor for better illustration, different to reality.

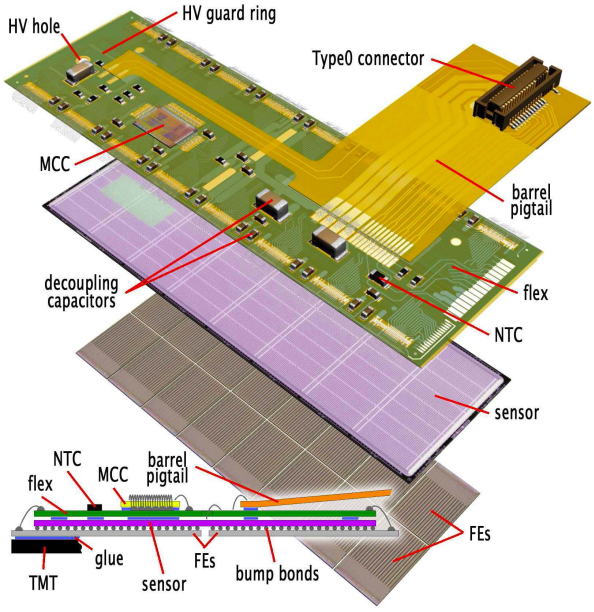


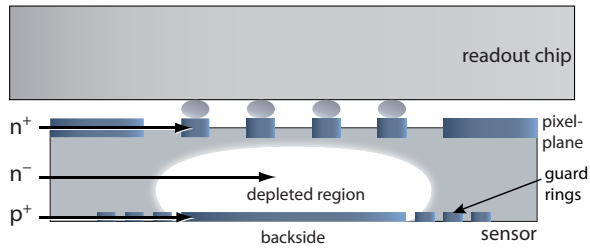
Fig. 8.54 Assembly details of an ATLAS pixel modules [3], perspective view and cross section (bottom). The module contains three layers: chip layer (bottom), sensor layer (middle), and a distribution layer (top). Chip and sensor layer are connected by employing bump and flip-chip technology. A thin, kapton-based flex circuit layer is glued to the sensor backside for connection and redistribution of the signals. A further flex circuit (pig tail, very top) connects the module with external cables. The dimension are 6.08 × 1.64 cm^2 (active area). (Source: CERN/ATLAS).

the event trigger arrives (up to 120 collisions later), upon which the buffers are either read out or otherwise cleared. The readout of pixel detectors is described in chapter 17. More detailed literature on pixel detectors is found for example in [837, 3, 442].

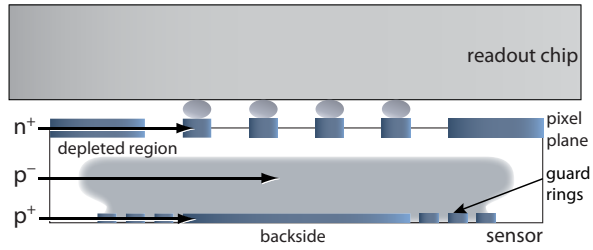
Pixel and strip detectors under high particle flux. At the position of the innermost detectors near the interaction point at the LHC, particle fluxes of up to 10^{14} particles per cm^{-2} and year occur. The time integrated particle flux is called *fluence* measured in number of particles per area. After a fluence of about 2–3 × 10^{12} cm^{-2} the silicon sensors (with substrate doping of $N_D \approx 2 \times 10^{12}$ cm^{-3}) face a so-called 'type inversion', meaning that the originally n-doped silicon becomes effectively p doped. The effective charge carrier concentration (originally $n \approx N_D$) and also the voltage necessary for full depletion, first decrease until the 'inversion point' is reached and then increase again thereafter with increasing fluence. This is shown in fig. 8.85 in section 8.12. The p^+n^- boundary (diode) of a p^+–n^-–n^+ pixel detector[10] hence moves to the opposite side after change of the charge sign of the space-charge region (from p^+–n^-–n^+ to p^+–p^-–n^+).

In order to keep the LHC pixel detectors operative after radiation damage and type inversion, they are designed as n^+–n^-–p^+ sensors ('n^+-in-n' sensors) which turn into n^+–p^-–p^+ structures after type inversion [837, 3]. The advantage of such an arrangement is that after irradiation with high fluence the sensor's depletion region grows from the pixel electrode side into the sensor (fig. 8.55). Hence operation is still possible even if full depletion can no longer be achieved within the provided bias voltage reach. Bias voltage is applied on the 'backside' (opposite to the pixel side). Guard rings bring the potential down toward the edges of the sensor to avoid high voltages reaching the cutting edges, which generally are conducting due to unavoidable damage of the atomic layer-structure from the cut, and would hence bring the potential to the

[10] p^+–n^-–n^+ denotes the sequence of doping layers seen from the electrode side of the sensor.



(a) n^- substrate doping before irradiation.



(b) p^- substrate doping after irradiation.

Fig. 8.55 n^- to p^- type inversion for n^+-in-n^- pixel sensors: (a) before irradiation, (b) after irradiation above 3×10^{12} particles per cm^2 (type inversion point for typical initial doping). The guard rings on the 'backside' (opposite to electrodes) bring the potential down to ground potential at the edge.

sensitive amplifier inputs on the pixel side.

Isolation of n^+-in-n pixels is achieved by employing the *p-spray* technique (see section 8.6) with an optimised doping profile [3] having a somewhat weaker doping at the edges. This isolation technique features lower field strengths at the pixel-to-substrate transitions than alternatives (e.g. p stop) and is hence more robust with respect to breakdown.

To be able to test the pixel sensor part, normally the readout chips must have been bonded to them in order to provide a defined potential for each pixel through the (virtual) ground of the amplifier (see section 17.2.1 on page 719). This is very problematic for pixel module production because, if the sensor must be rejected, all readout chips are lost as well. In oder to provide a fixed bias potential to every pixel, therefore, a special *bias grid* structure is distributed over all pixels (see fig. 8.41). By means of the punch-through technique the bias potential is put through to every pixel (see section 8.5.5).

Pixel detectors can tolerate higher irradiations than strip detectors because they are more finely segmented. Among other effects radiation damage leads to increased leakage current (see page 352 of section 8.12.1) which increases noise and finally also reduces the detection efficiency. The size of the leakage current I_L is proportional to the volume of the sensor and is distributed over all electrodes. In pixel detectors the leakage current is therefore shared by comparatively many electrodes (their preamplifiers) and hence is small per electrode compared to strip detectors. Figure 8.56 shows the cell layout of the ATLAS pixel sensors as well as the detector response to minimum-ionising particles after irradiation by a fluence of more than 10^{15} particles per cm^2. One recognises a rather homogeneous charge collection over the entire area with the

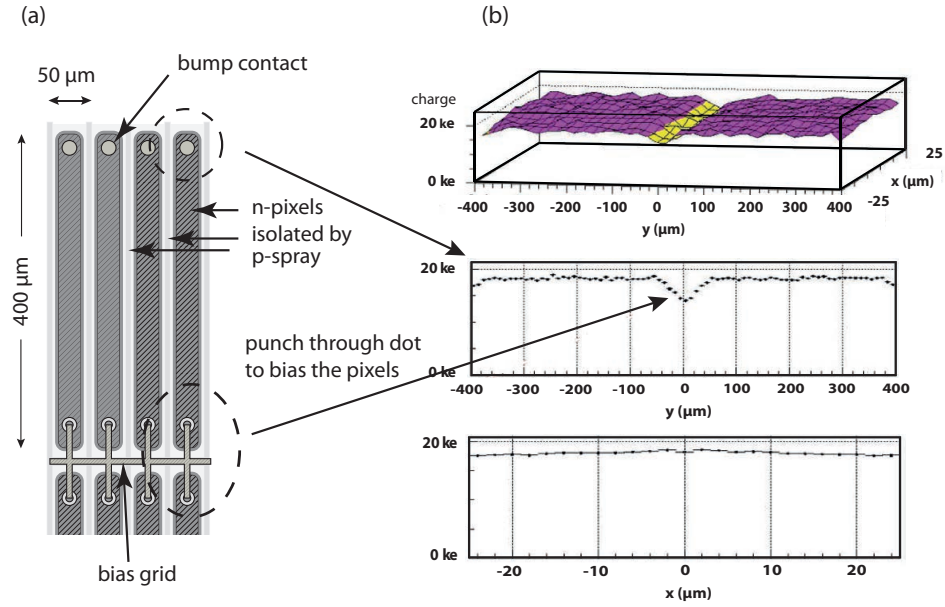


Fig. 8.56 (a) Layout of (a variant of) ATLAS pixel sensor cells (n^+ in n) [3]. The pixel cells are biased via the *bias grid* employing the *punch-through technique* in between the punch-through dot and the pixel implant. The sensor is connected to the readout chip via micro-bump contacts. (b) Measured charge (in units of thousand electrons, ke) from perpendicularly traversing particles in a double pixel cell (two neighbouring head-on mirrored cells with the bias grid being the mirror axis in the middle) after irradiation by 2×10^{15} protons/cm^2 corresponding to about 600 kGy. Shown is the detected charge in the cell when a minimum-ionising particle traverses the sensor (thickness 250 µm) in 2D representation (top) and in projected views (centre and bottom). The regions of charge loss are indicated by arrows. Source CERN/ATLAS.

exception of a region between the pixels where the bias grid, described in section 8.5.5, is located, and at the edges where the bump contacts are. But even in these regions the detector is still sufficiently sensitive, that is, the signal is still much larger than the noise which typically is in the order of 100 e^-.

8.8 The silicon drift chamber

For wire chambers the concept of exploiting the drift-time information in drift chambers (see section 7.10) for space-point reconstruction brought many advantages compared to the conventional wire detection method. The same technique can also be employed in semiconductor detectors, as will be discussed in this section. We will see that there are indeed rewarding applications for this principle, but the general gain over conventional readout of microstrip or pixel detectors without time information is much reduced compared to wire chambers.

8.8.1 Sidewards depletion

The realisation of a drift time measurement requires a technique called *sideward depletion*. On an n^--doped substrate p^+-doped areas (linear strips or rings) are implanted

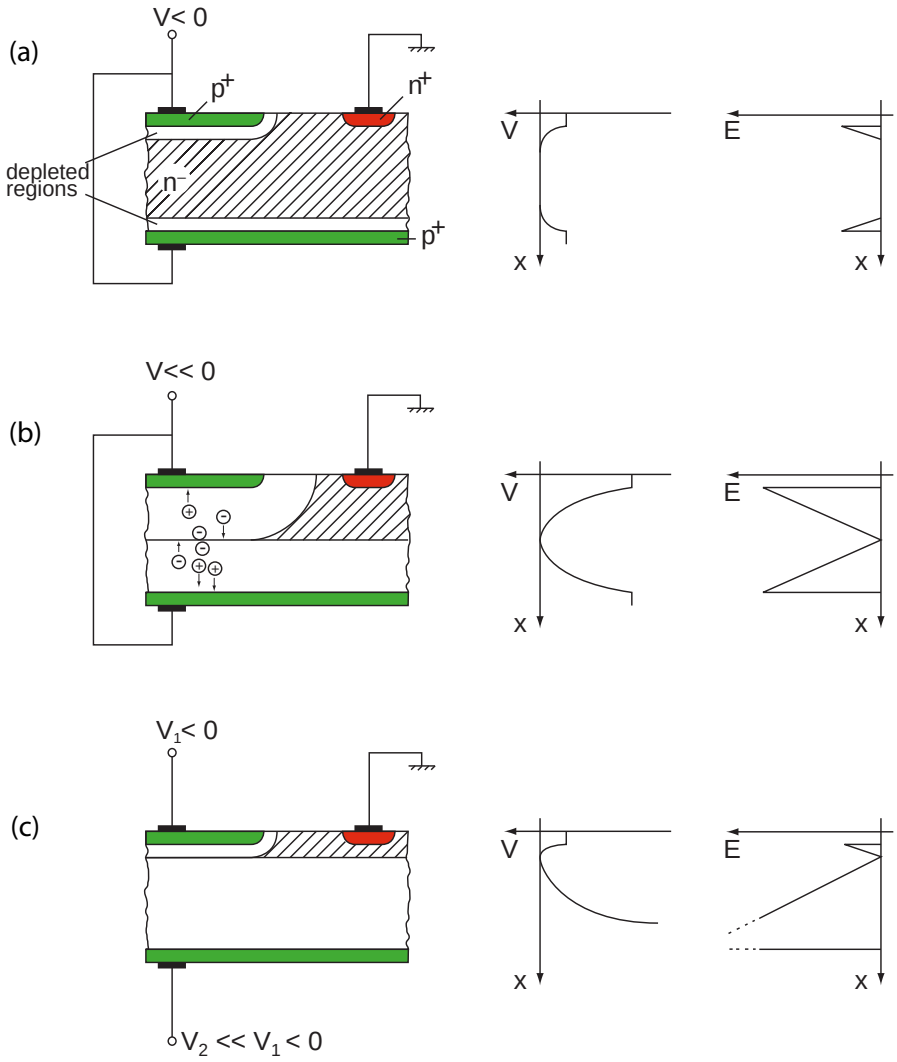
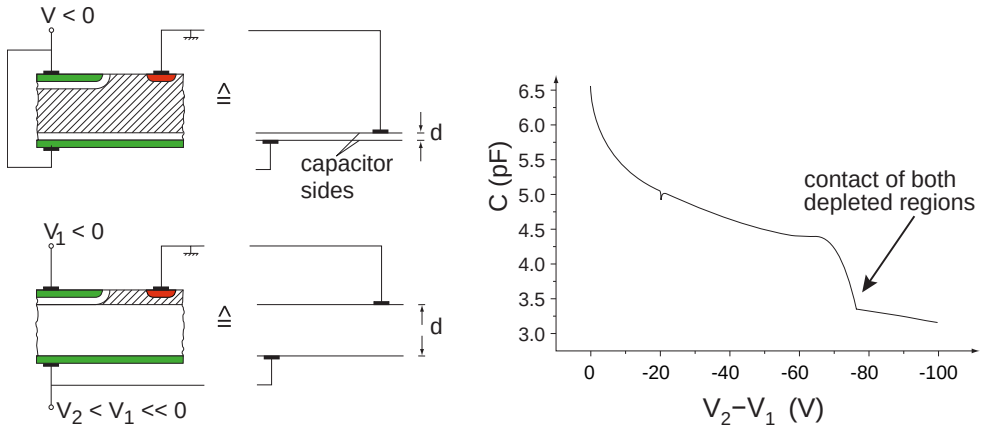


Fig. 8.57 Principle of sideward depletion. (a) $V < 0$, the sensor is not yet depleted, (b) $V \ll 0$, the depletion zones touch, (c) $V_1 < 0$, $V_2 \ll V_1 < 0$, shift of the potential minimum (for electrons) underneath the top surface.

on both sides (top and bottom) of the substrate (fig. 8.57(a)). In addition on one side an isolated n^+-implant region is created. If one applies negative bias on both p^+-regions, two depleted zones are created inside the semiconductor, each starting at the respective p^+–n^- boundaries on either side. Upon further increase of the reverse bias voltage (fig. 8.57(b)) both carrier-free regions touch. A situation is created as if two diodes (np–pn) were merged. The electric field linearly decreases from both sides towards the centre of the substrate. The potential is parabolic with a minimum (for electrons) in the centre.

When impinging radiation generates electron–hole pairs, the positive charge carriers drift towards the p^+-electrodes on the sides, while the electrons drift towards the



(a) Capacitance measurement. (b) C-V diagram for sidewards depletion.

Fig. 8.58 Principle of the capacitance measurement with sideward depletion: (a) measurement set-up, (b) C-V curve [727].

potential minimum in the middle of the semiconductor. Applying bias potentials of different height on both sides, as shown in fig. 8.57(c), the position of the minimum can be shifted inside the semiconductor substrate.

To determine the voltage necessary for full depletion of the detector the capacitance between the n^+-contact and the opposite side of the detector (backside) is measured. Figure 8.58(a) displays the method: the capacitance depends on the width of the depletion region and can be regulated by the applied voltage. This dependence is shown in the capacitance–voltage diagram (C–V diagram) of fig. 8.58(b). The kink at $-75\,\mathrm{V}$ is evidence that both depletion regions are in contact.

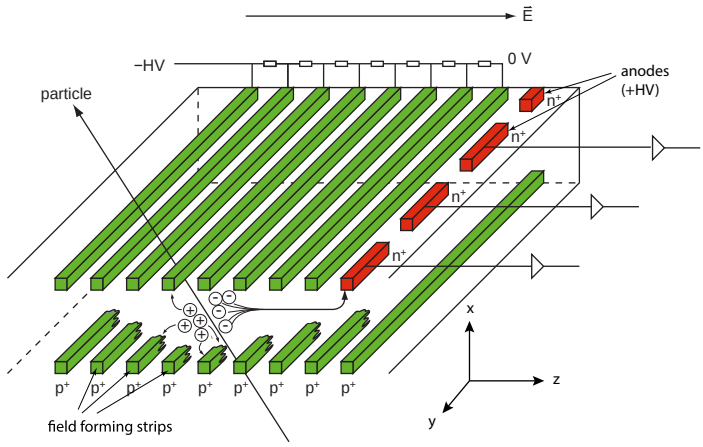
8.8.2 Silicon drift chamber

Based on the sideward depletion principle the *silicon drift chamber* (SiDC) has been developed [450, 811]. Figure 8.59(a) shows an illustration of the concept. The n-type substrate has strip-like p implants on top and bottom sides with linearly increasing potentials applied. The resulting linear field leads to the electrons accumulating in the potential minimum towards segmented n-type anodes. We choose as before the x coordinate for the detector depth. The electron drift direction is the z coordinate. In the y direction we assume the detector to be infinitely extended. The potential in y is constant for a given x–z. The potential in the x direction can be described by the one-dimensional Poisson equation:

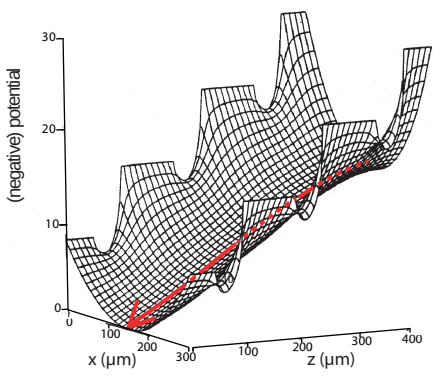
$$\frac{\partial^2\phi}{\partial x^2} = -\frac{\rho}{\epsilon_0\epsilon_r}\,, \tag{8.86}$$

where ρ is the space charge density inside of the sensor. Respecting the boundary conditions $\phi(x=-\frac{d}{2}) = \phi(x=\frac{d}{2}) = V_0(z)$ and for a detector thickness d and a bias voltage $V_0(z)$ the solution along the detector depth for full depletion is:

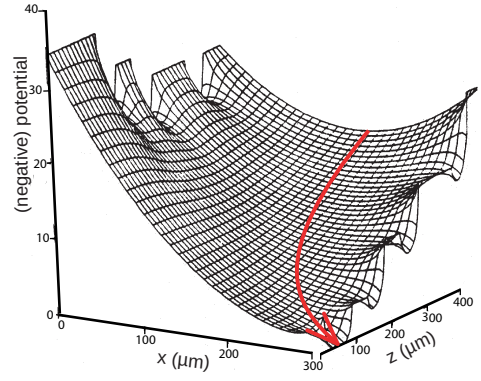
$$\phi(x,z) \;=\; V_0(z) - \frac{\rho}{2\epsilon_0\epsilon_r}\left(x^2 - \frac{d^2}{4}\right)\,, \tag{8.87}$$



(a) Silicon drift chamber principle.



(b) Region of the field forming strips.



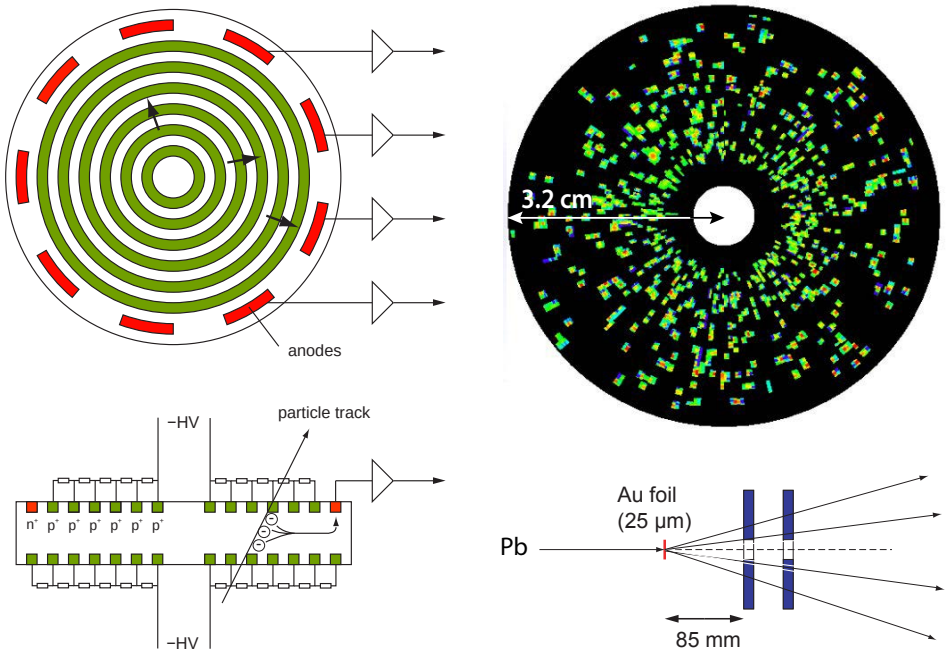
(c) Anode region.

Fig. 8.59 (a) Layout of a silicon drift chamber; (b,c) shape of the electric potential in silicon drift chamber (adapted from [920,917] with kind permission of Elsevier). Due to the potential gradient the electrons drift towards the anode (b) where they are collected (c).

where $V_0(z)$ can be z-dependent to superimpose a linear potential in this direction. The resulting potential is parabolic.

Figure 8.59(b) displays the shape of the potential in the drift region and fig. 8.59(c) in the region near the anodes; also illustrated is the line of the electron drift following the potential gradient and deflected in the anode region. Negative charge carriers generated in the detector volume are collected in the minimum of the parabola; holes drift towards the surface electrodes. By superposition of a potential slope in the z direction by means of a voltage divider chain, providing linearly decreasing voltages at the p strips in fig. 8.59(a), a constant drift field is created guiding the electrons collected in the minimum towards the readout anodes. The coordinates of the origin of the charge generation are obtained by measuring the drift time (z direction) and in the y direction by segmented anodes.

Obtainable space resolutions are in the range 20–40 μm in the drift direction and about 1/5 of the anode distance in the orthogonal direction (typically <50 μm). Challenging for a precise drift time calibration is the strong temperature dependence of



(a) Layout principle. (b) Recording of a Pb–Au reaction.

Fig. 8.60 Cylindrical silicon drift chamber (SiDC) [301]: (a) Layout of the SiDC ($r = 3.2\,\text{cm}$). (b) Hits of particle tracks of a heavy-ion reaction (with kind permission of the CERES/NA45 Collaboration).

the electron mobility in silicon. Also, the precise setting of the drift field by means of integrated resistors (voltage divider chain) requires great care.

The advantages of silicon drift chambers compared to other semiconductor detectors are for one the comparatively small number of readout channels needed for a truly 2D space point measurement. Furthermore, the collection electrodes (anodes) can be made geometrically very small, with small capacitance, leading to good signal-to-noise ratio. Due to comparatively long electron collection times (typically in the μs range) silicon drift chambers are not suited for high rate applications. The good 2D space resolution (typically about $20 \times 20\ \mu\text{m}^2$) and the good separation of close-by particles ($< 50\,\mu\text{m}$ [311]) is, however, excellent for resolving events with high particle multiplicities. Typical applications are heavy-ion reactions, see e.g. [301].

Besides the linear layout shown in fig. 8.59(a) a cylindrical layout can be chosen [301], as shown in fig. 8.60(a). On an n^--doped Si disk p^+ rings are implemented. The anodes can be realised on the outside as in fig. 8.60(a) or alternatively in the centre of the disk as in fig. 8.61, depending on whether spatially resolved particle detection (anodes at the outside) or spectroscopic application (anode in the centre) is the focus of the application.

The variant with a central anode [595] (fig. 8.61) offers the unique combination of a large collection area for radiation and a small collection capacitance, owing to the geometrically small dimensions of the anode leading to correspondingly excellent noise properties (see section 17.10.3). In [646] energy resolutions of 152 eV at 6 keV deposited energy (^{55}Fe source) at –20 °C have been obtained with diodes having active areas of

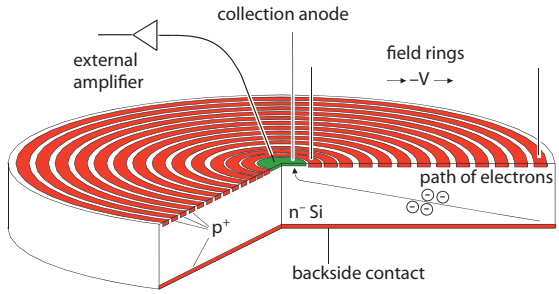


Fig. 8.61 Cylindrical silicon drift chamber (silicon-drift diode, SDD) featuring only one central anode for spectroscopic applications (adapted from [646] with kind permission of Elsevier).

several square centimeters.

8.9 Charge-coupled devices (CCDs)

Charge-coupled devices (CCDs) were invented in 1970 [242] and have been used for decades as optical sensors for media devices, for example as video cameras (CCD cameras). In recent years they have been largely replaced by CMOS cameras.

8.9.1 MOS CCDs

Figure 8.62 illustrates the CCD principle. Charge created in an 'active' layer (space charge region) is first stored locally in a potential minimum and is then transported to a readout node (see fig. 8.64). In MOS CCDs the potential minimum is created by a MOS structure with metal contacts as gates on the SiO_2 layer on top of the silicon substrate. In conventional MOS CCDs minority carriers (e.g. electrons in p-Si) are collected and transported to the output by moving them in the boundary layer between oxide and semiconductor bulk (see below). In more advanced techniques, among others for example in so-called *buried-channel* CCDs, storage and transfer is done in a somewhat deeper layer. We refer to the dedicated literature, for example [534, 929], for details.

In the standard CCD an epitaxial[11] layer is grown on a (usually p-doped) substrate (epi-Si) as a photoactive layer. On the top this layer is covered by an oxide layer on which metal electrodes (*gates*) are placed (fig. 8.62). Epitaxial silicon is used in chip electronics predominantly in optical applications as an absorption layer for light. The doping profile in the epitaxial layer can be well controlled and epi-Si can be deposited

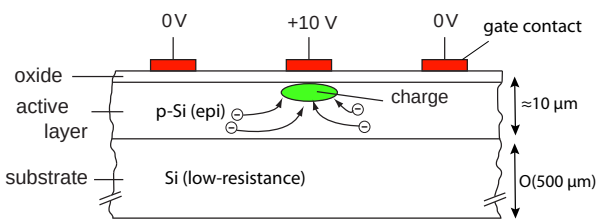


Fig. 8.62 Principal layout of a CCD cell. Charge (here electrons), which has been created by light absorption in the epi-layer under the oxide, is collected under the positive potential of the gate contact and from there is transported to the right (or left) by appropriate switching of the gate potentials.

[11] Epitaxial silicon is a silicon layer grown on a crystalline substrate by keeping the substrate's crystallographic orientation.

out of a gaseous atmosphere largely free of impurities such as oxygen or carbon. Hence it can be made high-ohmic, resulting in a higher mobility of the charge carriers created by the incident light.

The active epi layer is very thin, typically 10 µm, maximally about 30 µm. The layer is depleted by the potentials between the backside contact of the substrate and a diode strip on the top. The gate electrodes are arranged as strips similar to those of a microstrip detector. For the passage of a high energetic particle in the thin epi layer, comparatively few (typically some hundred) charge carriers are generated and collected under the electrodes. For the detection of individual charged particles with MOS CCDs high demands are hence imposed on the noise properties of the readout chain. For light (many photons), however, the charge carrier yield generally is large enough to obtain a sufficient signal that is processed further. A suitable voltage configuration applied to the gate electrodes keeps the charge carriers in the active layer underneath the oxide surface and enables their controlled movement.

Every third electrode is put on the same positive potential and the bulk is at negative ($-V_{\text{bias}}$) potential thus creating local energy minima for electrons at the Si–SiO_2 interface below the electrodes with the highest potential. For signal readout the collected charge carriers are moved stepwise to a readout electrode by applying appropriate clocking of the electrode potentials ('bucket chains'), as illustrated in fig. 8.63: (a) under one electrode (10 V) charge has been collected in a potential well. The well is bounded by the neighbouring electrodes carrying a lower voltage (2 V). In (b) the neighbour electrode to the right is set from 2 V to 10 V, the potential well hence extends and the charge carriers flow to the right (c). The potential well now has a larger geometrical extension. In (d) the electrode formerly having had a voltage of 10 V is now set to 2 V; the electrons now flow completely into the right part of the well. In (e), finally, all electrons are under the right electrode, the charges have been moved by the on gate position to the right. Figure 8.63(f) shows the clock pulses of the three-phase clocking mechanism.

Readout of 2D space information can thus be realised with several bucket chains and only one readout node, as illustrated in fig. 8.64. The charge carriers are moved down row-wise. Thereafter the last row is moved column-wise towards the readout node.

A faster, but circuit-wise more involved readout scheme operates in parallel, that is, in every row or column there is a readout electrode. The spatial resolution for charged particle detection is given by the size of the CCD cells. They are typically in the range 10–20 µm such that spatial resolutions in the order of a few micrometres can be obtained.

The clocking of CCDs with clock frequencies in the MHz range is possible with very good efficiency (charge loss is at the 10^{-5} level). The entire readout process, however, is comparatively slow. Another problem is caused by so-called ghost hits appearing when the detector is hit before readout is finished. One can eliminate ghost hits by copying all hit data in special storage cells assigned to every CCD cell (so-called vertical shift registers). For applications demanding high recording rates as is often the case for particle physics experiments, CCDs can thus usually not be used.

An exceptional situation existed for the vertex detector for the SLD experiment at the SLC accelerator at Stanford, USA (see table 2.2) [142, 338, 337] where beam conditions and demands were perfect for CCD detectors. Here the CCD technology was chosen in competition with microstrip detectors since CCDs have a better spatial

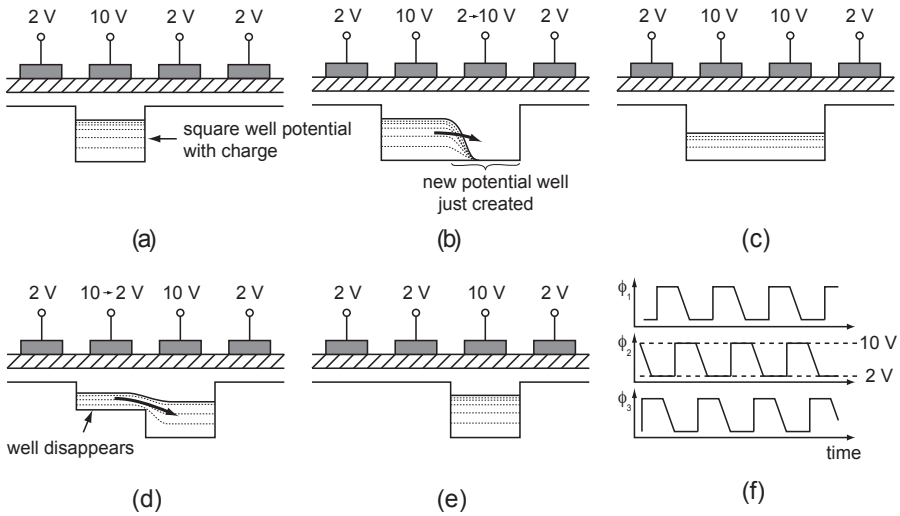


Fig. 8.63 Bucket-chain principle of the readout of a CCD detector (adapted from [192] [608] [837]). The sequence of operation from (a) to (e) is explained in the text. In (f) the asymmetric clock pulses of the three-phase clocking mechanism are shown.

resolution (cell size 22×22 μm^2, space resolution ≈5 μm) and because at the SLC there was sufficient time for readout between two successive collisions (160 ms). In CCD detectors very low amounts of charge ($\lesssim 1600$ e$^-$) are deposited in the epi-Si layer which is only about 20 μm thick. The noise of a CCD system, however, is with ∼30 e$^-$ also very low. Cooling to about –80 °C is needed to cope with leakage current generation (see section 8.12.1 on page 350). Figure 8.65 shows the two half-shells of the SLD vertex detector using MOS CCDs. It was installed and operated immediately outside of the beam pipe ($r = 13$ mm).

For experiments at colliders such as LEP, HERA, Tevatron, or even LHC, CCD detectors are much too slow. Table 8.3 compares some properties of MOS CCDs as particle detectors with microstrip and hybrid pixel detectors.

8.9.2 pnCCDs

The active layer of MOS CCDs is very thin. Therefore for traversing high energy particles the signal charge is rather small, as is the absorption probability for X-rays in this layer. By contrast pnCCDs [705] are built up on a much thicker, high ohmic and fully depleted substrate (fig. 8.66). As for the silicon drift chamber the layout is based on the principle of sideward depletion described in section 8.8. Different to MOS CCDs the substrate is high-ohmic and fully depleted from the top and bottom contacts. The electrodes are p$^+$n transitions. As for MOS CCDs the potential minimum is moved by shifting voltage levels at the top electrodes the potential minimum near the detector surface is moved and with it the charge. Besides the advantageous charge collection over the entire depleted substrate another advantage is that low energy ($\lesssim 10$ keV) X-ray photons can be detected with spatially homogeneous sensitivity by entering through a thin, large area p$^+$ entrance window on the backside. The pnCCD technique has therefore been applied for low energy X-ray detection as for example in the X-ray satellite XMM-Newton [919].

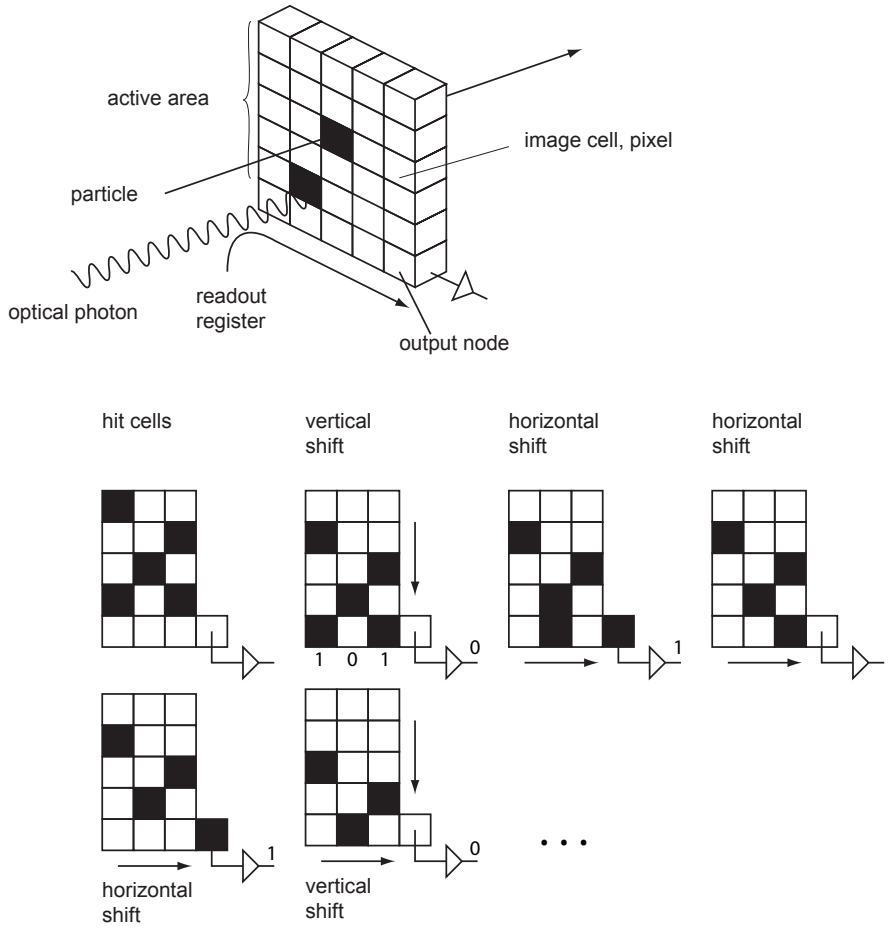


Fig. 8.64 Readout principle of a CCD matrix. The charge is first moved in the vertical direction, parallel in all cells; the bottom row is then clocked horizontally towards the readout node.

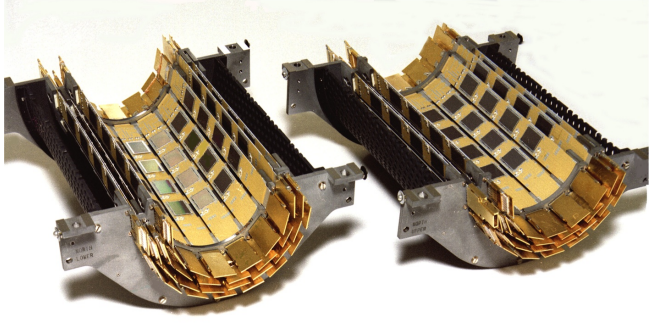


Fig. 8.65 The two half shells of the CCD-based vertex detector of the SLD experiment [338]. The detector consisted of four concentric layers and contained about 250 CCD detectors. Source: SLAC/SLD Collaboration.

Table 8.3 Some parameters and characteristics of microstrip, pixel and CCD detectors in comparison.

	Microstrips	(Hybrid) pixels	(MOS) CCD
typ. thickness	300 µm	250 µm	15 µm active + (< 100 µm) passive
typ. electrode pitch	50–100 µm	50–100 µm	10 µm
typ. module area	20 cm^2	10 cm^2	4 cm^2
energy loss (mip)	120 keV	100 keV	3 keV
noise equivalent	8 keV	1 keV	0.21 keV
signal : noise (mip)	> 10:1	100:1	15:1
dimensionality	max. 2×1 dim.	2 dim.	2 dim.
spatial resolution	5–10 µm	10–15 µm	2–5 µm
double-hit resolution	40–60 µm	50–100 µm	20–40 µm
typ. module readout	fast	fast	slow
readout time (size dep.)	µs	µs	ms

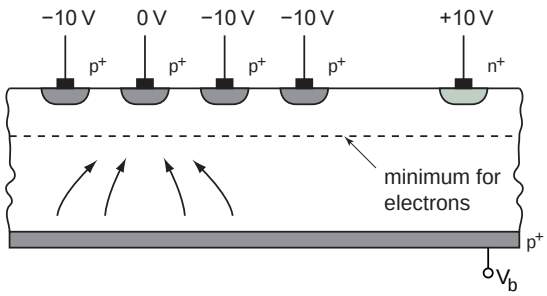


Fig. 8.66 In the pnCCD signal electrons are collected in a potential minimum and are guided by external electrodes.

8.10 Monolithic pixel detectors

8.10.1 Hybrid versus monolithic pixel detectors

The pixel detectors introduced in section 8.7 are hybrid structures for which the 'active' sensor and the 'passive' readout chip are separate components of the detector connected by micro-connections (e.g. bump bonds). This pixel detector type has been used very successfully in experiments with high rate and high radiation and is the only detector technology that can be operated near the interaction point at the LHC experiments under such conditions. A big advantage of the hybrid approach is the possibility to separately optimise both components which are complex entities by themselves, each having development times of several years connected with high costs. Due to the hybrid design a readout chip (see section 17.6.3), which has once been developed to maturity, can be used for different sensor types as for example for planar sensors, 3D-Si sensors or for diamond pixels.

Among the disadvantages of hybrid pixel detectors there is above all the laborious assembly, in particular the cost and labour-intensive bonding of pixel sensor and readout chip. Another disadvantage is the comparatively large material thickness since both parts need a certain thickness for reliable mating. Support and cooling structures add to it. The material thickness has detrimental effects on track reconstruction and momentum resolution due to multiple scattering (section 3.4) and secondary interactions (see chapter 3). The hybrid pixel detectors of the LHC experiments ATLAS

and CMS constitute a material thickness in excess of 3% of a radiation length X_0 per detector layer.[12]

Since both sensor and readout chip are based on silicon, it is in principle possible to fabricate them as a monolithic unit. In a self-supporting silicon structure, maybe even with integrated cooling, the amount of material can be reduced by about one order of magnitude compared to hybrid pixel detectors.

Development of monolithic pixel sensors for particle detection faces the following needs and challenges:

- If possible one should exploit industrially available IC technologies, especially CMOS technologies (see also section 8.3.5 on page 297), for reasons of costs and diversity.
- A prime target is good signal-to-noise ratio by realising a sufficiently large depletion zone (large signal) together with a small electrode capacitance (low noise, see section 17.10).
- For high rate applications, like at the LHC, high radiation tolerance is required against ionising radiation (up to dose levels in the MGy range) as well as against non-ionising radiation (up to fluences in the range $> 10^{15}\,\mathrm{n}_{eq}\,\mathrm{cm}^{-2}$, see section 8.12).

Not all CMOS technology lines are suited for monolithic detector development depending on technology details. So far commercial CMOS lines almost exclusively allow only for inexpensive low resistivity Si substrate.[13] Some special technologies also allow processing on high ohmic substrate wafers and/or provide processes that tolerate higher supply voltages, both being requirements to achieve depletion depths in the range 20–120 µm. Some CMOS technologies, especially those for optical applications (CMOS cameras), have an epitaxially grown silicon layer (typically about 10 µm) into which the electronics (transistors) is implemented. This epi-layer has higher resistivity than the carrier substrate, allowing charges released by absorbed photons to be more efficiently detected.

In the following sections monolithic or at least partially monolithic pixel detectors, which contain active elements in the sensor, are described. DEPFET pixel detectors (section 8.10.2) belong to the latter class and are included in this section featuring only one active transistor in every pixel. A dedicated non-commercial and non-CMOS sensor technology[14] is used for their fabrication. *Monolithic active pixel sensors* (MAPS) employ commercial CMOS technologies where either the epitaxial layer or depleted regions of the substrate (DMAPS) are used as detection regions (section 8.10.3).

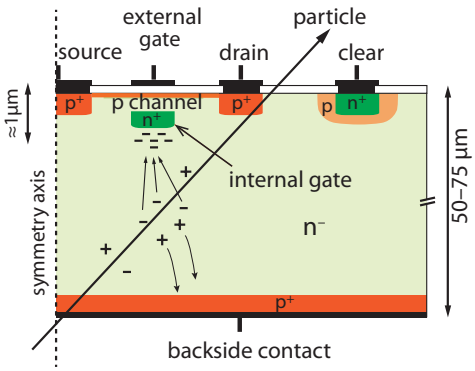
8.10.2 DEPFET pixel detectors

In a DEPFET pixel detector [594] a single transistor is implemented in every pixel cell, as shown in figs. 8.67(a) and (b) for a circular version of the pixel structure. The sensor substrate is sidewards depleted similar to the method described in section 8.8. Here the (sidewards) depletion evolves between the backside p contact and several p regions near the transistor (drain, source and a p region embedding an n^+ clear contact) and the n^- substrate. Electrons generated by impinging particles or radiation drift into a potential minimum which by proper biasing extends in the horizontal plane just underneath the (top) surface, while the holes travel to the backside (fig. 8.67(a)). The transistor is a p-channel MOSFET producing a hole current from source to drain being defined and

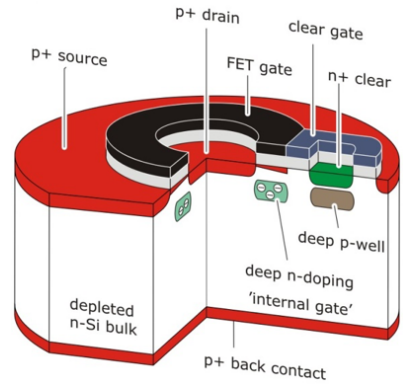
[12] In second generation LHC pixel detectors the total thickness was reduced to about 1.5%.

[13] The depletion depth d is proportional to $\sqrt{\rho}$, where ρ is the substrate resistivity.

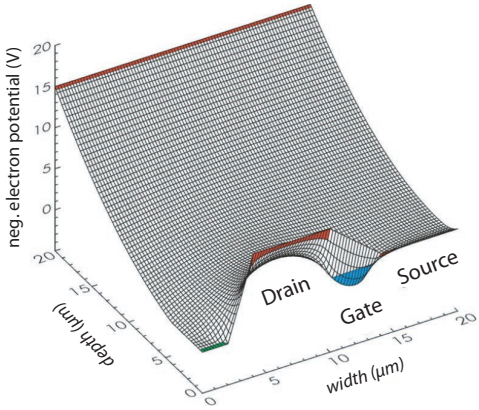
[14] Technology of the Semiconductor Laboratory of the Max-Planck Society, Munich.



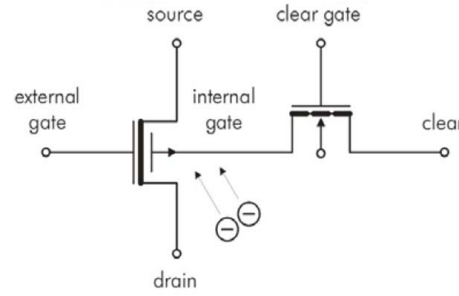
(a) Charge collection in the internal gate.



(b) Three-dimensional view.



(c) Electron potential.



(d) Electronic representation.

Fig. 8.67 DEPFET pixel: (a) cross section of one half of a circular DEPFET pixel cell with drawn symmetry axis; (b) three-dimensional view with visible additional elements needed for the operation like the *clear gate* [141]; (c) shape of the electrostatic potential for electrons near the transistor with a minimum (for electrons) at the internal gate (blue); maxima are indicated by the read areas; (d) electronic circuit representation.

controlled by the potential of the (external) gate. In addition the DEPFET structure features a deep n-implant located a few micrometres underneath the transistor channel on floating potential. Together with the external potential configuration this deep-n implant is (after a clear pulse) the most positive point in the structure and thus acts as a local minimum for electrons (fig. 8.67(c)). The collection of electrons directly underneath the transistor channel changes the potential at this place hence rendering this spot a second gate electrode, an *internal gate*.

The transistor is thus controlled by the external gate, but also by the internal gate. The corresponding representation as a circuit element has four electrodes with the internal gate electrode being connected via a pass transistor to a clear contact fig. 8.67(d). Primary electrons collected in the internal gate reside there until they are removed. The presence of their charge modulates the current flow in the DEPFET transistor channel. After the measurement has been performed the electrons are removed

from the internal gate by applying a positive voltage at the *clear* contact (fig. 8.67(a) and (c)). The clear contact is shielded by a surrounding p region (*deep p-well*), in order not to compete as a node for electron collection. Further elements appearing in the structure of fig. 8.67(b), which are not shown in fig. 8.67(a), serve to shape the potential or they aid better controlling of the operation sequence (*clear gate*). Because DEPFET pixels have very low noise (see further below) a bulk thickness of 50–75 µm is sufficient to obtain decent SNRs for traversing particle detection.

A MOSFET in saturation has a drain current [926]

$$I_D = \frac{W}{2L}\mu C_{ox}\,(V_G - V_{th})^2 \tag{8.88}$$

determined by the width W and the length L of the gate, the oxide capacitance per area C_{ox} as well as the saturation voltage determined by the gate voltage above threshold[15] $(V_G - V_{th})$; μ is the carrier mobility (here hole mobility). The amplification gain of the transistor is specified by the *transconductance*:

$$\begin{aligned} g_m = \frac{dI_D}{dV_G} &= \frac{W}{L}\mu C_{ox}\,(V_G - V_{th}) \\ &= \sqrt{\frac{2\mu C_{ox} I_D W}{L}}\,. \end{aligned} \tag{8.89}$$

Collected charge q_S in the internal gate of the DEPFET sensor capacitively couples into the channel of the transistor causing an effective gate voltage change [677]

$$\Delta V_G = \frac{\alpha q_S}{C} = \frac{\alpha\, q_S}{C_{ox} W L} \tag{8.90}$$

with αq_S ($\alpha < 1$ accounting for stray capacitances), where C denotes the gate capacitance which is represented by the oxide capacitance per area C_{ox} times the gate area $W \times L$. The gate voltage change ΔV_G modifies the drain current (8.88) to

$$\begin{aligned} I_D &= \frac{W}{2L}\,\mu C_{ox}\left(V_G + \frac{\alpha\, q_S}{C_{ox} W L} - V_{th}\right)^2 \\ &= \frac{W}{2L}\,\mu C_{ox}\left((V_G - V_{th})^2 + \frac{2\alpha\, q_S}{C_{ox} W L}\,(V_G - V_{th}) + \mathcal{O}(q_S^2)\right), \end{aligned} \tag{8.91}$$

where the term quadratic in q_S is orders of magnitude smaller than the linear term and will be neglected in the following. The change of the drain current for a given charge in the internal gate is the relevant conversion characteristic (gain) for DEPFET sensors:

$$g_q = \frac{dI_D}{dq_S} = \frac{\alpha\,\mu}{L^2}\,(V_G - V_{th}) = \alpha\frac{g_m}{C_{ox} W L} = \alpha\frac{g_m}{C}\,. \tag{8.92}$$

Typical values for g_q are 400–500 pA per electron in the internal gate. Due to the small capacitance (few fF) of the collection electrode (internal gate) the noise of the DEPFET sensor is low (see section 17.10.3). Even at room temperature noise figures of only few electrons ($\lesssim 2e^-$) have been measured [792] for DEPFET sensors when

[15] In MOSFETs the drain current saturates when the channel is pinched off at the drain saturation voltage V_{Dsat} which is given by the gate-to-threshold voltage difference.

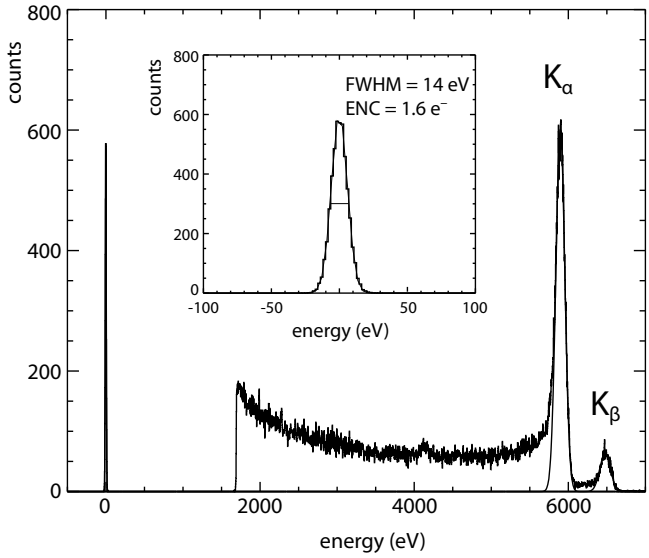


Fig. 8.68 X-ray spectrum of an ^{55}Fe radioactive source measured with a DEPFET pixel detector [678] (with kind permission of Elsevier). The Mn K_α and K_β lines are cleanly separated. The insert magnifies the section around the noise peak at 0 eV. The measured equivalent noise charge (ENC, eq. (17.109)) is 1.6 e^-.

operated with long ($\sim$ µs) shaping times of the subsequent electronics (see chapter 17). Figure 8.68 shows an X-ray spectrum from an ^{55}Fe radioactive source (K_α = 5.89 keV, K_β = 6.49 keV of the Mn daughter nucleus) recorded with a DEPFET pixel detector [678]. The spectrum is taken with a long (6 µs) shaping time of the readout electronics. Both K lines of the spectrum are nicely separated due to the good energy resolution of the DEPFET sensor due to its low noise. The inserted picture shows the noise peak which has a width (σ) of only 1.6 e^-.

The readout of a many-pixel matrix is very different from the readout of hybrid pixels discussed before. A typical arrangement of a DEPFET pixel matrix is shown in fig. 8.69. The drains of the individual DEPFET transistors are connected column-wise with each other, the (external) gate lines as well as the clear lines are connected row-wise. The matrix is steered row-wise, selecting a row by switching on the DEPFET transistors of that row via their external gates. The readout of the row is carried out column-wise via the drain lines of the transistors leading to a current-sensitive readout chip (located at the bottom in fig. 8.69). In the 'off' state no current flows in the pixels. After readout of a row the internal gates of the DEPFET transistors of that row are emptied by the clear lines. It is possible to readout the same row a second time to measure the pedestal values to be subtracted in the readout chip.

DEPFET pixel detector characteristics are a large signal (depending on the chosen thickness) because the entire detector substrate is depleted and very small noise due to the small input capacitance and the immediate in-pixel amplification.[16] Note that for monolithic pixel detectors large signals are often not easy to achieve, as will be discussed in section 8.10.3.

The pixel vertex detector of the Belle II experiment [25] at the SuperKEKB storage ring in Japan consists of DEPFET pixel detectors [428] using a rectangular variant of fig. 8.67(b). The good SNR is here exploited to design a detector with very low material content. Small amounts of material (more specifically small number of radiation

[16] In high rate experiments requiring intrinsically short shaping times of typically 50–100 ns (see section 17.3) extremely low noise values can, however, not be maintained.

lengths, see (3.88) on page 61) in the passage of particles are particularly important for the momentum measurement of particles with low momenta ($\lesssim 2\,\mathrm{GeV}$) where multiple Coulomb scattering dominates the measurement uncertainty. For the Belle II pixel detector a reduction of material is achieved without loss in stiffness by anisotropic etching of the DEPFET structures from the backside [97], by which only the centre part is etched down to 75 µm in a trapezoidal shape, as can be seen in fig. 8.70(b). By shingling the modules the total cylinder area is covered by the active thin regions of

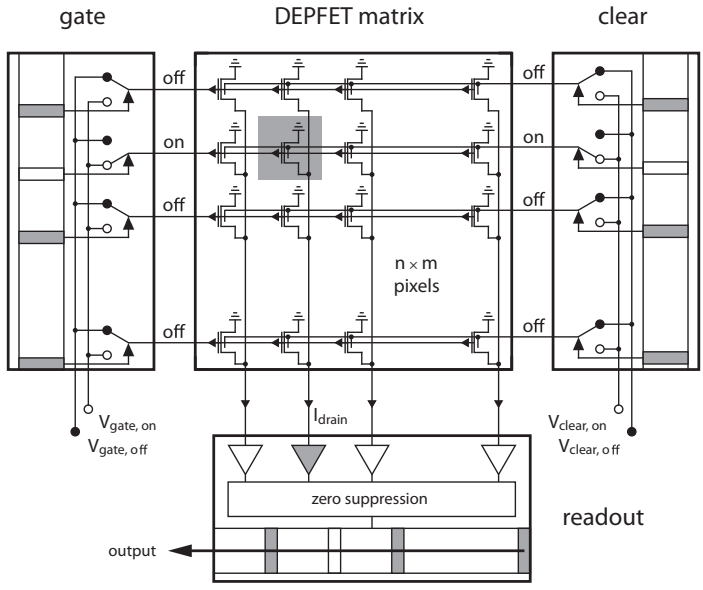
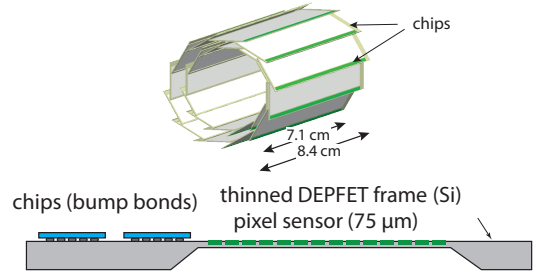


Fig. 8.69 Readout scheme of a DEPFET matrix. The gate chips address the rows one-by-one. The transistor drain lines (outputs) are routed to the readout chip(s) column-wise (bottom). Thereafter the charges in the internal gates of the row are cleared (clear).

(a) Pixel vertex detector (PXD) of Belle II (model)



(b) Thinned DEPFET pixel module

Fig. 8.70 DEPFET pixel vertex detector PXD (Belle II). (a) Model (source: Belle II PXD Collaboration). (b) Cross section through a pixel module perpendicular to the beam. The active area of the sensor is thinned from the backside by an etching process [97] to 75 µm. The thick regions serve as support structures. On them the steering and readout chips are placed. The active area has dimensions $44.8 \times 12.5\,\mathrm{mm}^2$. Module ladders contain two modules each. The radii of the DEPFET pixel layers are 1.4 cm and 2.2 cm, respectively.

Fig. 8.71 Radioactivity distribution of a leaf marked with tritium, recorded by a DEPFET pixel detector [953]. (a) Photograph of the leaf, (b) grey-coded activity (lighter color represents higher activity).

the module. The thick regions, on which the steering and readout chips are placed, are at the rim of the module and provide the needed stiffness. A 1:1 model of the detector is shown in fig. 8.70(a).

Apart from particle physics DEPFET pixel detectors find applications in X-ray astronomy [918]. They are particularly well suited for low-energy X-rays with energies down to below 1 keV [946] since the rate of X-ray quanta is low and DEPFET detectors readout with long filter times exhibit very low noise features. DEPFET pixel detectors are also well suited for tritium autoradiography of biological samples. Tritium is a favoured radio-marker replacing hydrogen in biomolecules. Figure 8.71 shows the image of the spatial ^{3}H distribution in a ^{3}H radio-marked leaf of a plant [953]. The detection of tritium by its β decay is challenging because the endpoint energy of the β spectrum is only 18.6 keV (the most probable electron energy is 2.5 keV), which in Si corresponds to only about 680 generated e/h pairs.

8.10.3 Monolithic CMOS pixel detectors

Employing commercial CMOS technologies to produce a monolithic (rather than hybrid) pixel detector, in which pixel sensor and electronics circuitry form one entity, was first proposed and realised in the early 1990s [759,598]. As a follow up CMOS technology lines were employed to develop so-called MAPS (*monolithic active pixel sensors*) detectors [706,952], first by exploiting an epitaxial silicon layer for (diffusive) charge collection and later by using fully depletable, high-resistivity substrates processed in CMOS lines [770,882,510,519] to develop DMAPS (depleted MAPS). The use of 'full CMOS', that is, using PMOS and NMOS transistors on the same substrate without any restrictions (see also section 8.3.5), is an essential ingredient of the development as it allows employing CMOS logic circuitry in monolithic detectors.

MAPS pixel detectors. Commercial technologies often use low cost, low-ohmic substrate wafers on which an epitaxial Si layer (epi-layer, see section 8.9.1 on page 323) is grown (fig. 8.72). In the epi-layer doping profile and conduction type can be controlled independently of the substrate and can be made chemically cleaner and with higher resistivity than the substrate. The epi-layer thickness typically is in the range 1–20 µm. Process lines for optical applications (CMOS cameras) often use (thick) epi-layers for higher light yield achievable with the higher resistivity. The electronics circuitry is implanted at the top of the epi-layer.

In light detection many photons impinge homogeneously distributed on a pixel cell. It is not important if only part of the total generated charge is detected as long as the fraction is constant on average. The effectively active fraction of the pixel area (*fill factor*), for instance in CMOS cameras, can hence be significantly smaller than 100%.

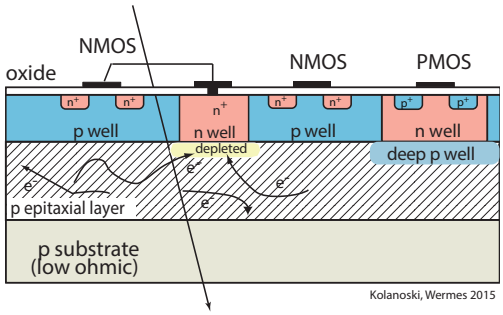


Fig. 8.72 Monolithic pixel detector with charge collection in a Si epitaxial layer (MAPS). Charge collection is largely governed by diffusion in this layer, here at an n^+ contact (n well). Other n wells, for example those hosting PMOS transistors, are competing for charge collection and must therefore be shielded, for example by a deep p well. Example transistors are shown in the electronics layer. The signal path from the collection node to the amplifier is also indicated.

Particles traversing the epi-layer can also be detected. Figure 8.72 shows the principle of such MAPS detectors. Electrons can arrive—in small parts by drift in the small depleted region, but mainly (and incompletely) by diffusion—at a collection electrode realised by an n^+ well.[17] Since there is no directional drift field in the epi-layer (except for the small depleted region directly underneath the collection node) charge collection mainly occurs by undirected diffusion and is hence incomplete and comparatively slow (of the order of 100 ns). Due to this and the small thickness of the epi-layer the deposited charge is very small (typically $\lesssim 1500\,e^-$). To achieve reasonable SNR values in the order of 10:1 the noise of the readout electronics must be correspondingly low.

The collection diode in fig. 8.72 is an n well. Other n wells would therefore compete in charge collection. Hence PMOS transistors which sit in n wells should also be avoided. As a consequence, complex CMOS circuits, based on equal usage of NMOS and PMOS transistors, can only be realised outside of the active area of the pixels. To use both NMOS and PMOS transistors in the active pixel area, further deep[18] wells are needed (see e.g. [146, 809]). In fig. 8.72 a highly doped deep p well, implanted underneath the n well containing the PMOS transistors, shields this n well from becoming a competing collection node. The substrate is not explicitly contacted and the p-doped region assumes a negative potential with respect to the collection diode and hence has a repulsive effect. As an advantage, the total thickness of the device can be made very thin, typically 50–100 µm, offering pixel detectors with a very small material budget.

For time uncritical applications the readout can be realised using only a few (three) NMOS transistors which do not need to sit in the pixel cell itself, as shown in fig. 8.73. The simple three-transistor readout (3T readout) has (i) a row *select*, (ii) a low-impedance signal output via a *source follower* (without amplification) and (iii) a *reset* input line. Further CMOS processing, such as signal discrimination and zero suppression, is done outside of the active pixel area. For LHC, MAPS detectors with 3T readout are too slow to cope with the data rates, in particular for the pp collision experiments. For heavy ion collisions, however, MAPS pixel detectors, using charge or voltage amplification and dedicated readout schemes, have been successfully employed in the STAR experiment at RHIC [318] and are also used for the ALICE upgrade at the LHC [682].

[17] Wells are doped regions in which MOS transistors are embedded in order to realise them in substrates with any doping: for example n wells to realise p-channel MOSFETs in p substrate, p-wells for n-channel MOSFETs in n substrate.

[18] 'Deep' here means deeper below the surface than the transistor wells.

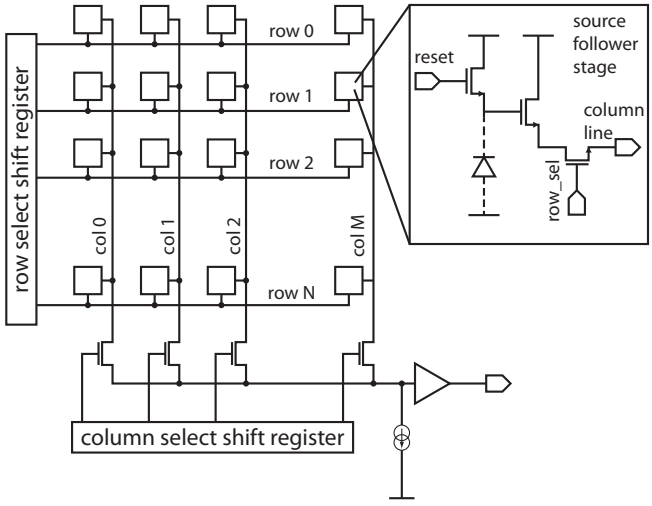
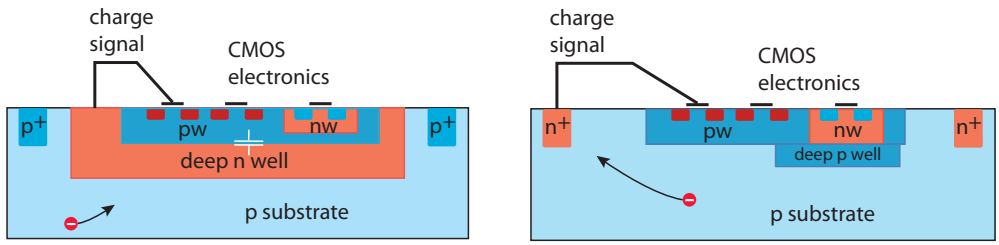


Fig. 8.73 Simple three-transistor readout scheme of a MAPS pixel matrix with row select, reset (not shown in matrix) and source-follower output as shown in the insert. The sensing pixels are represented by a diode drawn dashed. All pixels have row- and column-selects. The output is organised column -wise via the column-select line and is further amplified.

DMAPS pixel detectors. For high rate and high radiation environments, as for example at the LHC, fast ($\mathcal{O}(10\,\text{ns})$) and complete collection of deposited charge by drift in an electric field is mandatory. The depletion depth d in the Si substrate underneath a well depends on the substrate resistivity ρ and the bias voltage V (eq. (8.60)):

$$d \propto \sqrt{\rho V}\,.$$

Hence process technologies or process add-ons must be employed that allow applying higher voltages than usually allowed in CMOS technologies [770] or accept processing of high-ohmic substrate wafers [882, 510] or a combination of both. As a result the depletion region underneath the collection electrode can be increased (typically to 25–150 µm) to provide a sufficiently large and fast signal (*depleted* MAPS, DMAPS). In addition, the employed technology must provide multiple wells (typically four including deep n,p wells) to enable (a) shielding of wells that must not attract charge, and (b) to provide full CMOS circuit functionality by an unimpeded usage of both NMOS and PMOS transistors by having them decoupled from the substrate [510].



(a) Large collection electrode (b) Small collection electrode

Fig. 8.74 Two principal variants of CMOS cell geometries [442] (schematic): (a) Large electrode design: the charge collecting deep n well encloses the complete CMOS electronics. With a backside contact the depletion region builds up between deep n well and backside. (b) Small electrode design: the charge collection node is placed outside the CMOS electronics area. The red dot represents an electron.

In fig. 8.74(a) the cross section of a DMAPS pixel detector with a *large collection electrode* is shown realised by a deep n well which at the same time contains the entire CMOS circuitry. The depletion zone is either created from the top surface only or, by applying backside processing, a rear contact can be fabricated that allows depletion volume building up between the deep n well and the backside. Radiation tolerance benefits from the fact that the average drift distance of charges needed to reach the collection electrode is short and the trapping probability is therefore reduced. Owing to the large geometrical size, however, the total capacitance the electrode presents to the amplifier input is large, augmented by a significant additional capacitance contribution arising from the close distance of deep n and deep p well (indicated in fig. 8.74(a)). The latter is new in comparison with standard, non-monolithic pixels as it originates from the necessity to use deep wells to shield the embedded electronics. The total capacitance can amount to several hundred fF. A large amplifier input capacitance C_D increases the noise and decreases the timing performance of a pixel detector with charge sensitive readout (see section 17.10) as:

$$ENC^2_{thermal} \propto \frac{4}{3}\frac{kT}{g_m}\frac{C_D^2}{\tau_{sh}} \qquad \text{and} \qquad \tau_{CSA} \propto \frac{1}{g_m}\frac{C_D}{C_f}\,, \tag{8.93}$$

where τ_{CSA} characterises the CSA's time behaviour (e.g. the rise time dependence), C_f is the feedback capacitance, g_m the transconductance, τ_{sh} the shaping/filter time and kT the Boltzmann constant times the temperature. The new capacitance unfortunately also allows that fast transient signals appearing in the embedded electronics can couple into the sensing part, thus mimicking signals. Dedicated circuitry typically introducing some current limitation is needed to prevent transient couplings causing artificial signals.

An alternative shown in fig. 8.74(b) has a small charge collecting electrode (n well) set apart from the CMOS electronics; in the 2D view the electrode is in the centre of the pixel. The *small electrode variant* can have very small collection node capacitances C_D of only 5–20 fF, which is very beneficial for noise and timing performance

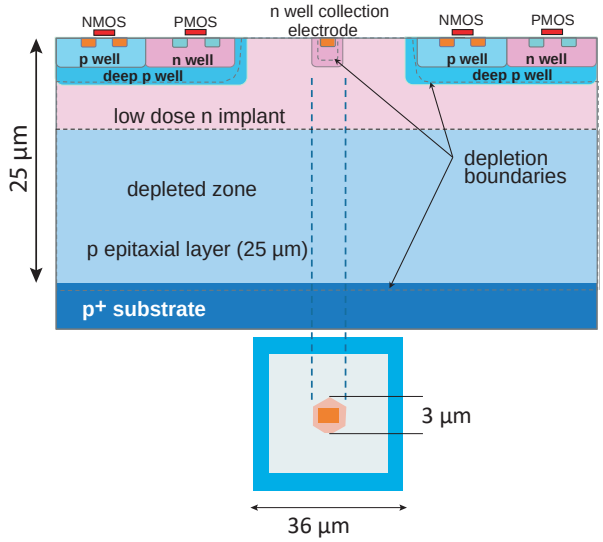


Fig. 8.75 Small electrode design with lateral field strengthening by means of an n-doped layer [901]. The device is depleted from two diode structures, between the deep p-wells and n implant and between n implant and p epitaxial layer. The collection electrode is on positive potential, as high as the technology allows. The epi-layer thickness is 25 µm. The layout features a very small input capacitance.

(see eq. (8.93)). However, radiation tolerance is more difficult to achieve than in the large electrode design due to on average longer drift distances to the collecting electrode (for the same pixel cell size) and hence larger trapping probability. Therefore small pixel sizes are preferred for small electrode designs at the expense of a higher power density, because the power needed is roughly constant per pixel, driven by the amplifier. Figure 8.75 shows a small-electrode design realised in a technology employing a 25 µm thick epitaxial layer. The technology has been modified [901] by adding an n-doped layer that serves to strengthen the electric field laterally. The epi-layer is depleted from two pn junctions (the deep p well / n layer and the p epi / n layer junctions) similar to the sidewards-depletion technique discussed in section 8.8. This creates a potential minimum for electron collection underneath the deep p wells with a field direction towards the (positive) n collection well, thus strengthening collection of charges laterally. The small charge signal achieved in the thin epi-layer (typically 1600 e^-) becomes a sizable voltage signal (about 50 mV) due to the small (∼5 fF) capacitance according to $dv = dQ/C_D$. Therefore voltage (rather than charge) amplification is employed for the readout. The signal-to-noise ratio in terms of voltage can be written as:

$$\frac{S}{N} = \frac{Q/C_D}{\sqrt{\langle v_n^2 \rangle}} \propto \frac{Q/C_D}{\sqrt{g_m}}\,, \tag{8.94}$$

where Q/C is the voltage signal and $\langle v_n^2 \rangle = \frac{4}{3} kT \frac{1}{g_m} \frac{1}{\tau_{sh}}$ is the equivalent-noise voltage (thermal) referred to the input (see section 17.10.3.3 for noise considerations) with the other quantities as defined in (8.93).

DMAPS pixel designs have been realised in large [527, 178] as well as in small electrode [179] configurations showing high rate and radiation performance capability to levels encountered at the LHC.

Monolithic SOI pixels. Another approach to combine the electronics layer and the depleted sensing part in a monolithic way is achieved by employing the *silicon-on-insulator* (SOI) technology [116]. Figure 8.76 shows the principle. The transistor layer is separated from the bulk by a SiO_2 layer called *buried oxide* (BOX). In addition the transistors are individually isolated by vertical trenches. The bulk wafer in SOI technology is normally just a (low resistivity) handle wafer.[19] If a high resistivity wafer is chosen instead the volume underneath the BOX can be depleted and used to provide a comparatively large charge signal upon particle passage. Connection between

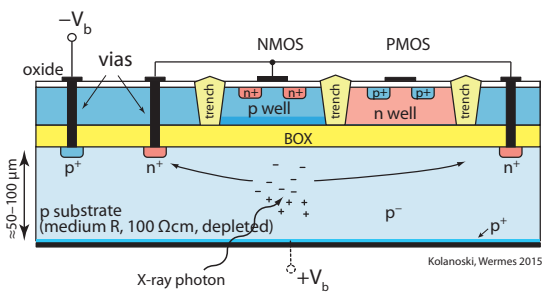


Fig. 8.76 SOI MAPS: Concept based on *silicon-on-insulator* technology. The CMOS electronics layer and the sensor are separated by a 'deep' (≈ 1 µm below the surface) Si dioxide layer (BOX). The transistors are also isolated from each other by vertical trenches. In the example the deposited energy is generated by an absorbed X-ray photon.

[19] Handle wafers are used for handling operations in wafer technology, for example to aid handling of very thin wafers.

substrate and CMOS electronics layer is obtained by vertical connections (*vias*). Also here a particular challenge is to avoid capacitive coupling of the electric field in the sensor layer through the BOX into the electronics layer hence leading to significant transistor threshold shifts (back gate effect). The BOX also traps positive charges which have low mobility in SiO_2 and tend to be fixed at the interface, hence causing radiation effects (see section 8.12). Using technology variants these problems can be solved or at least be mitigated [709, 519]. Shown in fig. 8.76 is an SOI approach that has doped volumes underneath the electronics above the BOX (partially depleted SOI) [519]. This prevents the direct coupling from the substrate into the electronics mentioned above.

8.11 Precision timing with silicon detectors

Over decades silicon detectors have been used for precision space point measurements or in nuclear physics for energy measurement. Precision time measurement was not a prime focus of silicon detector research for a long time. This is due to the fact that charge collection in silicon detectors takes place on typical time-scales of a few nanoseconds and typical amplifier rise times are in the range 25–50 ns. For competitive time measurements, for instance in comparison to scintillators or RPCs (see sections 14.2.1 and 7.7.3), resolutions in the range 10–100 ps are required, 2–3 orders of magnitude below the characteristic time-scales of charge collection and amplification. To enter these timing regimes, faster collection times and large signals together with a minimisation of all possible time jitter contributions is necessary.

So-called *low gain avalanche diodes* (LGADs) with mm^2 size pattern structures have been developed to cope with this challenge [843, 845]. In order to minimise time fluctuations in the signal generation process, an amplification structure with moderate gain ($m \approx 10$–50) is realised by a p^+ implantation ($N_A \approx \mathcal{O}(10^{16}/cm^3)$) right underneath the n^{++} electrode,[20,21] as shown in fig. 8.77 [274]. The p^+n^{++} (metallurgical) junction creates a very high electric field of up to 300 kV/cm, leading to the amplification.

The fast signal is obtained by fast e/h movement in the high fields of thin detectors biased at high voltage. In addition, the amplification structure provides a large amplitude with fast rise time (large slew rate dV/dt). Interestingly, this is mainly achieved by holes rapidly moving away from the amplification structure—very reminiscent of the signal development in wire chambers (see section 5.3.2), which is governed by the movement of the positive ions. Here, individual electrons generated by the traversing particle along its track drift towards the amplification region where they are amplified creating m electron–hole pairs. The electrons are collected at the n^{++} electrode being less than about a micrometre away from the amplification point. Their contribution to the signal therefore is very small. Instead, the amplified holes moving (fast) away from the junction towards the p^{++} back electrode constitute the main part of the signal which governs the timing resolution.

The signal development can be compared to the signal generation of a wire chamber described in section 5.3.2 on page 142 having similar features. The composition of the signal induced on the readout electrode is shown in fig. 8.77. All requirements to

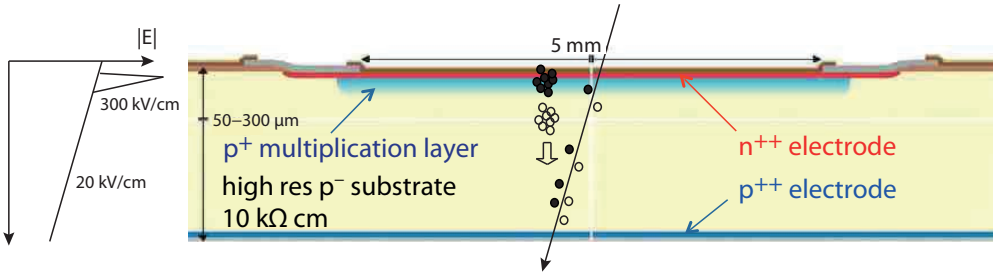
[20] We here adopt the notation p^{++} for $N_A \gg 10^{16}/cm^3$, likewise for n^{++}.

[21] The amplification junction can also be fabricated at the side opposite to the electrodes, see e.g. [769].

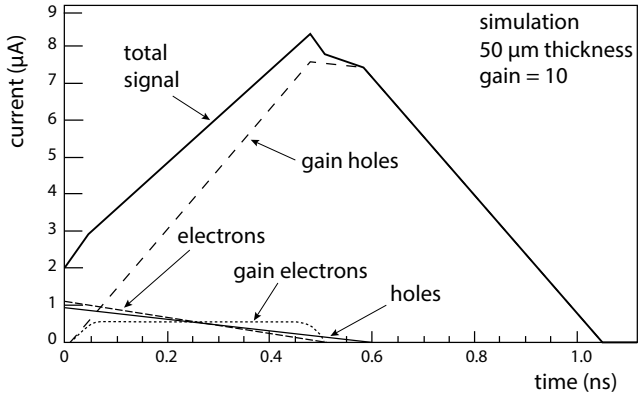
obtain large slew rate (high fields of order 20 kV/cm, fast drift movement of both carrier types, short bunching of electrons from the primary deposition) benefit from using thin (50–100µm) silicon sensors. Gains m are kept below 50 in order to avoid amplification excess noise (compare section 10.3.2 and fig. 10.19 on page 425). The described fast signal development is obtained for a homogeneous weighting field, that is, a large ratio of electrode area to pitch, avoiding low field regions which would occur for example between small pixels with pitch much smaller than thickness.

The achievable time resolution has several contributions [275] which must be quadratically summed:

$$\sigma_t^2 = \underbrace{\left(\frac{(\Delta v_s^{th})_{rms}}{dV/dt}\right)^2}_{\sigma^2_{\text{time walk}}} + \underbrace{\left(\frac{v_n^{rms}}{dV/dt}\right)^2}_{\sigma^2_{\text{noise}}} + \sigma^2_{\text{arrival}} + \sigma^2_{\text{dist}} + \sigma^2_{\text{TDC}} \tag{8.95}$$



(a) LGAD structure.



(b) Contributions to the signal of a minimum-ionising particle.

Fig. 8.77 Low gain avalanche diodes (LGADs) are capable to provide high time resolutions. (a) LGAD structure featuring a high ohmic p-type bulk, and an amplification junction. The electric field is sketched on the left. (b) LGAD signal pulse detailing the contributions from electrons and holes before and after amplification following weighting field simulations [283]. The contributions from both electron and hole signals are shown before and after amplification (denoted 'gain electrons' and 'gain holes') as well as the total signal. Adapted from [274] with kind permission from Elsevier.

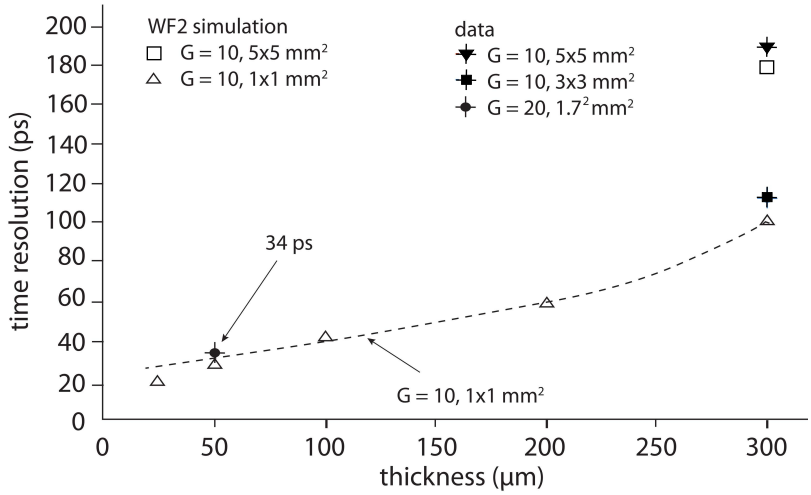


Fig. 8.78 Comparison of time resolutions from simulations (WF2) [283] and from test beam measurements. Data points are from [274, 844, 276, 277, 275].

The first term represents time walk introduced by signal pulse height variations (Landau fluctuations) in the energy loss process with $(\Delta v_s^{th})_{rms}$ being the rms of the signal height variations at discriminator threshold. Minimisation of this term is possible for example by employing constant fraction discrimination (see section 17.5) or by applying corrections using amplitude information. The second term is noise jitter with $v_n^{rms} = \sqrt{\langle v_n^2 \rangle}$, that is, time fluctuations due to noisy signals which can also be expressed as $t_r/(SNR)$ where t_r is the rise time. Both contributions are made small when large pulse slew rates (dV/dt) are achieved. An irreducible contribution comes from fluctuations resulting from non-uniform depositions of charge along the particle track moving towards the amplification junction, hence causing an intrinsic jitter in the arrival time σ_{arrival} (third term). Fluctuations from secondary ionisations and from the amplification process enter here as well. The thinner the detector the less disturbing is this effect. The fourth contribution is signal distortion due to non-uniform weighting field regions and variations in (non-saturated) drift velocities. The final term denotes time fluctuations due to uncertainties in the time digitisation. The latter contribution can be made negligible (below 10 ps) with GHz TDCs.

The LHC experiments ATLAS and CMS explore exploiting the precision time measurement offered by LGADs in detector applications, ATLAS in a forward calorimeter [83] and CMS at the outside of the tracker (see e.g. [235]). The major problem concerning operation in a high radiation environment is the fact that the doping concentration of the p^+ layer of the amplification structure will change with radiation fluence and hence also will the gain of the device.

Weighting field simulations [283] are compared with test beam measurements in fig. 8.78 for different sensor thicknesses. Simulations and measurements agree and point to achievable resolutions in the range of 30 ps for very thin detectors.

8.12 Radiation damage

During operation a particle detector is exposed to an intense flux of charged and neutral particles as well as γ- and X-rays. Even neutrons can abundantly occur from backscattering in dense materials as present in calorimeters or absorbers. At the LHC the ratio of pions to neutrons is 50:50 at about 30 cm distance from the interaction point and 90:10 at 5 cm distance. Depending on the particle flux level this leads to damage of the exposed material with time, so-called radiation damage. We shall discuss the physics and phenomena of radiation damage for silicon as a detector material. They are however, with some restrictions, largely also valid for other semiconductors.

Semiconductor detectors are often operated close to a target or a collision point at an accelerator experiment and are thus exposed to particularly high particle fluxes. We distinguish in the following:

– Damage of the silicon crystal (substrate volume damage) by particles impinging on the lattice. This damage has a dominantly non-ionising nature (*non-ionising energy loss*, NIEL), meaning that it is caused by collisions with the nuclei of the lattice atoms. These can lead to phonon excitations of the lattice on the one hand, which do not cause any damage, but also to dislocations of the lattice atoms or more complex distortions of the crystal lattice.
– Surface damage and damage of boundaries and interfaces (e.g. Si–SiO_2) of semiconductor sensors and of readout chips by means of ionisation energy loss of impinging radiation (*ionising energy loss*, IEL).

Substrate damage in the crystal volume affects the sensor characteristics, in particular its charge collection properties, whereas surface and boundary damage cause above all transistor threshold shifts in the CMOS electronics or can give rise to unwanted parasitic currents between transistors that may compromise the functionality of the electronics.

Conventionally, for silicon sensors the NIEL substrate damage is normalised to the damage level caused by 1 MeV neutrons (see section 8.12.1). The total radiation fluence, which is the number of particles per cm^2 traversing a material over some amount of time (usually the assumed life time of the detector), is hence quoted in 1 MeV neutron equivalents n_{eq}/cm^2. At the LHC, for example, the fluence ϕ_{eq} after an assumed 10-year running time will rise to levels of the following order of magnitude, depending on the distance from the collision point: $\phi_{eq} \approx 10^{15}$ n_{eq}/cm^2 for the innermost pixel layers (at a distance of $r > 4$ cm) and $\phi_{eq} \approx 10^{14}$ n_{eq}/cm^2 for microstrip detectors ($r > 20$ cm). For the planned high luminosity upgrade of the LHC these numbers must be multiplied by a factor of roughly ten.

8.12.1 Substrate damage

Damage processes. At the high luminosity LHC upgrade in the region close to the interaction point, every silicon lattice cell is traversed by about 50 particles during its operation lifetime. Depending on the type and the energy of the radiation, different defect mechanisms can appear in semiconductor materials. Ionisation is a dominant energy loss mechanism for most practical particle momenta (see also section 3.2). Since in semiconductors ionisation is a largely reversible process, no permanent damage remains inside the crystal except at isolation layers and implant boundaries.

The situation is different in the case of non-ionising damage, resulting from direct collisions with the atomic nuclei of the crystal lattice. These can knock-off atoms from

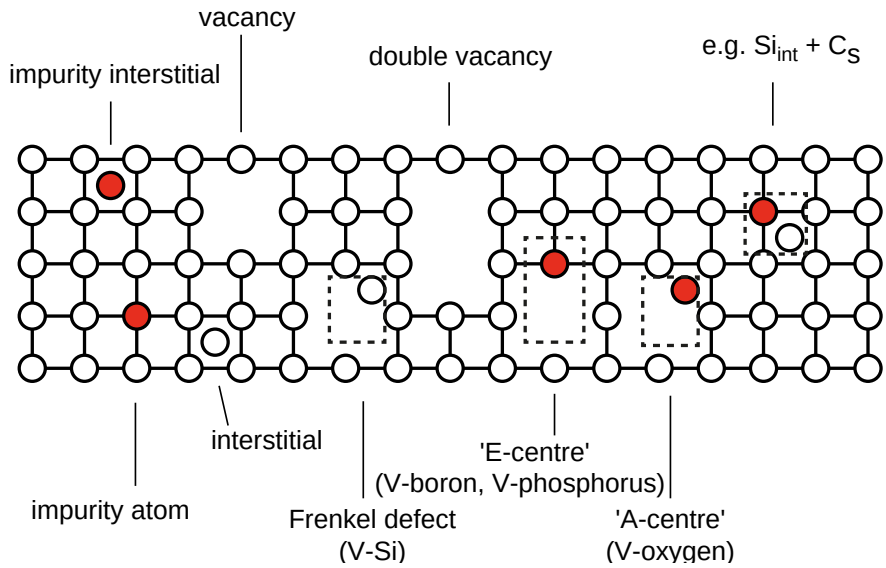


Fig. 8.79 Different defect types in a silicon lattice caused by particle radiation. The filled red circles are impurity atoms. More complex defects are marked by the dashed rectangles. V marks a vacancy. Si_{int} is a lattice interstitial. C_S is a carbon atom substituting a lattice atom. Impurities can enter a crystal for instance during the crystal growth.

their position inside the lattice, thereby creating vacancies and interstitials in between lattice atoms. Some of these primary point defects are illustrated in fig. 8.79. In addition neutron captures and nuclear transmutations can occur. A considerable amount of these primary point defects is unstable and dissolves again by recombination[22]. Since they are mobile in the lattice they can then form stable defect complexes together with existing impurities (fig. 8.79, see also [910]).

The collision processes are Coulomb scattering off nuclei (for electrons, charged pions, and protons) as well as elastic (also inelastic) scattering off nuclei for neutrons. To kick-off an atom from the silicon lattice with more than 50% probability, a minimum energy of 25 eV is necessary [662]. The maximum energy that can be transferred in a central elastic collision from a non-relativistic particle with kinetic energy T and mass m to another particle with mass M is using classical scattering theory:

$$T_{max} = 4\frac{Mm}{(M+m)^2}T\,. \tag{8.96}$$

Table 8.4 shows cross sections as well as maximal (for $T = 1\,\mathrm{MeV}$) and average energies for e, p, n and knocked-off Si nuclei in silicon. For nucleons and nuclei T_{max} is more than three orders of magnitude larger than the minimal energy of 25 eV needed to kick off an atom from the lattice. The released atom thus has sufficient energy to kick off further atoms before it comes to rest, and so can do the secondary atoms. Instead of a point defect, defect clusters with typical dimensions of 10 nm × 200 nm along the

[22] Point defects can recombine via different mechanisms, among others for example individual recombination by defect diffusion to the surface or by mutual annihilation of interstitial defects with vacancies.

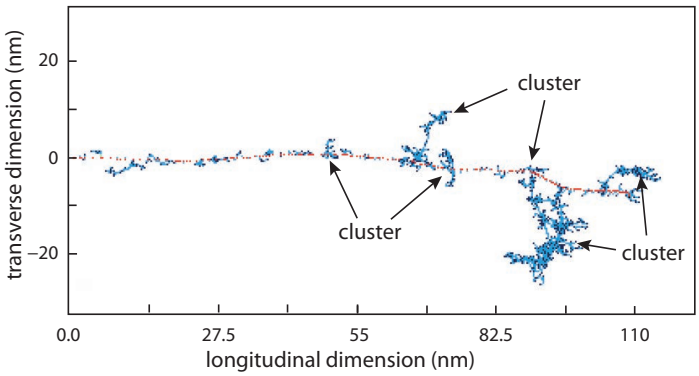


Fig. 8.80 Simulation of lattice damage in silicon [579] caused by a primary lattice atom, kicked off at $(x, y) = (0,0)$ with a kinetic energy of 50 keV, a typical recoil energy from 1 MeV neutrons.

Table 8.4 Comparison of maximum and average energy transfers for kinetic energies of 1 MeV as well as cross sections (integrated from a threshold energy $T_{min} = 25$ eV to T_{max}) for electrons, protons, neutrons, and for (knocked-off) silicon nuclei. The cross sections for Si nuclei and neutrons are relevant for the creation of damage clusters (after [1008] and references given therein.)

Particle species cross section	Electrons Coulomb scat.	Protons Coulomb scat.	Neutrons elastic scat.	Si^+ nuclei Coulomb scat.
T_{max}	155 eV	134 keV	134 keV	1 MeV
$\langle T \rangle$	46 eV	210 eV	50 keV	265 eV
σ (10^{-24} cm^2)	44	17 950	3.7	502 500

direction of the impact are created, consisting of many vacancies and interstitials. A simulation of cluster damage in silicon caused by a kicked-off lattice atom is shown in fig. 8.80. For electrons the formula for T_{max} is given by the maximally transferable energy appearing in the relativistic equation (3.21) on page 28 for the energy transfer in an ionisation process—with exchanged roles of the scattering partners, the electron and the heavy particle ($m_e \leftrightarrow M$). Neglecting then in (3.21) the first two terms in the denominator compared to the third, the maximum energy that an electron with kinetic energy T_e can transfer to an atom becomes

$$T_{max} \approx 2\frac{T_e + 2m_e}{M} T_e \,, \tag{8.97}$$

which at $T = 1$ MeV is about three orders of magnitude smaller than for collisions with protons or neutrons (table. 8.4).

While the distribution of the energy transfer in elastic neutron–nucleus scattering is almost flat, for protons the Coulomb scattering off lattice nuclei causes a steep decrease towards large energy transfers according to the Rutherford scattering formula. Hence the cross section averaged energy transfer $\langle T \rangle$ is much smaller for protons than for neutrons. On the other hand the total cross section for Coulomb scattering is much larger than the one for elastic neutron scattering off nuclei. Table 8.4 compiles maximum and average energies of particle species as well as cross sections integrated from a minimum energy transfer for lattice damage T_{min} to the maximum possible transfer T_{max}. A direct consequence is the fact that electrons (and also gamma rays) as well as protons and charged pions create many more point defects than do neutrons.

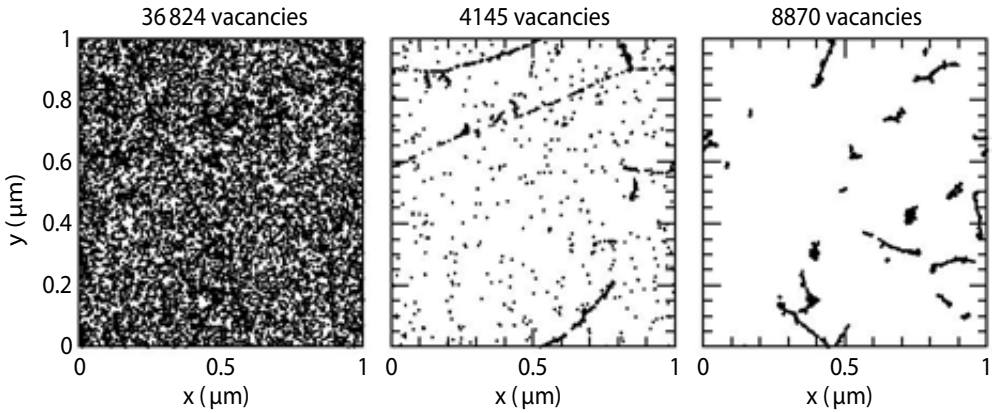


Fig. 8.81 Initial distribution of lattice vacancies in a silicon volume of 1 μm^3 for (a) 10 MeV protons, (b) 24 GeV protons and (c) 1 MeV neutrons after a fluence of 10^{14} particles per cm^2. Shown is the x–y projection at a depth of 1 μm [548] (with kind permission of Elsevier).

Figure 8.81 give an impression of the different damage forms and levels obtained after a fluence of 10^{14} particles per cm^2 in a volume of 1 μm^3, respectively for 10 MeV and 24 GeV protons and for 1 MeV neutrons.

The NIEL hypothesis. The non-ionising energy loss or NIEL denotes the energy lost by reactions other than ionisation, which is, almost exclusively, by the production of atomic displacements in a material of a given density and is specified in the same units as used for the stopping power (see section 3.2.1.2), namely MeV/g/cm^{-2} [910]. It is that part of the energy introduced to the material via elastic (both Coulombic and nuclear) and nuclear inelastic interactions that produces the initial vacancy–interstitial pairs and phonons (e.g. vibrational energy) in the lattice. The kicked-off primary atom (PKA) with recoil energy E_R leaves a vacancy and releases its energy by ionisation and by creating further displacement damage until it comes to rest as a lattice interstitial.

NIEL can be calculated for electrons, protons, neutrons and pions as [908, 645]:

$$\left.\frac{dE}{dx}\right|_{\text{NIEL}}(E) = \frac{N_A}{A} D(E)\,, \tag{8.98}$$

where N_A is the Avogadro number, A the atomic mass in g/mol of the target material and $D(E)$ is an energy and particle type dependent *damage function* taking into account the cross sections [645]:

$$D(E) = \sum_i \sigma_i(E) \int_{E_d}^{E_R^{max}} f_i(E, E_R)\, P(E_R)\, dE_R\,. \tag{8.99}$$

Here E and E_R are the kinetic energies of the impinging particle and of the recoil atom, respectively, while E_d is the minimum energy required for dislocation damage (in Si $E_d \approx 25$ eV). The index i runs over all occurring reactions with cross sections σ_i. The function $f_i(E, E_R)$ denotes the probability that in reaction i a recoil atom of energy E_R is produced, whereas $P(E_R)$ is a partition function returning the fraction E_{dam}/E_R available for further displacement damage. The other part $1 - E_{\text{dam}}/E_R$ is mostly for ionisation. At small energies ($<$ keV) the E_{dam}/E_R fraction is about 80–90%, but for

energies in the MeV range it drops below 20% (see [1008]). The damage curves $D(E)$ for silicon are shown in fig. 8.82 for neutrons, protons, pions and electrons.

For silicon as a material, the so-called *NIEL hypothesis* is a good first order description of the observed damage effects (see e.g. [714]), in particular what the leakage current generation due to NIEL damage is concerned (see page 350). It states that all lattice radiation damage in silicon linearly scales with the NIEL and can be traced back to the abundance of primary defects (point defects and clusters), irrespective of their initial distribution over energy and space, that is, whether they are homogeneously scattered over a large volume (typical for low-energy proton or gamma ray damage) or whether they are densely clustered in small regions (typical for neutron damage). Assuming the NIEL hypothesis, the observed differences in damage caused by neutrons, protons, pions and electrons can be scaled to each other using fig. 8.82 (*NIEL scaling*).

The damage effect of 1 MeV neutrons corresponds to $D_n\,(1\,\mathrm{MeV}) = 95\,\mathrm{MeV\,mb}$ [660] and is used as the normalisation point in fig. 8.82 for the damage caused by radiation from different particles at different energies. NIEL scaling means that the efficiency of a damage created by a different particle species at a different energy is described by the ratio κ of D_x for a radiation species x at an energy E to neutron damage D_n at 1 MeV:

$$\kappa = \frac{\int_{E_{min}}^{E_{max}} D_x(E)\phi(E)dE}{D_n(1MeV)\ \int_{E_{min}}^{E_{max}} \phi(E)dE}\,. \tag{8.100}$$

κ is often called *hardness factor* or *damage factor* [714]. As an example, for 24 GeV protons a value of $\kappa \approx 0.62$, for 25 MeV protons $\kappa \approx 2.0$ is used [353]. The *equivalent fluence* for 1 MeV neutrons then is the damage-weighted real fluence that a detector has received from a given particle species at a certain energy:

$$\phi_{eq} = \kappa\,\phi\,. \tag{8.101}$$

Typical (kinetic) energies of the particles mainly causing bulk damage at the inner zones ($r < 20\,\mathrm{cm}$) of the LHC experiments are 0.6–1 GeV for protons and 10 MeV for neutrons.

It should be mentioned that NIEL scaling, while being very powerful for many aspects of damage descriptions, has shortcomings arising for example from the fact that point and cluster defects contribute differently to various damage parameters important for detectors. Damage parameters which are not well described by the NIEL hypothesis, are for example changes in the effective space charge concentration N_{eff} (see below on page 349) which depends also on the impurity concentrations (especially carbon and oxygen) as well as the trap introduction rate, which is particle-type dependent. Also, annealing effects (see section 8.12.4) are not included when assuming NIEL scaling (see [714] and references therein).

The action of intermediate levels: the Shockley–Read–Hall framework. Substrate damage can result in new energy levels within the silicon band gap that may become electrically active, that is, they can generate or absorb freely moving charge carriers by transitions to or from the conduction or valence bands. Depending on their location within the band gap they can act more as donors or as acceptors and their effect can thus be different. Three main defect types and their effects are sketched in fig. 8.83 which cause (a) generation of donor and acceptor centres, (b) leakage current through generation-recombination centres, and (c) creation of trapping centres.

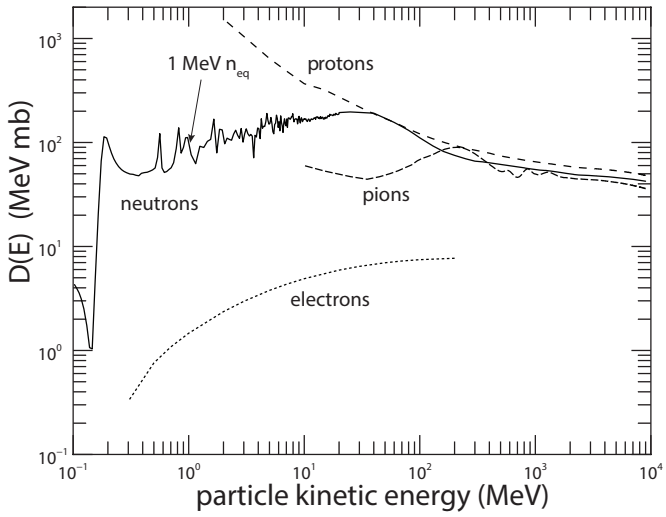


Fig. 8.82 Damage function $D(E)$ for atom dislocations in a Si-lattice for different particles: neutrons, protons, pions, electrons (after [661], with kind permission of Elsevier). The normalisation point for 1 MeV neutrons is marked by the arrow.

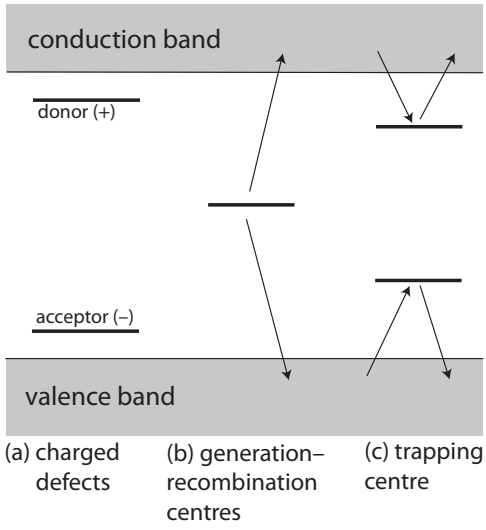


Fig. 8.83 Characteristic locations of energy levels of substrate defects caused by radiation damage.

Mathematically the consequences of these lattice defects are worked out in the framework of the so-called *Shockley–Read–Hall* (SRH) statistics [493, 894] describing generation-recombination processes with intermediate defect levels inside the band gap acting as 'stepping stones' in the process (see also e.g. [486,929] and [714]). This model has been quite successful in describing a large variety of phenomena in semiconductors.

SRH statistics can be straight forwardly understood by considering that (a) the probabilities for transitions from/to defect levels E_t (for simplicity called 'trap levels' or 'trapping centres' (TC)) into the conduction (CB) or valence bands (VB), respectively, are proportional to their abundance and their occupation probabilities, and (b) that the capture/emission probabilities follow exponential energy laws. As illustrated in fig. 8.84, we consider rates for capturing of electrons from the conduction band to the trapping centre CB $\rightarrow$ TC ($r_{c,n}$), emission of electrons TC $\rightarrow$ CB ($r_{e,n}$), as well as the corresponding processes for holes, capturing VB $\rightarrow$ TC ($r_{c,p}$) and emission TC $\rightarrow$ VB

($r_{e,p}$). The signal injection rate is denoted R_{inj}.

In the following we follow [486] and calculate first the rates of these four processes, again considering n-bulk silicon. The electron capture rate $r_{c,n}$ must be proportional to the electron concentration n_n in the CB and the concentration of centres/traps n_t which are not (yet) occupied:

$$r_{c,n} \propto n\, n_t\, (1-f)\,,$$

where $f = (1+\exp\frac{E_t - E_F}{kT})^{-1}$ is the centre occupation probability by electrons following Fermi statistics, as discussed in section 8.2.3. The capture rate for holes correspondingly is $r_{c,p} \propto p_n\, n_t\, f$, $p_n =$ hole concentration in VB, since hole transition from the VB to the centre is equal to electron transition from the centre to the VB and hence proportional to f. The proportionality constants are given by the respective products of thermal velocities $v_{th,(n,p)}$ and capture cross sections $\sigma_{n,p}$ which have here atomic dimensions of order $10^{-15}\,\mathrm{cm}^2$. We then obtain for the capture rates:

$$\begin{aligned} r_{c,n} &= c_n\, n_n\, n_t\, (1-f)\,; & c_n &= v_{th,n}\, \sigma_n\,; \\ r_{c,p} &= c_p\, p_n\, n_t\, f\,; & c_p &= v_{th,p}\, \sigma_p\,, \end{aligned} \tag{8.102}$$

with $c_{n,p}$ being the capture coefficients. The emission rates $r_{e,n}$ (TC $\rightarrow$ CB) and $r_{e,p}$ (TC $\rightarrow$ VB) are readily calculated accordingly:

$$\begin{aligned} r_{e,n} &= e_n n_t\, f \\ r_{e,p} &= e_p n_t\, (1-f) \end{aligned} \tag{8.103}$$

with $e_{n,p}$ being the yet to determine emission probabilities. They can be determined by considering that in the equilibrium case (i.e. without external injection) $r_{c,(n,p)} = r_{e,(n,p)}$ must hold. Hence equating (8.102) and (8.103) and using the equilibrium carrier concentrations of (8.27) and (8.28) we obtain the emission probabilities:

$$\begin{aligned} e_n &= c_n\, n_i\, \mathrm{e}^{(E_t - E_f)/kT} \\ e_p &= c_p\, n_i\, \mathrm{e}^{(E_f - E_t)/kT}\,, \end{aligned} \tag{8.104}$$

where E_t is the energy of the defect state and E_f is the intrinsic Fermi level.

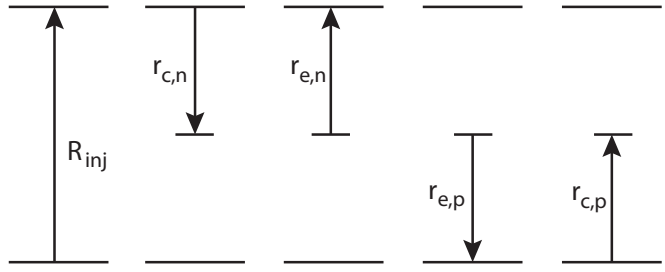


Fig. 8.84 Injection, emission and capture rates occurring in the Shockley–Read–Hall model under non-equilibrium conditions (adapted from [486]), that is, with charge injections.

If charge injection takes place, for example incurred by a traversing particle, non-equilibrium processes are introduced (VB → CB transitions). In steady state,[23] the rate by which electrons (holes) enter the CB (leave the VB) will equal the rate of electrons leaving the CB (enter the VB) and therefore

$$\begin{aligned} \frac{dn_n}{dt} &= R_{\text{inj}} - (r_{c,n} - r_{e,n}) = 0\,, \\ \frac{dp_n}{dt} &= R_{\text{inj}} - (r_{c,p} - r_{e,p}) = 0\,, \end{aligned} \tag{8.105}$$

with R_{inj} being the injection rate. For steady state non-equilibrium conditions we thus have

$$r_{c,n} - r_{e,n} = r_{c,p} - r_{e,p} \tag{8.106}$$

and can substitute the respective rates from (8.102) and (8.103) and solve for the occupancy factor f. Note that while neither the expression for $f = (1 + \exp\frac{E_t - E_F}{kT})^{-1}$ nor for n and p in (8.27) and (8.28) are meaningful under non-equilibrium conditions, under injection the electron and hole concentrations will depend on the injection level thus making f also dependent on R_{inj}:

$$\begin{aligned} f &= \frac{c_n\, n_n + e_p}{c_n\, n_n + e_n + c_p\, p_n + e_p} \\ &= \frac{c_n n_n + c_p n_i \mathrm{e}^{(E_f - E_t)/kT}}{c_n \left(n_n + n_i \mathrm{e}^{(E_t - E_f)/kT}\right) + c_p \left(p_n + n_i \mathrm{e}^{(E_f - E_t)/kT}\right)}\,. \end{aligned} \tag{8.107}$$

By substituting these values of f into the rates appearing in (8.106) we obtain the steady-state net recombination rate (generation rate) through the action of the intermediate levels:

$$\begin{aligned} G_t &= r_{c,n} - r_{e,n} = r_{c,p} - r_{e,p} \\ &= \frac{c_n c_p n_t \left(n_n p_n - n_i^2\right)}{c_n n_n + e_n + c_p p_p + e_p} \\ &= \frac{c_n c_p n_t (n_n p_n - n_i^2)}{c_n \left(n_n + n_i \mathrm{e}^{(E_t - E_f)/kT}\right) + c_p \left(p_n + n_i \mathrm{e}^{(E_f - E_t)/kT}\right)}\,. \end{aligned} \tag{8.108}$$

Simplifying by setting[24] $c_n = c_p = \sigma v_{th}$ and abbreviating $n_n = n$ and $p_n = p$ one obtains the form known in the literature (see e.g. [486])

$$G_t = \sigma v_{th} n_t \, \frac{np - n_i^2}{n + p + 2n_i \, \cosh\left(\frac{E_t - E_f}{kT}\right)}\,, \tag{8.109}$$

which shows that the deviation from equilibrium $np - n_i^2$ is the 'driving force' for recombination. G_t is smaller if the doping $n + p$ is large and decreases also if the cosh term becomes large by E_t moving away from the middle of the band gap E_f.

[23] Note that 'steady state' does not imply equilibrium.

[24] Note that both v_{th} and σ are different for electrons and holes due to their different effective masses (see [474]).

Consequences of substrate damage for Si detectors. We use the SRH formulation described in the previous paragraph for the three main effects on detector silicon by writing the defect occupation probability f of (8.107) as

$$f = \frac{r_{c,n}n + r_{e,p}}{r_{c,n}n + r_{e,n} + r_{c,p}p + r_{e,p}} \approx \frac{r_{e,p}}{r_{e,n} + r_{e,p}}\,, \tag{8.110}$$

where the approximation assumes that the carrier densities are very low in the depletion region of a detector [714]. We then can identify three main effects arising from the damage:

(1) *Change of the effective doping concentration*
While in undamaged silicon the bulk doping with the much larger concentration (that is e.g. boron, N_A, or phosphorous, N_D) govern the effective space charge density $\rho = e(N_D - N_A) = eN_{\text{eff}}$, for damaged silicon the *effective doping concentration* or better the *effective space charge density* of the detector bulk silicon changes as a function of the duration of the damaging radiation. On the one hand, radiation can deactivate donor atoms (like P or As) or acceptor atoms (like B). The most likely donor removal process is their deactivation by capturing mobile single vacancies resulting in V–P or V–As complexes (E-centres, see fig. 8.79), whereas boron acceptors are mainly removed by becoming first interstitial boron B_i (substitutional B_s removed from its lattice position by an Si_i) which can be captured by intestitial oxygen O_i forming a B_iO_i complex [580].[25] On the other hand, new donor- or acceptor-like states can be created as discussed above.

The concentration change is proportional to the radiation fluence. Even a sign change (*type inversion*) of the space charge is possible (see also section 8.7), in which initial n-type silicon becomes effectively more p-type with continued irradiation (fig. 8.85). For high ohmic substrate material this can already occur at fluences of $\phi_{eq} \approx 10^{12-13}\, n_{eq}\,\text{cm}^{-2}$ and is predominantly caused by acceptor-like defects that trap electrons from the conduction[26] band, thus contributing to negative space charge similar to p doping (see e.g. [786]).

The space-charge concentration in the depletion zone changes and hence also the voltage necessary for full depletion of the silicon volume. The effective space charge N_{eff} is given by the sum of positively charged donors and negatively charged acceptors

$$N_{\text{eff}} = \sum_D (1 - f)\, n_t - \sum_A f\, n_t \tag{8.111}$$

where n_t is the centre density as in (8.102) and f is given in (8.110).

At the fluence levels encountered by the LHC experiments, type inversion occurs, depending on the luminosity of the accelerator this may be already after a few years of operation (see also section 8.7). With the change of the effective space charge with fluence also the bias voltage needed for a complete depletion of the detector volume is lowest at the inversion point. Since the effective space charge continues to increase with radiation (normally it becomes more negative), the full depletion voltage also increases again (fig. 8.85, y-axis on the left). In the extreme case it exceeds the maximum allowed voltage of the power supplies,

[25] Also B_iB_s defect complexes can be formed.

[26] At room temperature thermal energy is in general insufficient to fill these levels with electrons from the valence band.

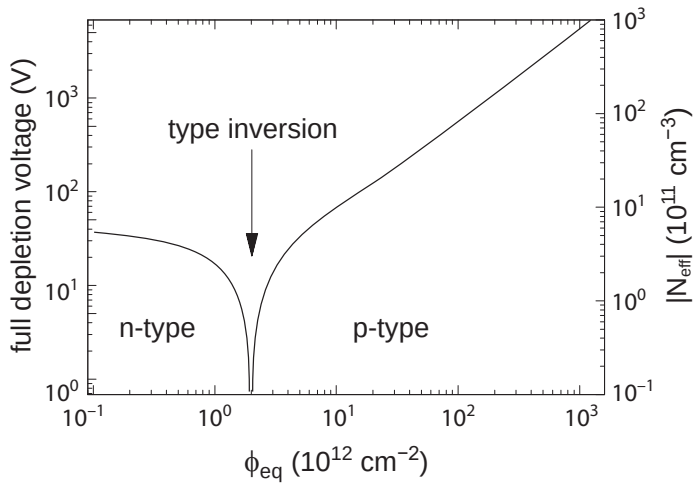


Fig. 8.85 Effective charge carrier concentration (right-hand y-axis) and full depletion voltage (left hand y-axis) as a function of the particle fluence ϕ_{eq}, normalised to the damage caused by 1-MeV neutrons (after [1008]).

such that the detectors can only be operated in partial depletion, hence leading to a reduction in signal charge.

(2) *Leakage current*
Defect levels near the middle of the band gap act as generation-recombination centres. Due to their shorter distance to the respective bands (CB and VB) excitation of an e/h pair (generation) is much easier from these levels due to the exponential energy dependence. Conversely, an electron (in the CB) and a hole (in the VB) can recombine in such a state. The generation rate G_t of (8.109) can for an individual defect type and neglecting free carriers be written in the simplified form for f (eq. (8.110)) by means of emission rates only [714]:

$$G_t = n_t\, f\, r_{e,n} = n_t\,(1-f)\, r_{e,p} = \frac{e_n e_p}{e_n + e_p}\,. \tag{8.112}$$

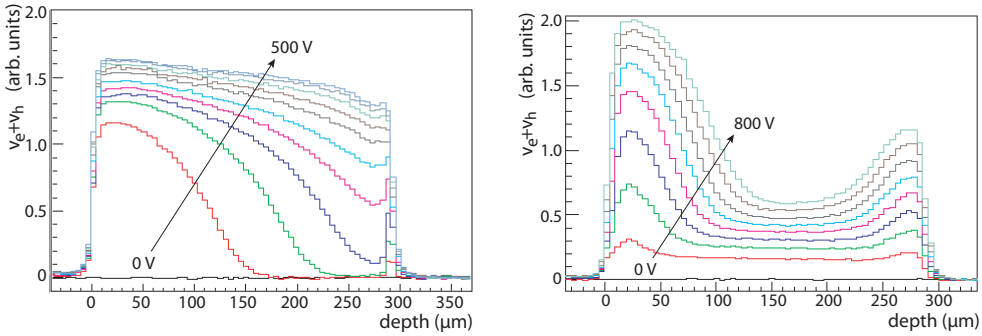
The generation rate leads to detector *leakage current* which per electrode is given by the sum over all centres in an 'active' volume V under the electrode:

$$I_L = e\,V \sum_{\text{defects}} G_t\,, \tag{8.113}$$

where e is the elementary charge.

(3) *Trapping*
Electrons and holes can be trapped by defect levels and be released again after some time. Consequences are decreased carrier lifetimes and carrier mean free path lengths, resulting in decreased signal amplitudes if the detrapping time constant is longer than the time of signal formation.

Carrier trapping is described by an effective trapping time $\tau_{\text{eff},e/h}$ (inverse capture rate) given as



(a) Non-irradiated detector (b) Irradiated detector ($\phi_{eq} = 10^{16}\,\mathrm{cm}^{-2}$)

Fig. 8.86 Development of a 'double junction' profile in irradiated silicon detectors. Sum of the drift velocities of both charge carrier types as a function of depth measured employing the edge-TCT technique (see text). (a) Detector before irradiation. The arrow indicates bias voltages raised from zero to 500 V in steps of 50 V. (b) Detector after irradiation to a fluence of $\phi_{eq} = 10^{16}\,\mathrm{cm}^{-2}$. Bias voltages are raised from 0 V to 800 V in steps of 100 V (from [631] with kind permission of IOP).

$$\frac{1}{\tau_{\mathrm{eff},e}} = \sum_{\text{defects } i} c_{n,i}\,(1-f)\,n_t \qquad \text{and} \qquad \frac{1}{\tau_{\mathrm{eff},h}} = \sum_{\text{defects } i} c_{p,i}\,f\,n_t \tag{8.114}$$

where $c_{n,i}$ and $c_{p,i}$ are the individual electron and hole coefficients of defect i according to (8.102).

8.12.2 Implications for the detector operation

Change of the substrate doping. Macroscopic consequences of radiation damage manifest themselves as changes in the detector response. The change of the effective doping concentration potentially leads to type inversion, as described above. This also means a change of the electric properties of the detector. In silicon detectors the shape of the electrical field changes from that of an undamaged detector (see fig. 5.10 on page 148), because the distribution of the space charge in the depletion region is no longer homogeneous [713]. In fact, a kind of a double junction builds up with higher field strengths at both ends than in the centre of the detector substrate [378,631]. This can be seen in fig. 8.86 where the sum of the drift velocities $(v_e + v_h)$ is measured as a function of the depth from the electrodes when shooting a thin laser beam into the polished edge of a detector parallel to the strips or pixel electrode plane (edge-TCT technique [632]). The initial (before trapping takes place) induced pulse (see eq. (5.26) on page 133) is proportional to the drift velocities:

$$i_S(x, t \sim 0) \approx eN_{e/h}\frac{v_e + v_h}{d}\,, \tag{8.115}$$

where $eN_{e/h}$ is the total drifting charge and $E_w \approx 1/d$ is the approximate weighting field. Notice, however, that for detector operation, except for adjustments to the height of the applied bias voltage, no other parameters need to be changed to operate the detector 'through type inversion'.

The operation voltage necessary for full depletion of a Si detector with thickness d is directly proportional to the doping concentration

$$V_{dep} \approx \frac{e|N_{\text{eff}}|d^2}{2\epsilon\epsilon_0}\,, \tag{8.116}$$

where[27] the original doping concentration $N_{A,D}$ of (8.59) is replaced by the effective doping concentration N_{eff}. At very high fluences large bias voltages are needed ($\mathcal{O}$(500–1000 V)) to reach full depletion of the detector and for fluences above 5×10^{15} cm^{-2} full depletion for standard bulk thickness is in general no longer achievable. In the context of the development of very radiation hard silicon detectors for the high-luminosity upgrade of the LHC accelerator [285], sensors fabricated from p-type wafers are used. It has been found[28] that n$^+$-in-p sensors are more radiation tolerant than p$^+$-in-n sensors [279] and can be more easily (and more cheaply) fabricated than n$^+$-in-n sensors for which double-sided processing is necessary [713, 680]. Sensors fabricated from p-type bulk generally[29] also do not 'type invert' as the original space charge is already negative before irradiation.

Leakage current. The creation of deep-level generation centres increases the detector leakage current proportional to the fluence, as fig. 8.87 demonstrates for various silicon sensor processing techniques and resistivities:

$$\Delta I_L = \alpha\ \phi_{eq}\, V\,. \tag{8.117}$$

Here V is the volume under the electrode, ϕ_{eq} the (1 MeV) neutron equivalent fluence, and α the (current related) *damage rate*, which according to (8.117) can be determined by measuring the change in leakage current as a function of radiation fluence. For a fixed annealing (see section 8.12.4) duration and fixed temperature (see section 8.12.4) α is a universal constant when normalised to a certain temperature (20 °C), neither depending on the characteristics of the silicon material (crystal growth) nor on the doping or the type of the damaging radiation (neutron, proton, pion) [660]. The value is $\alpha_{80/60} = 4.0 \times 10^{-17}$ A cm^{-1} ($\pm$ 5%) for 80 min annealing time at 60 °C, conditions that were fixed by convention [660].

The strong temperature dependence of the leakage current has been discussed in section 8.3.2 on page 284. In origin it is dominated by deep-level defects residing in the middle of the band gap $E_G/2$ (see also fig. 8.83 (case b)):

$$I_L^{gen} \propto T^2 \exp\left(-\frac{E_a}{2kT}\right)\,, \tag{8.118}$$

where the parameter E_a, called *activation energy*, replaces E_G in the empirical description having best-fit values around $E_a = 1.19 - 1.21$ eV [474, 308, 1005].

Leakage current leads to increased electronic noise (*shot noise*, see section 17.10.3.1) in the readout chain. The current flow also heats up the detector and the danger of 'thermal runaway' exists, a chain reaction caused by a leakage current increasing the temperature, which in turn causes more thermally generated current, which heats up further, and so on. Thermal runaway can ultimately destroy a detector.

[27] Equation (8.116) is only valid if N_{eff} is constant throughout the sensor volume. For radiation damaged silicon the space charge density becomes non-uniform because free charge carriers from high dark currents can change the occupation of the defects.

[28] The reason that p$^+$-in-n sensors have been chosen for the first generation of semiconductor sensors at the LHC is that pure p-type wafers were difficult and expensive to obtain.

[29] Exceptions under special fabrication and irradiation conditions have been observed [714].

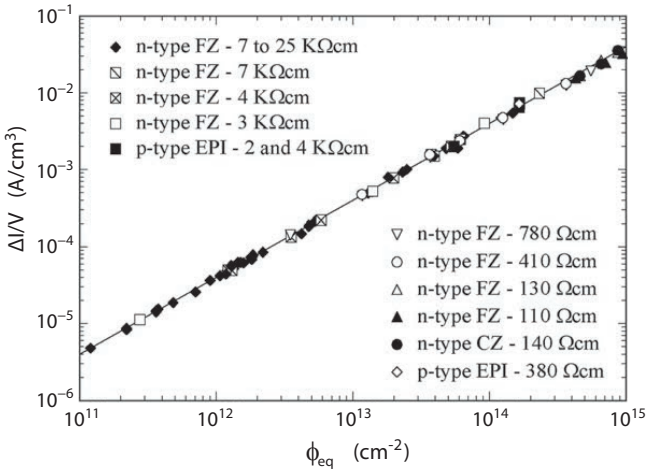


Fig. 8.87 Linear dependence of radiation induced volume leakage current on particle fluence ϕ_{eq}, normalised to the damage caused by 1 MeV neutrons, for different silicon resistivities and fabrication techniques. Measured at 60° C after annealing for 80 min [712].

Carrier trapping. In silicon, even to fluences of $\phi_{eq} = 10^{15}\,\mathrm{cm}^{-2}$, the charge collection in a depleted volume is close to 100% efficient. In other sensor materials (for particle detectors in high radiation fluxes diamond is particularly relevant here, see section 8.13.4), and also in silicon after very high fluences (typically well above 10^{15} $\mathrm{n}_{eq}/\mathrm{cm}^2$) charge trapping is dominantly responsible for the deterioration of the detector response after radiation.

An effective time $\tau_{\mathrm{eff,e/h}}$ describes charge loss due to trapping:

$$Q(t) = Q_0 \exp\left(-\frac{\mathrm{t}}{\tau_{\mathrm{eff,e/h}}}\right), \tag{8.119}$$

where an empirical fluence dependence is used for $\tau_{\mathrm{eff,e/h}}$ as a parametrisation:

$$\frac{1}{\tau_{\mathrm{eff,e/h}}} = \frac{1}{\tau_0} + \beta_{e/h}\,\phi_{eq}\,, \tag{8.120}$$

($\tau_0 =$ lifetime before damage) thereby defining the *effective trapping time* constant $\beta_{e,h}$. The larger $\beta_{e/h}$ the more severe is the damage. In contrast to α of (8.117), which relates to carrier emission, $\beta_{e/h}$ relates to carrier capturing. While α is strongly temperature dependent (depends on n_i) and must be quoted under defined conditions (see above), $\beta_{e/h}$ depends only little on temperature. At 293 K it has values $\beta_{e/h} \approx 4 \times 10^{-16}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [633].

Typical trapping times and distances $\lambda_t = v_{\mathrm{drift}}\,\tau_{\mathrm{eff,e/h}}$ at a fluence of $\phi_{eq} = 10^{15}\,\mathrm{n}_{eq}/\mathrm{cm}^2$ are [633]: $\tau_{\mathrm{eff,e/h}} = 2.5\,\mathrm{ns}$; $\lambda_t \approx 125 - 250\,\mu\mathrm{m}$, assuming a typical electron drift velocity of $v_{\mathrm{drift}} \approx 50 - 100\,\mu\mathrm{m/ns}$, which inversely scales with ϕ_{eq} according to (8.120).

Charge collection distance and charge collection efficiency. For materials for which the mean free path λ for the charge carriers is small compared to the thickness of the device, it is useful to define quantities that characterise the average distance over which charges can reach the electrodes. This situation is given for example for GaAs or CVD diamond substrates (see section 8.13) and—with some restriction—also for heavily damaged silicon.

The total induced signal is determined by the duration of the separation movement of electrons and holes, which in the presence of charge losses is governed by the average lifetime $\tau_{e/h}$ until trapping occurs. During this time the charge carriers separate by the distance

$$d_{sep} = v_e\tau_e + v_h\tau_h = (\mu_e\tau_e + \mu_h\tau_h)\ E = \mu\tau E\,, \tag{8.121}$$

with

$$\mu = \frac{\mu_e\tau_e + \mu_h\tau_h}{\tau} \tag{8.122}$$

being the lifetime weighted mobility. The distance d_{sep} is the sum of the mean free path lengths $\lambda_{e/h}$ of both carrier types and is also called *charge collection distance* (CCD) in a detector with thickness d. For $d_{sep} \ll d$, the ratio d_{sep}/d corresponds to the fraction of the maximum induced signal. In this case CCD can be very simply determined experimentally:

$$\text{CCD} \approx \frac{Q_{det}}{Q_{ion}}\, d = \text{CCE} \times d\ . \tag{8.123}$$

Here Q_{ion} and Q_{det} are the produced ionisation and induced (detected) charges, respectively, and their ratio is called the *charge collection efficiency* (CCE).

More generally we can derive the ratio of detected and deposited ionisation charge from the rate (continuity) equation (8.46) on page 278 with some simplifying assumptions which are particularly valid for mono crystalline diamond, but only very limited in silicon. We calculate the total induced signal due to the drift motion of deposited charges [1021], first considering only electron carriers with density n and assuming that the excess carriers, generated for example by a passing particle track, dominate the total number of free carriers ($n \gg n_i, N_D$). We can then write (8.46) as

$$\frac{\partial n}{\partial t} = -R_n(t) = -\frac{n}{\tau_e}\,, \tag{8.124}$$

where $R_n(t)$ is the recombination rate mostly due to charge trapping and τ_e is the trapping time. The generation rate $G_n(t)$ in (8.46) can be set to zero for times $t > t_0$, since after instant charge deposition by the passing track at $t = t_0$ on the time-scale of picoseconds no further carriers are generated. Furthermore, a divergence-free drift current density $\vec{\nabla}\vec{j} = 0$ is assumed, that is space charge is neglected. The trapping time τ_e relates to the mean free path $\lambda_e = v_e\tau_e$ such that (8.124) becomes

$$\frac{dn}{n} = -\frac{dt}{\tau_e} = -\frac{dx}{v_e\tau_e} = -\frac{dx}{\lambda_e}\,. \tag{8.125}$$

Here and in the following we use a coordinate system as in fig 5.5 on page 137 for a detector of thickness d with electrons drifting in positive x direction towards the readout electrode and x_0 denoting the starting point of the drift motion. The solution to (8.125) is

$$n(x, x_0) = n_0 \exp\left(-\frac{x - x_0}{\lambda_e}\right) \tag{8.126}$$

with n_0 being the primary ionisation density. Employing the Shockley–Ramo theorem (5.26), derived in chapter 5 for a two-electrode system with weighting field $|\vec{E}_w| = 1/d$,

the induced charge is given by the drifting charge in the volume $V = A\,d$ of thickness d and area A:

$$\begin{aligned} dQ_n &= -\left(en(x,x_0)\,dV\right)\frac{1}{d}\,dx \\ \Rightarrow Q_n &= \int dQ_n = -\int_{V=Ad} dV' \int_{x_0}^{d} e\,n(x,x_0)\,\frac{dx}{d} \\ &= -\underbrace{en_0A}_{Q_0/d}\int_0^d dx_0 \int_{x_0}^{d} \exp\left(-\frac{x-x_0}{\lambda_e}\right)\frac{dx}{d} \\ &= -Q_0\,\frac{\lambda_e}{d}\left(1-\frac{\lambda_e}{d}\left(1-\mathrm{e}^{-d/\lambda_e}\right)\right), \end{aligned} \tag{8.127}$$

where we denote with Q_0 the total deposited ionisation charge along the track. This equation is also known as the Hecht relation [514]. The corresponding signal contribution from the hole movement is identical to (8.127) with λ_e being replaced by λ_h, such that we obtain:

$$\mathrm{CCD} = \frac{Q_n + Q_h}{Q_0}\,d \;=\; \sum_{i=e,h} \lambda_i \left(1-\frac{\lambda_i}{d}\left(1-\mathrm{e}^{-d/\lambda_i}\right)\right). \tag{8.128}$$

For $\lambda_{e,h} \ll d$ we retain (8.121)

$$\mathrm{CCD} = \lambda_e + \lambda_h\,. \tag{8.129}$$

The loss of signal due to charge carrier trapping changes the mean free path lengths of both carriers:

$$\frac{1}{\lambda_{e/h}(\phi)} = \frac{1}{\lambda_{e/h}(\phi=0)} + \beta_t\phi_{eq}\,. \tag{8.130}$$

Here we have defined with β_t another damage constant due to trapping. Different to β in (8.120), β_t does depend on the type of radiation (neutrons, protons, pions) and their energy; it can be determined from the signal loss as a function of the radiation fluence [950] using (8.123), (8.128) and (8.130): for diamond β_t lies in the range between about 3×10^{-18} cm^2 µm^{-1} (25 MeV protons) and 0.7×10^{-18} cm^2 µm^{-1} (24 GeV protons), respectively, and is about 2–3 times smaller than for silicon (see section 8.13.4).

8.12.3 Surface damage

As surface damage we denote damage to the surface of silicon sensors or the (MOS) electronics, especially the silicon dioxide layer and the Si–SiO_2 interface. Also here radiation causes ionisation and/or dislocations of lattice atoms. Different, however, to bulk damage discussed in the previous section causing serious effects for sensor operation, in the Si oxides of MOS transistors the damage impact from ionisation is more severe than from lattice dislocations because the troubling consequences are caused by charged defect states in the oxide or interface. NIEL damage, by contrast, does not become electrically active in the anyway amorphous structure of the oxide. As for bulk silicon in detectors, high electric fields can exist in the oxides of MOS transistor structures, especially in gate oxides. They separate the charge carriers created by ionisation and hence prevent their complete immediate recombination. The mobility

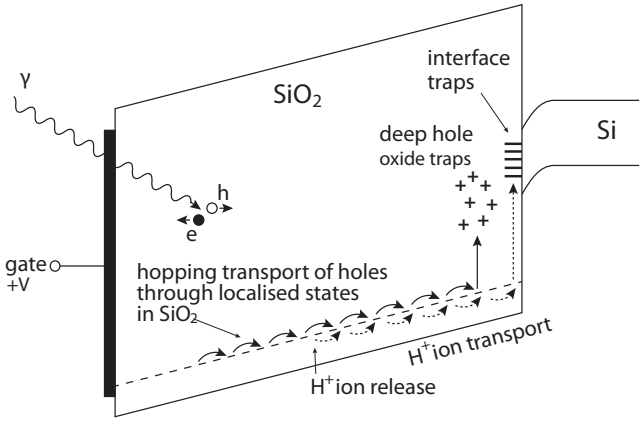


Fig. 8.88 Illustration of damage in SiO_2 by ionising radiation using a MOS band–diagram. Oxide traps (for holes, positive) and interface traps (negatively or positively charged) exist (see text). The hole 'hopping' movement creates ions from impurity (atomic) hydrogen which can become traps at the interface (adapted from [909])

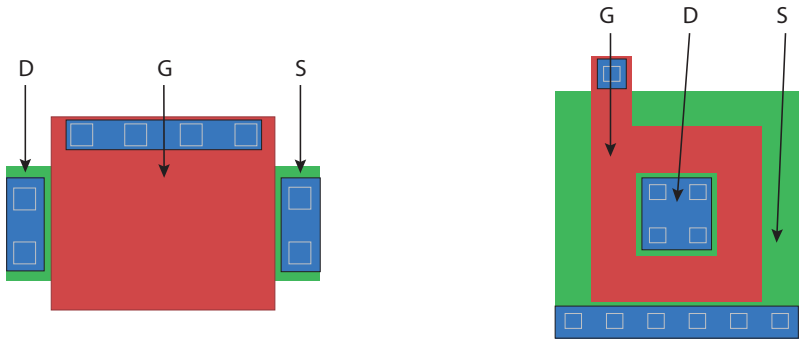
of holes in SiO_2, however, is about 10^6 times lower than the corresponding electron mobility, due to the large hole capture cross section by shallow levels in the silicon oxide. Holes therefore reside much longer in the oxide than electrons and move only very slowly.

Oxide and interface traps. Figure 8.88, which is a detail of the MOS band-diagram of fig. 8.30(b), illustrates the effects resulting from this particularity for holes [909, 876]. Due to their slow movement they tend to become 'self-trapped holes', that is, they are localised by their own effect on the lattice potential surrounding them[30] (see also section 13.3.1 on page 516). This local lattice distortion moves along with the hole movement, not steadily but hopping from one lattice location to the next, an effect known as 'polaron hopping', thereby increasing the holes' effective mass and further lowering their mobility. This polaron hopping makes the hole transport very dispersive, that is, it happens on very long time-scales compared to the charge injection, and it is very temperature and, in particular, oxide thickness dependent. If the applied bias to the gate is positive the holes will be transported to the Si–SiO_2 interface opposite to the gate. Near the interface the density of oxygen vacancies is large due to lattice mismatches between Si and SiO_2 near the interface and due to oxygen out-diffusion from the oxide. These oxygen vacancies can act as trapping centres for the arriving holes (*deep hole oxide traps*). The positive charge associated with trapped holes near the interface causes a negative transistor threshold shift in both NMOS and PMOS transistors.

It is also believed [876] that, induced by the hole hopping, impurity hydrogen ions (H^+) are likely to be released from the lattice (see fig. 8.88). They can also drift towards the interface where they may react forming interface traps, as discussed below.

A second ionisation induced effect is the mentioned formation of *interface traps* at the Si–SiO_2 interface [1002]. These traps (or interface states) are physically located at the interface bonds. They are defects with energy levels inside the silicon band gap which can act as acceptors (negative when filled) if residing in the lower part of the gap, as donors (positive when filled) if in the upper half of the gap, or can be amphoteric (both acceptor- or donor-like behaviour). The charge state of these levels depends on the applied bias. It has been found [909] that for n-channel MOSFETs

[30] The created strain field is known as a *polaron*.



(a) Linear layout (b) Enclosed layout

Fig. 8.89 Linear (a) and round (b) geometrical structures of a MOSFET transistor (*layout*) with labels *source* (S, green), *drain* (D, green) and *gate* (G, red). The blue areas mark the metal contacts for the S,G,D implants which lie underneath and are closed structures.

under positive bias the interface traps are predominantly negative, while for PMOS under negative bias they are charged positively.

Since oxide-trapped charges are positive, interface-trapped charges act compensating for n-channel transistors whereas they increase their impact on the operation of p-channel devices [909]. The time-scales for both charge build-ups are, however, different. Interface-trap build-up occurs on still much slower time-scales than oxide-trap build-up, up to about 10^5 s for saturation [876]. Both oxide and interface trapping effects create essentially stationary charges in the oxide and at the interface. Although their neutralisation starts immediately, the time-scales for complete annealing can reach many months. Annealing happens through electron tunnelling from the silicon into the oxide traps and/or thermal emission of electrons from the oxide valence band into the trap levels.

Influence on the transistor operation. The fixed oxide charges influence the threshold characteristics of MOS transistors according to eq. (8.80) in section 8.3.5 or can generate parasitic transistor structures, which can significantly disturb the functionality of integrated circuits [837]. The rapid progress of IC miniaturisation has very positively influenced the development of radiation hard electronics. Small feature sizes (~ 250 nm or less) in modern IC technologies also have very thin gate oxides, only a few nanometres thick, leading to a strong reduction in the trap building processes. As a consequence the importance of radiation-induced charge build-up rapidly decreases, such that their influence on the transistor characteristic is much reduced. However, similar effects in other oxides occurring in microelectronics (e.g. field oxide, shallow trench isolation oxide) still exist and can cause effects accordingly in narrow channel or short channel MOS transistors (see e.g. [389]).

At critical transistor circuit nodes one can replace the usual *linear* transistor geometry of fig. 8.89(a)—that is the *gate* is realised by a rectangular implant structure in between *source* and *drain*—by an *enclosed* geometry, by which the *drain* implant is surrounded by the *gate* implant and then the *source* implant shown in fig. 8.89(b). Parasitic currents, which in linear geometry can pass by the *gate* and hence lead to unwanted currents elsewhere (parasitic transistors), are also strongly suppressed in enclosed structures.

8.12.4 Measures for radiation hardening

The operation of semiconductor detectors in radiation intense environments is a major research and development focus, in particular since the deployment of such detectors for the LHC experiments operating at very high luminosities is a necessity.

All of the above mentioned damages and their development with time are temperature dependent. Raising the crystal's temperature can for example cause the electrical activation of defects. The treatment of irradiated samples with heat causing physical changes in the solid is called *annealing*. In the case of silicon, the initial effective decrease of acceptor-like states N_{eff} when keeping the damaged sensor for a few hours at about 60°C is called 'short-term' or 'beneficial annealing' and is likely caused by an increase of donor-like defects rather than by acceptor removal [784]. If one keeps the crystal at this higher temperature for too long, however, other defects, acceptor-like, which so far did not influence detector operation, become active, thus counteracting the beneficial effect of the initial annealing. The latter is called *reverse annealing* [660]. The behaviour of these effects with time after irradiation is shown in fig. 8.90.

To describe the different time and temperature dependent annealing processes the following expression for the change in effective doping concentration (effective space charge) is used with three terms, which can each again be parameterised [1008], also [837, 714]:

$$\Delta N_{\text{eff}}(t;\phi,T) = N_{\text{eff}} - N_{\text{eff}}^{\phi=0} = N_A^{\text{ben}}(t;\phi,T) \;+\; N_C(\phi) \;+\; N_Y^{\text{rev}}(t;\phi,T)\,, \tag{8.131}$$

where $N_{\text{eff}}^{\phi=0}$ is the value before irradiation, ϕ and T are fluence and temperature, and the three terms account for the two annealing contributions, N_A (beneficial) and N_Y (reverse), respectively, and also include a stable component N_C that does not change with time after irradiation and largely accounts for dopant removal. Figure 8.90 shows that ΔN_{eff} is measured to be positive, meaning that the radiation-induced damage produces an overall negative space charge, in agreement with the observed type inversion in n-type bulk silicon.

The three contributions for the time development of the incurred damage are parameterised as [714]:

$$\begin{aligned} N_A(t;\phi,T) &= g_A \exp\left(-\tfrac{t}{\tau_A}\right)\phi\,, \\ N_C(t;\phi) &= N_{C,0}\left(1-\exp\left(-c\phi\right)\right) + g_C\phi\,, \\ N_Y(t;\phi,T) &= g_Y\left(1-\exp\left(-\tfrac{t}{\tau_Y}\right)\right)\phi\,. \end{aligned} \tag{8.132}$$

Here g_A, g_C and g_Y are respective introduction rates for the space charges, for instance from reverse annealing $N_Y = g_Y\phi$. $N_{C,0}$ accounts for the fact that often an incomplete dopant removal is observed. The annealing time constants τ_A and τ_Y are temperature dependent, following a so-called *Arrhenius* equation,[31] $\tau_{A,Y} \propto \exp(E_a/T)$, with activation energies E_a of 1.09 eV and 1.33 eV, respectively [712], rendering cooler temperatures to cause much longer time constants. To suppress *reverse annealing* the silicon detectors at the LHC experiments are held at −6 to −10 °C when operated and also during time intervals without beam.

[31] In chemistry kinetics the Arrhenius equation describes the temperature dependence of a reaction rate constant used to analyse the effect of temperature on the reaction rate.

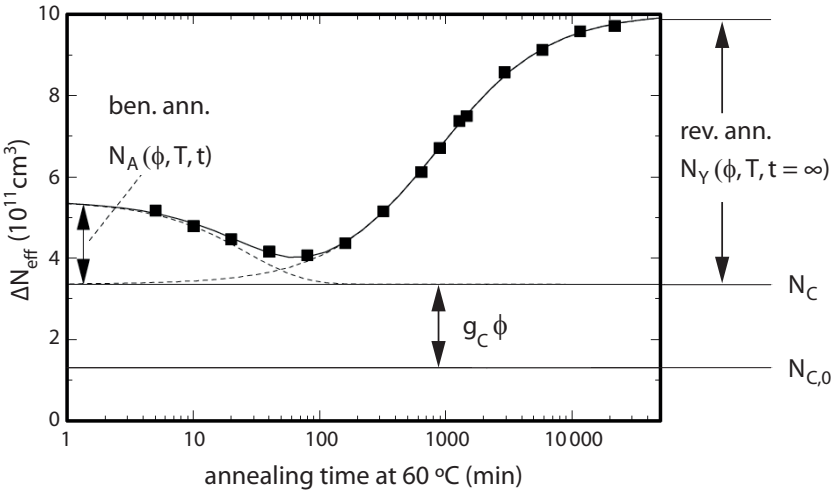


Fig. 8.90 Behaviour of the effective doping concentration upon annealing at a temperature of 60 °C (after [715], with kind permission of Elsevier). The y-axis shows the difference of N_{eff} before and after irradiation. The decrease of the effective doping concentration on the left hand side of the figure is known as 'beneficial annealing', the rise on the right-hand side as 'reverse annealing'.

Measures to increase the radiation tolerance of silicon detectors are:

- After type inversion the detector must still be functional. It is advantageous if after type inversion the n^+p boundary is located at the side of the readout electrodes. Since the depletion region grows from there into the bulk the region near the electrodes will still be depleted when for other parts of the sensor complete depletion can no longer be maintained. This can for instance be achieved by starting with n^+-in-n sensors (or also n^+-in-p, since p-type silicon usually does not undergo type inversion) with n implants for pixel or strip electrodes embedded in weakly doped n or p substrate (see fig. 8.55 in section 8.7). After the inversion the readout electrodes are then at the side of the n^+p diode and can also be operated underdepleted.
- Due to the high voltages needed for full depletion of the sensors after irradiation, the potential must be brought down towards the edge of the sensor by appropriate guard ring structures.
- If wanted, some stronger n-substrate doping can be chosen to protract the type inversion point.
- In order to avoid high fields at n^+n boundaries, so-called p-spray techniques are used rather than p-stop implants in between strips or pixels (see section 8.6.1).

Many implications of radiation damage increase with the area of the readout electrode and hence favour pixel over larger area pad or strip detectors. Furthermore, the signal-to-noise ratio is comparatively high for pixel detectors, because in addition to the smaller input capacitance that pixels pose to the preamplifier (see section 17.10), the leakage current, which is proportional to the volume underneath the electrode, is also reduced and thus contributes less to the electronic noise than do larger area electrodes.

Oxygen enriched silicon. Intensive studies [660, 505, 786] have shown that enrichment of silicon wafers with oxygen has a positive effect on the detector operation.

Oxygen content can be enriched for example by crystal growth in an oxygen-rich atmosphere using a quartz crucible (Czochralski pulling method). Depending on the type of radiation it can suppress or even prevent type inversion. The physical origin traces back to complex interactions of donor- and acceptor-like defects [786]. In simple terms one may say that the lattice vacancies created by the impinging radiation can form for example V–P complexes (vacancy-phosphorus), thus removing the phosphorus atom as a donor. If, however, oxygen atoms are more abundant than phosphorus atoms, 'harmless' V–O complexes are more likely to be formed instead, hence suppressing the donor removal. Similarly, in p-type material a defect labelled B_i–O_i (boron-oxygen interstitial at $E_t = E_C - 0.23\,\text{eV}$ [685], see fig. 8.91) effectively removes acceptors (boron) and thus influences the space charge in p substrates [580, 360].

To understand the contribution and effects of the various defects one must be able to study point defects and defect clusters separately. As γ radiation, for example by a ^{60}Co source, cannot generate clusters it can, for example, be used to identify observed defects as point defects. Dedicated methods like the *deep level transient spectroscopy* (DLTS) technique or the *thermally stimulated current* (TSC) technique are used to determine the position of defect levels in the band gap. The chemical composition of defects can be resolved by *transmission electron microscopy* (TEM). A concise overview of these methods can be found in [506].

Figure 8.91 shows the energy levels of the most important point and cluster defects. Defects can be electrically charged or neutral. For example, a donor level will be neutral, if it has an electron (is occupied or has captured an e^- or is 'not ionised'), or is positively charged if it is ionised (emitted an electron into the CB and is hence 'empty' of electrons); an ionised acceptor level is occupied by an electron (has emitted a hole) and is hence negatively charged; it is neutral when it is empty (not ionised). Also doubly charged defects (++ or –; for example the *bistable donor* BD in fig. 8.91) and defects with more than two charge states (e.g. $0, +, ++$) exist as well as defects which have (structure dependent) different energy levels which then can act either as donor or acceptor.

Prominent point defects are: (a) the so-called I_p defect [648, 786], the structure of which is still unknown, corresponding to an acceptor-like state near the band centre ('deep' acceptor); (b) a 'shallow' donor in the upper half of the band gap appearing in two stable charge states, hence being called BD-centre (*bistable donor*) and being copiously generated in oxygen-rich float zone silicon.

Now it is important that the I_p centres become less abundant with a high oxygen concentration, whereas the BD centre is enhanced in oxygen-rich material. The deep acceptor I_p creates a (p-type) negative space charge and also contributes to an enhancement of the leakage current via the generation/recombination mechanism (see fig. 8.83). The shallow BD donors are responsible for an increase of positive space charge (n-type). In oxygen-enriched material the creation of BD centres dominates and hence overcompensate the influence of the I_p centres on the space charge. Therefore, oxygen-enriched detectors with n-substrate material keep their overall positive space charge even at high radiation doses,[32] and do not type-invert.

Irradiation of silicon with hadrons produces defect clusters (see section 8.12.1 on page 342) which change the detector performance in addition to point defects. These

[32] Tested up to 5 MGy [786].

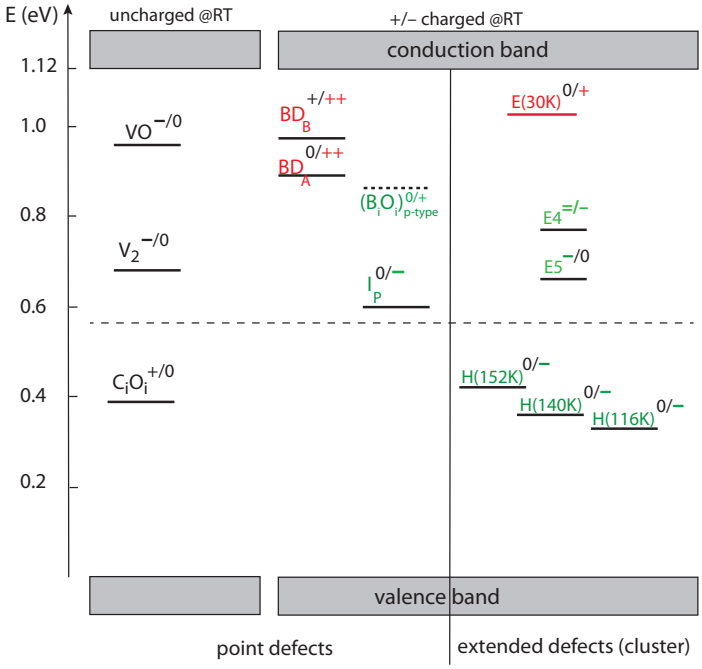


Fig. 8.91 Radiation-induced defects in silicon (drawn from information given in [713] and [802]) Shown are *point defects* (left) and *cluster defects* (right). The +,-,0 labels as superscripts indicate the possible charge of the states before/after emission or capture (also doubly charged ++ or =). Relevant for the effective space charge are electrically active defects. V = *vacancy*. See also explanations in the text. (@ RT = at room temperature).

clusters are characterised by several defects,[33] which are acceptors with levels in the lower half of the band gap and thus acting as hole traps. They contribute fully with negative space charge to N_{eff}. In addition a shallow donor defect E(30K) is created with a large cross section upon irradiation with charged hadrons, which partly compensates the p-type character of the change in space charge after irradiation.

For neutrons, however, point defects—being dominantly created by Coulomb scattering—play a minor role relative to cluster defects (see table 8.4). As a consequence the E(30K) centre is created much less strongly as with proton or pion irradiation and the acceptor-like defects remain effective. Therefore, after radiation damage of oxygen-enriched material with neutrons one does not observe any difference to standard (oxygen-lean) material (fig. 8.92), whereas for proton and pion irradiation the change in space charge is significantly smaller than for material without oxygen enrichment. With sufficient oxygen concentration type inversion can even become completely suppressed.

Also the phenomenon of *reverse annealing* is assigned to the same acceptor (cluster) defects as origin, as their abundance increases with annealing time [785]. In the LHC tracking detectors in regions near the beam, especially pions and protons are responsible for radiation damage, since these particles are created more abundantly in the collisions than neutrons. In the outer parts of the trackers, however, neutrons are abundant originating from back scatters in the calorimeters, in which the primary particles undergo hadronic interactions with nuclei.

Columnar electrodes: 3D-Si pixel detectors. With a changed geometrical arrangement of the charge collecting electrodes in so-called '3D-Si' pixel detectors [597,

[33] H(116K), H(140K), H(152K) [786].

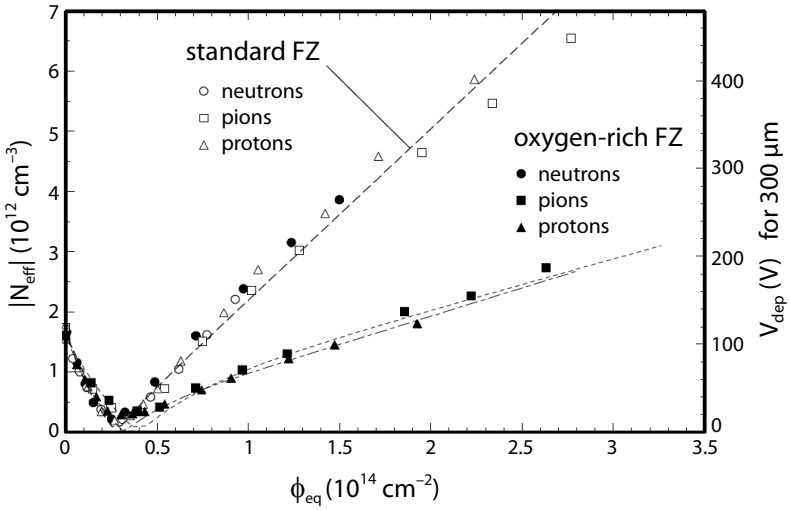


Fig. 8.92 Voltage necessary for full depletion of a 300 µm thick silicon detector as a function of the radiation fluence (1-MeV-neutron equivalent) for standard silicon material in comparison to oxygen-enriched silicon, fabricated in both cases using the *float-zone* technique (FZ). For the irradiation reactor neutrons as well as protons and pions from accelerators have been used. Apart from 'type inversion', already described in fig. 8.85, the figure shows that oxygen enrichment of the material leads to a significantly reduced effective p doping for protons and pions, whereas for neutrons no difference to standard FZ-material is observed (from [660], with kind permission of Elsevier).

334] (fig. 8.93) shorter drift distances and faster charge collection can be achieved, thus significantly increasing the radiation hardness. In contrast to conventional 'planar' pixel detectors (e.g. in fig. 8.35(c)) with electrode implants at the surface of the sensor, oriented perpendicular to particles at normal incidence to the sensor volume, in 3D-Si sensors particles typically run parallel to the electrodes. The distance between the columnar n electrodes and the p electrodes is 71 µm in both directions. Hence the sensor volume is fully depleted at comparatively small voltages, even after irradiation. The average field strength at typical bias voltages is larger than with planar detectors.

Since the mean drift distance $\lambda_{e/h}$ of the charge carriers becomes shorter with radiation fluence, signal decrease or even loss occurs when $\lambda_{e/h}$ reaches the same order of magnitude as the mean distance to the collection electrodes. For 3D-Si sensors this happens only at higher radiation fluences compared to planar sensors with a thickness of 250–300 µm. The measured signal loss for 3D-Si sensors after a fluence of for example $\phi_{eq} = 10^{16}\,\mathrm{cm}^{-2}$ is therefore smaller than for conventional silicon pixel detectors [335, 622]. For very high values of the electric field (>10 V/µm [955]) even an enhancement of the signal due to charge multiplication has been observed [622].

Using the same electrode fabrication technique the edges of the sensors can also be implanted as electrodes, leading to sensors with a very high active to passive volume ratio (see fig. 8.93).

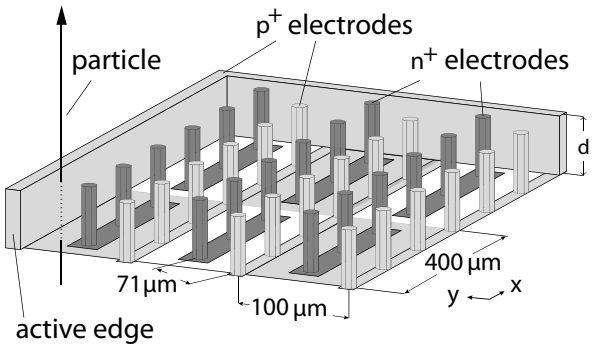


Fig. 8.93 Schematic drawing of a so-called 3D-Si pixel sensor [597] with a special electrode configuration (adapted from [699]). p^+ and n^+ implants are oriented vertically in the volume of the n^- substrate. The edges of the sensors are also fabricated with p^+ implants, such that the sensor remains sensitive for particles up to the very edge.

8.13 Other semiconductor materials

For most applications silicon is the semiconductor material best suited for detecting particles or radiation. Its characteristics are:

- In comparison to most other semiconductors silicon has a small density and a large radiation length, due to the low atomic number ($Z = 14$). While this is an advantage for particle physics experiments, the corresponding photon absorption probability is small (roughly $\propto Z^5$) and is disadvantageous when high photon detection efficiency is important, for example for X-ray detectors.
- As sensor substrates relatively large silicon wafers can be used,[34] having relatively few impurities and lattice inhomogeneities. This results in long carrier lifetimes and high charge collection efficiencies, such that charges generated by particles or radiation are usually all collected at the electrodes as long as these characteristics are not changed by radiation damage.
- Silicon is cheap and abundantly available. Its economic usage and industrial processing have matured over decades.
- Silicon can be n and p doped in a systematic and selective way. Hence the electric field in a sensor volume can be shaped for dedicated charge collection paths.
- The number of created e/h pairs per MeV deposited energy is comparatively large (about 23 000 e/h pairs for a minimum-ionising particle in a 300 µm thick detector).

Semiconductors other than silicon become attractive as detectors when special demands come into play that outbalance the advantages of silicon. Often a large absorption probability for high energetic ($>$ keV) photons is required, which is not achievable with silicon. Another special demand is radiation tolerance of the material, as for example at the LHC.

For X-ray detection compound semiconductors with high Z are attractive, in particular CdTe or CdZnTe ($Z \approx 50$) as well as GaAs ($Z = 32$) having much larger photon absorption cross sections than silicon (see section 3.5.1). Germanium (Z=32)is used exclusively in experiments where very good energy resolution of the absorbed radiation is required, for example in experiments seeking to detect weakly interacting particles in nuclear or in astroparticle physics (see chapter 16).

[34] Detector grade silicon wafers are available with typically 4" (10 cm), recently up to 8" (20 cm) diameter, IC wafers with up to 12" diameter.

For particle physics experiments artificial diamond is also an interesting material, even if it is classified as an insulator in the resistivity based definition of fig. 8.4. The band gap of diamond is 5.5 eV, resulting in a small charge signal for minimum-ionising particles in comparison to silicon. Because of its large radiation length, its high radiation tolerance, and also because of its excellent thermal conductivity, diamond is attractive as a particle detector base material, in particular in environments with high radiation fluences (see section 8.13.4).

In the following we give an overview over semiconductor materials which find applications as particle detectors besides silicon.

8.13.1 Germanium

As the only element semiconductor apart from silicon, germanium is also used for particle detection, albeit much less often. In comparison to silicon (see table 8.2) germanium has larger Z ($Z = 32$), a smaller band gap (0.7 eV) and higher carrier mobilities by a factor 2–3. The small band gap and, related to it, the smaller average energy needed for the creation of an e/h pair of $w_i = 2.96$ eV (and hence the larger number of charge carriers per energy deposit) are responsible for the better energy resolution of germanium compared to silicon. An impression of the remarkable energy resolution of Ge can be obtained from fig. 17.53 on page 778. For the same reason (smaller bad gap) germanium detectors must be cooled to keep the thermally generated leakage current small.

As a detector with high energy resolution germanium is used in nuclear physics and for photosensors in the near infrared, occasionally also for X-ray detection because of its large atomic number ($Z = 32$). As a position sensitive detector in particle physics germanium is rarely used. An exception are experiments requiring a high sensitivity for small signals, for example experiments on double-beta decay or to detect the nuclear recoil from a collision with a WIMP particle.[35] The expected energy deposited by ionisation of the (charged) recoil nucleus is in the order of 10 keV, which can be detected with Ge detectors as done for example in the CDMS-II experiment [518] (see section 16.7.2).

8.13.2 Gallium arsenide

Gallium arsenide (GaAs) is a III–V compound semiconductor. Its band structure is that of a direct semiconductor within which transitions between valance and conduction bands are possible without momentum transfer to the lattice (see section 8.2.2). At 1.43 eV the band gap is larger than in silicon (see fig. 8.14 and table 8.2). For radiation detectors again the large atomic numbers are interesting (Ga: $Z = 33$, As: $Z = 31$). GaAs contains 48.2 % of gallium and 51.8 % of arsenic in mass proportions.

Crystal growth of GaAs is more complicated than for silicon. The process techniques established for silicon must be modified such that vapour pressure equilibrium of the two parts is maintained during the melting. As with all compound semiconductors, fabricated GaAs crystals are impure comparison wise. They contain defects and impurities in the order of about $10^{15}\,\mathrm{cm}^{-3}$ leading to very short carrier lifetimes of only a few 10 ns (see table 8.2). The intrinsic carrier concentration without doping at room temperature is about $n_{\mathrm{intr}} \simeq 2 \times 10^6\,\mathrm{cm}^{-3}$. This is four orders of magnitude smaller than that of intrinsic silicon and is of similar order in number density as for

[35] WIMP = *weakly interacting massive particle*

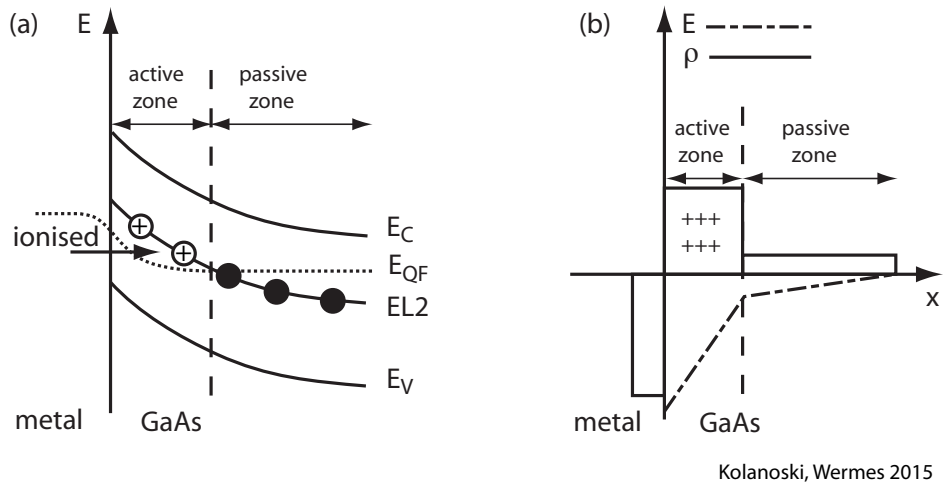


Fig. 8.94 Effect of the EL2 defect in a GaAs Schottky diode: (a) energy-band diagram under reverse bias. E_{QF} denotes the *quasi-fermi level*, which can be employed instead of E_F in non-equilibrium conditions and is not constant over the device (see also footnote 4 on page 279). (b) Resulting space charge density and electric field as a function of the substrate depth. High- and low-field regions lead to active and passive zones for charge collection.

Si in the depletion region of a pn junction (see fig. 8.19 on page 281). Undoped GaAs can hence be regarded as already depleted of free charge carriers, in contrast to Si and Ge. Carrier motion and carrier loss are thus less determined by recombination rather than by interaction with and trapping by defects and impurities.

Some relevant impurities in GaAs are shown in fig. 8.14 on page 273. The most important ones are carbon, acting as a shallow acceptor with a very small energy distance to the valence band edge, and silicon as a shallow donor. Very important for GaAs as a detector material is an intrinsic[36] defect called EL2, an antisite, where an arsenic atom takes the place of a gallium site. It acts as a deep double donor (two extra electrons) sitting slightly below the centre of the band gap ($E_C - E_{\mathrm{EL2}} = 0.75\,\mathrm{eV}$). The reverse situation, a Ga atom on an As site, acts as shallow acceptor which is less relevant for the detector operation. The EL2 defect controls the GaAs sensor properties for charge collection because it creates regions of different field strengths inside the substrate, as is discussed further below. The concentration of EL2 defects is therefore always specified for GaAs material characterisations.

The defect and impurity density in GaAs is typically nine orders of magnitude larger than the intrinsic carrier concentration. They are often electrically active (i.e. ionisable) and hence can lead to mobile charge carriers. While in silicon the active detector layer is created by a reverse biased pn diode with very different doping levels of both parts, in GaAs detectors the space charge region is usually created by operating them in Schottky mode (see section 8.3.4). The energy bands are bowed up towards the metal–semiconductor boundary, as is shown in fig. 8.94. This way zones are created in which the EL2 levels lie above the Fermi level E_{F} and are thus ionised, as well as zones in which they lie below E_{F}. The latter part is hence not depleted. The resulting space charge density thus has zones with high and with low field strengths, respectively,

[36] Not caused by any impurity.

leading to regions with high and low charge collection efficiency (active/passive zones). In the low-field regions the signal carriers are not removed sufficiently fast, such that their recombination is enhanced and the signal decreases.

In recent times process technologies have been successfully applied [137], by which the bothersome EL2 defects are compensated after GaAs crystal growth by chromium doping diffused into the crystal at high temperatures [332]. Chromium atoms on gallium lattice sites act as deep acceptors in GaAs with a similar energy distance to the valence band as the EL2 level has to the conduction band (see fig. 8.14). If the density of the Cr atoms N_{Cr} is chosen larger than the density N_D of existing (shallow) donor levels from the n-type doping, the donor electrons fill the deep EL2 levels and are finally compensated themselves by the Cr acceptor centres (see [498], page 48). The result is high-ohmic p-doped GaAs of detector-grade quality.

In electronics GaAs is used for fast switching applications because of its high mobility and high saturation drift velocity. Being a direct semiconductor (see section 8.2.3) it also absorbs and emits photons more efficiently than silicon, making it better suited for some applications in photonics (LEDs, laser, solar cells). For particle detection GaAs was studied intensively in the 1990s, in particular because of its potentially better radiation resistance compared to silicon due to its larger band gap and the resulting better leakage current behaviour. The large impurity density, however, causes short carrier lifetimes and impedes complete charge collection due to trapping.

The detector quality of GaAs is characterised by the *charge collection efficiency* (CCE) defined in (8.123) or by the *charge collection distance* (CCD). CCEs of 90–100% for 250–300 µm thick detectors or CCDs of more than 250 µm were achieved with GaAs in single detectors [830].

GaAs appears to be radiation hard to high doses for gamma radiation and for electrons, which makes it interesting for electronics and sensors in military applications. In environments with high hadron fluences as at the LHC, however, NIEL damage is relevant (see section 8.12), which in GaAs appears to be more severe by an order of magnitude than in oxygen-enriched silicon [715].

A further problem for the use of GaAs as a particle detector lies in the fact that the drift velocity for electrons has a maximum at a certain value of the electric field strength (see fig. 4.16 on page 123) and that the field itself is not constant in GaAs but varies strongly over the depth of the detector (see fig. 8.94). Therefore electrons with various drift velocities can be underway when drifting towards an electrode. This can lead to complications in pulse shape interpretation and time assignments.

8.13.3 Cadmium telluride and cadmium zinc telluride

The II–VI compound semiconductor CdTe as well as $Cd_{1-x}Zn_xTe$ (known as CZT) are of interest as semiconductor detectors, in particular for X-ray imaging because of the high atomic numbers ($Z_{Cd} = 48, Z_{Te} = 52, Z_{Zn} = 30$) and the corresponding good photon absorption property, together with good energy resolution. The band gap increases from 1.44 eV for pure CdTe to 2.2 eV for pure ZnTe (without Cd); values in between depend on the Zn fraction. In use are CdTe detectors (without Zn) as well as CdZnTe detectors with a Zn fraction x of up to 15% corresponding to a band gap of about 1.6 eV. They can be operated as ohmic (mostly for CZT) or as Schottky (often for CdTe) detectors depending on whether both metal–semiconductor boundaries act as ohmic contacts or if one is a Schottky contact (see section 8.3.4). The numbers of created e/h pairs are comparable to those in silicon. The average energy needed for

the creation of an e/h pair is 4.43 eV, only about 20% larger than for silicon (compare table 8.2). The energy resolution for a 100 keV energy deposit neglecting electronic noise is about 200 eV from the statistical fluctuations of the charge carriers using a Fano factor of 0.089 (see section 17.10.2 on page 781).

Characteristics of CdTe and CdZnTe are their low hole mobility ($100\,\mathrm{cm}^2/\mathrm{Vs}$) and the small mobility–lifetime product for both carrier types:

$$(\mu\tau)_e \approx 10^{-3}\,\mathrm{cm}^2/\mathrm{V}\,,$$
$$(\mu\tau)_h \approx 10^{-4}\,\mathrm{cm}^2/\mathrm{V}\,.$$

For comparison, $\mu\tau$ for silicon is larger by 2–3 orders of magnitude.

The intrinsic carrier concentration n_{intr} is in the order of 10^7. Therefore additional depletion is not needed, similar to GaAs. However, in the production process this low intrinsic carrier concentration is usually not achieved because free charge carriers introduced by lattice defects or impurities outweigh the intrinsic concentration. Often chlorine or indium is added as a dopant with donor character to compensate crystal defects by forming chlorine–vacancy complexes [966]. While in GaAs ionised EL2-defects are responsible for high and low field regions in the detector, in CdTe the strength of the electric field depends on the occupation level of the trapping centres (see also section 8.12.1) and hence on the time that the detector has been exposed to radiation. Compared to silicon detectors, CdTe and CZT detectors often also show incomplete and inhomogeneous charge collection properties.

CdTe detectors can be operated as ohmic or Schottky devices by the choice of the electrode metal, depending on how the metal's work function $e\phi_m$ and the electron affinity $e\chi_s$ of CdTe are related (see section 8.3.4). Figure 8.95 shows ohmic and Schottky contacts for weakly p-doped and Cl-compensated CdTe, a typical choice. The electron affinity $e\chi_S$ in CdTe is 3.35 eV and a typical contact metal like Pt has $e\phi_m = 5.8\,\mathrm{eV}$ as work function. For the contact characteristic for p doping (holes are the majority carriers) the conditions near the valence band edge are relevant. For weak p doping and after contact with the metal, there is no potential barrier for holes existent at the boundary in either direction. The contact has 'ohmic' behaviour (fig. 8.95(a)).

If the platinum contact is exchanged on one side by an indium contact, which has a relatively low work function of 4.1 eV, a potential barrier forms at the boundary which prevents holes flowing from the metal into the semiconductor (physically speaking this means that electrons cannot flow from the semiconductor into the metal). A rectifying Schottky contact is formed as described for electrons and n-doped Si in section 8.3.4, while in fig. 8.95 p-doped CdTe is considered.[37] By applying a reverse bias voltage the leakage current can be kept small in a CdTe detector in Schottky mode, hence increasing the resistance by an order of magnitude compared to the ohmic contact ($\approx 10\,\mathrm{G\Omega\,cm}$). Such CdTe Schottky detectors, however, develop negative space charge regions, caused by charge injection at the Schottky contact, which change the electric field in the sensor in a time dependent way. This leads to a decrease of the signal with time. A cure of this effect can be for example a periodic depolarisation by removal of the bias for a few seconds on a time-scale of 10 minutes [160].

Therefore both contact types are found in applications. Pixel detectors with good spatial resolutions and with CdTe as a substrate material are particularly interesting

[37] For n-doped CdTe the contact relations are reverse to those for p-CdTe: ohmic for n-CdTe/In, Schottky for n-CdTe/Pt or n-CdTe/Au [929].

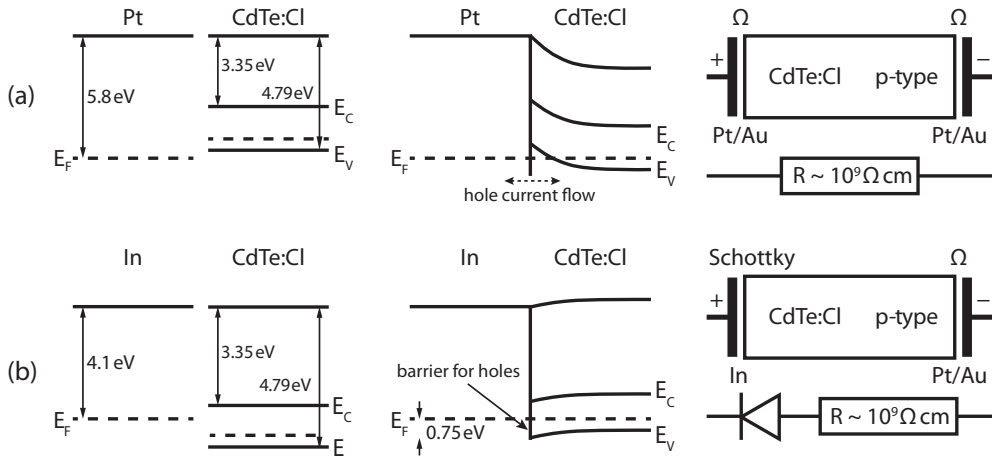


Fig. 8.95 Ohmic (a) and Schottky- (b) contacts in Cl-compensated, p-doped CdTe: shown are on the left the energy-band relations before the contact, in the middle after the contact, and on the right the arrangement of the contacts and the replacement circuit. Indicated are the band edges E_C and E_V, the Fermi level E_F, resp. E_f, for p doping as well as the energy distances to the 'continuum'. The type of contact depends on the metal, here platinum and indium, respectively. Also indicated are in (a) the ohmic current flow and in (b) the barrier for holes (after [407]).

for applications in X-ray astronomy [970] and also as imaging X-ray detectors [668, 977, 265, 757].

8.13.4 CVD diamond

Carbon in the form of the diamond crystal lattice is an insulator with a resistance of 10^{13}–$10^{16}\,\Omega\,\mathrm{m}$ which traces back to the large band gap of diamond. Natural diamond is formed in large pressure and temperature environments around 50 000 bar and 1300 K in the Earth's mantle and under stable tectonic plates. By volcanic activity diamondiferous rock penetrates up into depths accessible to mining.

A process technique to fabricate artificial diamond, successfully applied for the first time in 1962, is deposition of diamond out of the vapour phase (*chemical vapour deposition*, CVD) in layers deposited on a substrate (see for example [608, 575]). The growth starts by nucleation of carbon from a mixture of methane (CH_4), molecular hydrogen (H_2) and optionally an oxygenated compound like acetone $(CH_3)_2CO$ on a nucleation substrate consisting of silicon or a metal. By means of microwave radiation a plasma of hot electrons (5000 K) is created which activates the gas components. In this process atomic carbon is created by removal of the hydrogen atoms from the methane which then condenses on the substrate. Details of this gas phase chemistry is described in [755].

The process allows for the fabrication of polycrystalline diamond structures with thicknesses of up to about 1 mm. The growth process causes vertical structures, originating from the original seed grains as shown in fig. 8.96. The typical extension of these structures at the top surface (in growth direction) amounts to about 100 µm [644], towards the substrate side they are much smaller.

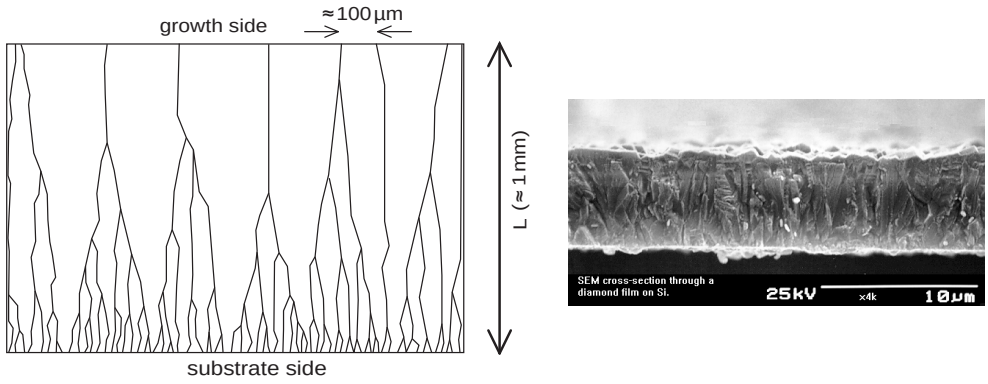


Fig. 8.96 Growth structure of CVD diamond created by chemical vapour deposition (left: schematic, right: SEM photography [702], credit: CVD diamond group, School of Chemistry, University of Bristol, UK). The growth process creates vertical crystal grain structures.

The energy necessary to kick an atom off its lattice position is 43 eV [623], which is substantially higher than that for silicon (25 eV [662]). Also due the strong lattice bonds diamond cannot be doped by impurity atoms. Both facts together render diamond principally a very radiation hard material. Consequently the damage constants β_t (see section 8.12.1, page 355) are smaller than for silicon by a factor 2–3 [950], depending on the particle energy.

The small atomic number ($Z = 6$) renders diamond only slightly attractive for photon detection but all the more for tracking detectors, as the radiation length X_0 is inversely proportional to Z^2 (see eq. (3.89) on page 61). In comparison to silicon ($n_i \approx 10^{10}\,\mathrm{cm}^{-3}$) and even to depleted silicon (typically $n \sim 10^{5-6}\,\mathrm{cm}^{-3}$) diamond is free of charge carriers[38] ($n_i \approx 0$, table 8.2). Furthermore, because the mass density is larger by a factor of two, the distribution of deposited energy (Landau distribution) is narrower for diamond than for silicon at the same thickness.

To create an e/h pair in diamond an average energy of $w_i = 13.1$ eV is needed, that is, almost four times more than in silicon. As a consequence the amount of signal charge is much smaller than in silicon. The energy deposited per particle path length, however, benefits from the 50% higher density of diamond compared to silicon. Per micrometre on average about 500 eV of energy is released in diamond. The number of created e/h pairs depends on the assumed energy cut-off value T_{cut} up to which the ionisation electron remains in the detector and is available for e/h pair creation. This restricted energy loss is given in eq. (3.33) on page 34. The Sternheimer parametrisation [913,765] can be used to compute the δ term of (3.33). In [1021] a value of $T_{cut} = 7.5$ keV has been found optimal for a diamond thickness of 500 µm leading to 32 e/h pairs per micron.

Both carrier mobilities have very high values (1800 cm^2/Vs for electrons, 1600 cm^2/Vs for holes), whereas the carrier lifetimes are low (≈1 ns), see table 8.2. The challenge for diamond detectors therefore is to fabricate sensors with longer carrier lifetimes leading with high electric fields to CCDs of about 200 µm or more.

[38] From (8.16) we find that diamond without impurities with effective state densities of $N_C \approx 10^{20}\,\mathrm{cm}^{-3}$ and $N_V \approx 10^{19}\,\mathrm{cm}^{-3}$ and a band gap of $E_G = 5.5$ eV has a charge carrier density at 300 K of $n_i \approx 10^{-27}\,\mathrm{cm}^{-3}$.

Fabrication and development of artificial diamond as particle detectors was at first successful in the form of polycrystalline (pCVD) diamonds [582, 46, 150] and later also in monocrystalline form (sCVD, *single crystal*) [150], for which successively first tiny, then larger mono-crystals are used as seeds for further growth of monocrystalline pieces. Detector-grade pCVD diamond can meanwhile be fabricated in wafer sizes [582].

The undoped substrate material does not build up space charge regions when put in contact with a metal. A typical electrode configuration uses a Ti/W layer coated with gold for having a solder wettable surface to place a contact (e.g. a *bump-bond*, see section 8.5.5).

Diamond detectors behave in principle as parallel-plate detectors (see fig. 5.5 on page 137) filled with a dielectric. In a pure monocrystal the electric field in the sensor is constant. In polycrystalline diamond, however, vertical polycrystal grain structures exist. At their boundaries charges can be trapped leading to a (largely horizontal) polarisation field inside the detector which is then superimposed on the applied drift field. This leads to systematic distortions in the space point reconstruction of traversing particles [644]. The lateral extent of the polycrystalline grain structures is irregular and fabrication-dependently varying. The average grain size at the electrode side has been measured to be about 100 µm [644]. Their irregular pattern superimposes on the regular pattern of the electrodes (pixels or strips), thereby impacting the spatial resolution. For pCVD pixel detectors with a pixel pitch of 50 µm, resolutions of about 20 µm have been measured [149] which must be compared with the binary resolution (eq. (8.83) on page 300) of $50\,\mu\text{m}/\sqrt{12} \approx 15\,\mu\text{m}$. Detectors made of single crystal diamonds (sCVD) are of course free of grain-boundary effects.

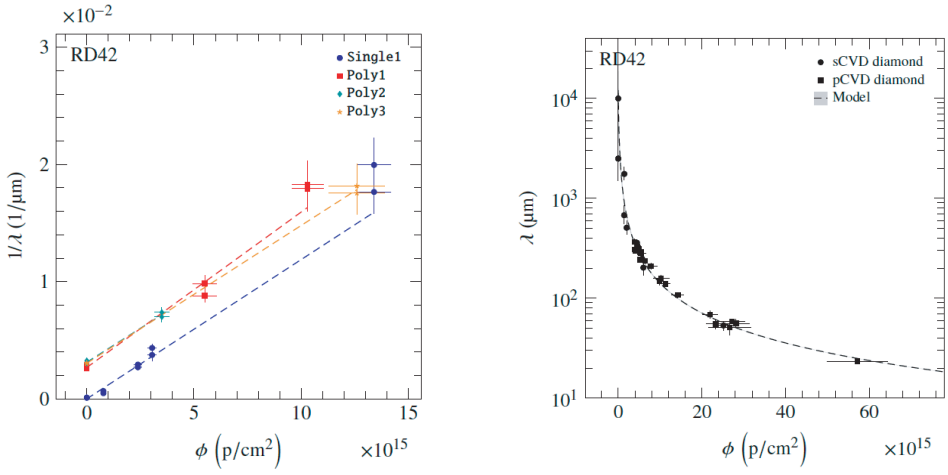
Of particular interest is the radiation hardness of diamond at fluences in the order of more than $10^{15}\,\text{n}_{eq}\,\text{cm}^{-2}$. Due to the large band gap diamond detectors are completely free of leakage current and can be operated at room temperature even under intense irradiation. Furthermore, as already mentioned above, the strong lattice bonds and the subsequently high energy transfer required to release carbon atoms from the lattice (43 eV versus 25 eV in Si) renders diamond more robust against lattice damage.

Signal losses are mainly traced back to charge trapping. As a positive side effect of the irradiation one observes, however, that at moderate irradiation levels of pCVD diamonds, trapping centres with long relaxation times located at the vertical grain boundaries become neutralised by trapped carriers. This causes a relative increase of the signal and can be provoked by deliberate pre-irradiation using radioactive sources, a treatment known as *priming* or *pumping*. This moves the threshold for signal loss by charge trapping to fluence values above the order of $10^{15}\text{n}_{eq}/\text{cm}^2$ [150], depending somewhat on radiation type and energy.

Charge trapping reduces the mean free path $\lambda = \lambda_e + \lambda_h$ of charge carriers as described before by (8.130):

$$\frac{1}{\lambda(\phi)} = \frac{1}{\lambda(\phi = 0)} + \beta_t \phi_{eq}\,, \tag{8.133}$$

with β_t being the damage constant due to trapping and $\lambda_{e/h}(\phi = 0)$ describing the mean free path before irradiation. In fig. 8.97(a) the inverse mean free path, determined using (8.128) and (8.123), is plotted for sCVD and pCVD diamond in comparison as a function of proton fluence [147] (shown here for 800 MeV protons) yielding a damage constant $\beta_t(800\,\text{MeV p}) = 1.21 \pm 0.06 \times 10^{-18}\,\text{cm}^{-2}/(\text{p}/\mu\text{m})$. The damage is particle



(a) Inverse mean free path

(b) Mean free path for sCVD and pCVD

Fig. 8.97 (a) Inverse mean free path λ plotted as a function of the fluence for 800 MeV protons comparing single-crystalline (sCVD) and polycrystalline (pCVD) diamond detectors. (b) Mean free path λ for pCVD and sCVD after correcting for a damage offset in pCVD diamond [147].

Table 8.5 Relative damage constants in diamond for protons at different energies and for neutrons. Data from [147], 25 MeV p data from [950].

Radiation	24 GeV p	800 MeV p	70 MeV p	25 MeV p	fast n
relative β_t^{rel}	1	1.85 ± 0.13	2.5 ± 0.4	4.7 ± 0.7	4.5 ± 0.5

type and energy dependent which can be quantified by a relative damage constant normalised to that of 24 GeV protons, as compiled in table 8.5. The damage level is more severe for lower energy protons and for neutrons. In comparison to silicon before irradiation and at low fluences, diamond falls short in mean free path λ by an order of magnitude, but features a much slower decrease for higher fluences. The damage constant β_t (slope) for diamond is 2–3 times smaller than for silicon.

While in fig. 8.97(a) the mean free path remains larger for single crystal diamond, the parallel slopes for sCVD and pCVD indicate the same damage mechanism to be in action for both: compared to sCVD diamond pCVD material features a shorter λ from the start. This is supported by fig. 8.97(b) where λ is shown versus fluence for 24 GeV protons again comparing sCVD and pCVD. Here the polycrystalline (pCVD) data have been corrected by an offset fluence $\phi_0 = 1/\beta_t\lambda(\phi = 0)$. Hence pCVD material can be regarded as already damaged sCVD material having a smaller mean drift distance $\lambda_(\phi = 0)$ (8.130) from the start.

9

Track reconstruction and momentum measurement

9.1 Charged particles in a magnetic field

The momentum vector $\vec{p}$ of a charged particle can be determined by measuring its trajectory in a magnetic field. In a magnetic field $\vec{B}$ a particle with mass m, charge q, velocity $\vec{v}$ and Lorentz factor γ will be deflected by the Lorentz force:

$$\vec{F} = \dot{\vec{p}} = q\left(\vec{v} \times \vec{B}\right) \qquad \Rightarrow \qquad \dot{\vec{v}} = \frac{q}{\gamma m}\left(\vec{v} \times \vec{B}\right) . \tag{9.1}$$

The right-hand side uses $\vec{p} = \gamma m \vec{v}$ derived from (2.6) with the time derivative $\dot{\vec{p}} = \gamma m \dot{\vec{v}}$. Note that $\dot{\gamma} = 0$ because the magnetic field does not change the particle energy.

The solution of the differential equation (9.1) describes a rotating velocity vector $\vec{v}_T$ in the plane perpendicular to $\vec{B}$, while the component parallel to the magnetic field remains unchanged. If we restrict ourselves to homogeneous magnetic fields, we find in a coordinate system where the magnetic field has only one component, for example with $B_1 = B_2 = 0,\ B_3 = B > 0$:

$$\begin{aligned} v_1 &= v_T \cos(\eta\, \omega_B\, t + \psi_0)\,, \\ v_2 &= -v_T \sin(\eta\, \omega_B\, t + \psi_0)\,, \\ v_3 &= v_3\,. \end{aligned} \tag{9.2}$$

Here ω_B is the cyclotron frequency (or gyro-frequency), η the sign of the charge and v_T the absolute value of the velocity component perpendicular to the magnetic field:

$$\omega_B = \frac{|q|\, B}{\gamma m}, \qquad \eta = \frac{q}{|q|}, \qquad v_T = \sqrt{v_1^2 + v_2^2}\,. \tag{9.3}$$

Although (9.1) is not written in a covariant form it is also valid for relativistic velocities (a detailed discussion can be found, for example, in [564]). Relativity enters only via the Lorentz factor γ, hence via the cyclotron frequency $\omega_B = q\, B/(\gamma m)$ which is constant in the non-relativistic case ($\gamma \approx 1$) and dependent on the energy γm in the relativistic case.

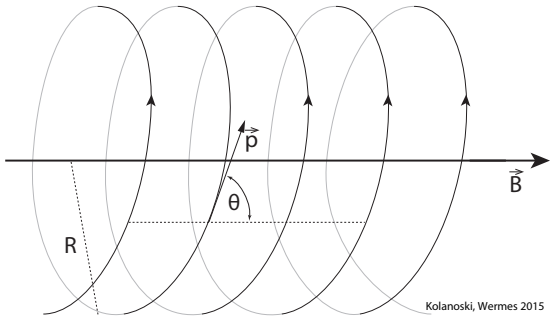


Fig. 9.1 Helix trajectory of a positively charged particle in a magnetic field. For positively charged particles the rotational direction is anticlockwise when looking into the direction of the $\vec{B}$ field. The angle θ between the momentum and a line parallel to the magnetic field is also shown ($\cos\theta = \vec{p}\,\vec{B}/(|\vec{p}||\vec{B}|)$).

Integration of (9.2) yields the particle trajectory in position space:

$$\begin{aligned} x_1 &= \frac{v_T}{\eta\,\omega_B}\,\sin(\eta\,\omega_B\,t + \psi_0) + x_{10}\,, \\ x_2 &= \frac{v_T}{\eta\,\omega_B}\,\cos(\eta\,\omega_B\,t + \psi_0) + x_{20}\,, \\ x_3 &= v_3\,t + x_{30}\,. \end{aligned} \tag{9.4}$$

This is the representation of a helix (fig. 9.1) which lies on a cylinder surface coaxially with the magnetic field and whose projection onto the plane perpendicular to $\vec{B}$ describes a circle with radius

$$R = \sqrt{(x_1 - x_{10})^2 + (x_2 - x_{20})^2} = \frac{v_T}{\omega_B} = \frac{\gamma\,m\,v_T}{|q|\,B} = \frac{p_T}{|q|\,B}\,. \tag{9.5}$$

For a given momentum the radius depends only on the charge but not on other properties of the particle, like for example the mass. The measurement of R determines the so-called rigidity, that is, the ratio $p_T/|q|$ from which follows the transverse momentum if the charge is known. In particle physics one can usually assume $q = \pm e$; in other fields, like nuclear, heavy-ion or astroparticle physics, the charge must be determined independently. An independent measure of the charge is, for example, the energy loss of a particle in a medium because, according to the Bethe–Bloch formula (3.25), for a given momentum the energy loss is proportional to the charge (this is for example demonstrated for the detection of different nuclei in cosmic radiation in fig. 14.8 and in section 16.2).

Since the component p_3 of the momentum parallel to the magnetic field and the absolute value of the transverse momentum p_T do not change with time, the angle θ defined by

$$\cos\theta = \frac{p_3}{p} \qquad \text{or} \qquad \tan\theta = \frac{p_T}{p_3} \qquad \text{or} \qquad \sin\theta = \frac{p_T}{p} \tag{9.6}$$

is a constant of motion. In these equations θ is the angle between the instantaneous momentum direction and the direction of the magnetic field (fig. 9.1). It can assume values between 0 and π which are uniquely mapped by the tangent and cosine functions but are ambiguously mapped by the sine for θ and $\pi - \theta$. On the unrolled cylinder surface θ is the angle between the particle trajectory and the magnetic field. The complementary angle, $\lambda = \pi/2 - \theta$, is called *dip angle* and can be expressed by

$$\cos\lambda = \frac{p_T}{p}\,. \tag{9.7}$$

The angle θ is preferred in experiments at collider rings with the magnetic field in the beam direction since in this case it is the polar angle of the particle relative to the beam. In experiments with the magnetic field perpendicular to the beam this polar angle is the dip angle λ.

Employing the reconstructed radius of curvature R and slope θ of the helix, the transverse and the total momentum can be determined according to (9.5) and (9.6):

$$p_T = |q|\, B\, R\,, \qquad p = \frac{p_T}{\sin\theta}\,. \tag{9.8}$$

For practical use the transverse momentum can also be written

$$p_T = 0.3\,|z|\, B\, R \tag{9.9}$$

where the quantities are given in the following units: p_T in GeV/c, B in T and R in m; z is the particle charge in units of the elementary charge, $z = q/e$. Equation (9.9) means that in a field of 1 T the trajectory of a 0.3-GeV/c particle has a curvature radius of 1 m. For the conversion of (9.8) to (9.9) we use $\mathrm{T} = \mathrm{V\,s/m^2}$ and the velocity of light (slightly rounded) $c \approx 0.3 \times 10^9\,\mathrm{m/s}$ to get:

$$\mathrm{GeV}/c = \frac{\mathrm{GeV}}{0.3 \times 10^9\,\mathrm{m/s}} = \frac{1}{0.3}\,\frac{\mathrm{eV\,s}}{\mathrm{m}} = \frac{e}{0.3}\,\mathrm{T\,m}\,. \tag{9.10}$$

The factor 0.3 occurs whenever we change from the momentum unit eV/c to eV s/m. The latter would be a normal SI unit if e is expressed in units of coulomb, but in most cases, as in (9.9) above, e cancels out when writing charges in units of e.

9.2 Magnetic fields

Modern particle experiments often have magnets with volumes up to many cubic-metres and field strengths of several tesla. Large field strengths with long path lengths of the particles in the field are desirable for good momentum resolution up to high energies. With large field volumes multi-particle reaction products can be detected more completely and precisely than in small ones. Usually for momentum measurements the magnetic field resides in an air volume where it should be as homogeneous as possible. As the field lines must be closed the field is returned through an iron yoke with high permeability. Sometimes magnetic fields are employed which are fully contained in iron (magnetised iron), in particular for muon detection. The highest magnetic fields are reached with superconducting magnets [1011] which could even be operated without a return yoke.

The magnet configurations differ according to the requirements of the experiment. For example, dipole magnets are usually employed in fixed-target experiments (section 2.2.2) where particles are measured in the forward direction. In storage ring experiments, in contrast, usually solenoid magnets with the field in beam direction are preferred. The reason is that in this way the transverse momentum measurements are rotationally symmetric with respect to the beam, that is, without dependence on the azimuth, and with good solid angle coverage.

9.2.1 Dipole magnet

Figure 9.2 shows a typical dipole magnet. The volume usable for the momentum measurement is the air gap between the pole pieces. The magnetic field in this gap

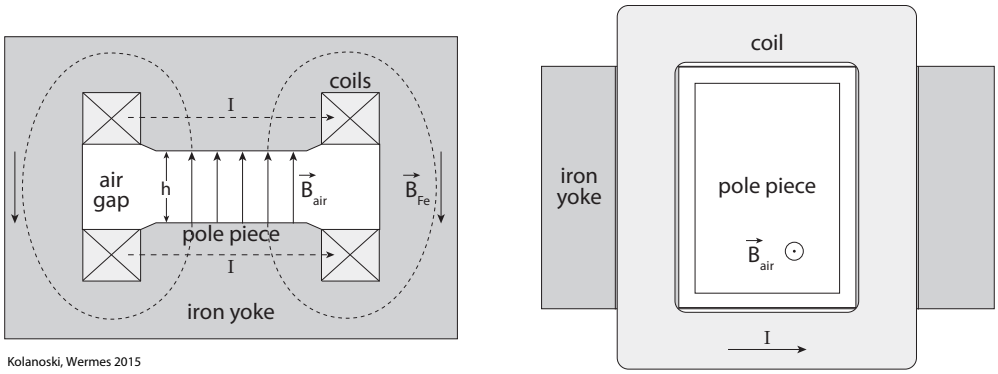


Fig. 9.2 Cut view of a typical dipole magnet. Left: Cross section perpendicular to the pole pieces, view into the preferred direction of the particles. In an air gap between the two pole pieces an approximately homogeneous magnetic field is generated by currents which run in coils around the pole pieces. The magnetic field lines are closed on both sides by an iron yoke. Right: Cross section parallel to the pole pieces through one of the coils.

is generated by coils which are wound around the pole pieces and through which a current I runs in a total of N turns. The current generates a magnetic field according to Ampere's law:

$$N\,I = \frac{1}{\mu_0\,\mu_{\text{air}}}\int_{\text{air}} \vec{B}\,d\vec{l} + \frac{1}{\mu_0\,\mu_{\text{Fe}}}\int_{\text{Fe}} \vec{B}\,d\vec{l}. \tag{9.11}$$

Here $\mu_0 = 4\pi \times 10^{-7}$ T m/A is the vacuum permeability and μ_{air} and μ_{Fe} are the relative permeabilities in air and iron, respectively. The integration paths through air and iron together form a closed path. For $\mu_{\text{air}} \approx 1 \ll \mu_{\text{Fe}} \approx 1000$–$10\,000$ the magnetic field in the air gap of height h is approximately

$$B \approx \frac{\mu_0\,N\,I}{h}\,. \tag{9.12}$$

The air gap should be large enough to yield a good angular acceptance and usually also to allow for the installation of detectors in the field volume. However, with the enlargement of the field volume the inhomogeneous fringe fields also increase. Inhomogeneities can be reduced by 'shimming', that is, by appropriately forming the iron structures, like pole pieces and iron yokes, or by employing additional coils.

9.2.2 Solenoid magnet

The magnetic field in a solenoid magnet (fig. 9.3) is generated by a current-carrying coil. As in the case of dipole magnets, here the field lines are also usually closed via an iron yoke. With a similar ansatz as for (9.11) for dipole magnets the magnetic field inside the solenoid is obtained by

$$B \approx \frac{\mu_0\,N\,I}{L}\,. \tag{9.13}$$

The current I flows through N turns over a length L around the magnetic field. The longer the coil relative to its diameter the more homogeneous is the magnetic field. Inhomogeneities at the coil ends, which would hamper particle tracking in these regions, can be much reduced by properly adjusting the iron yoke.

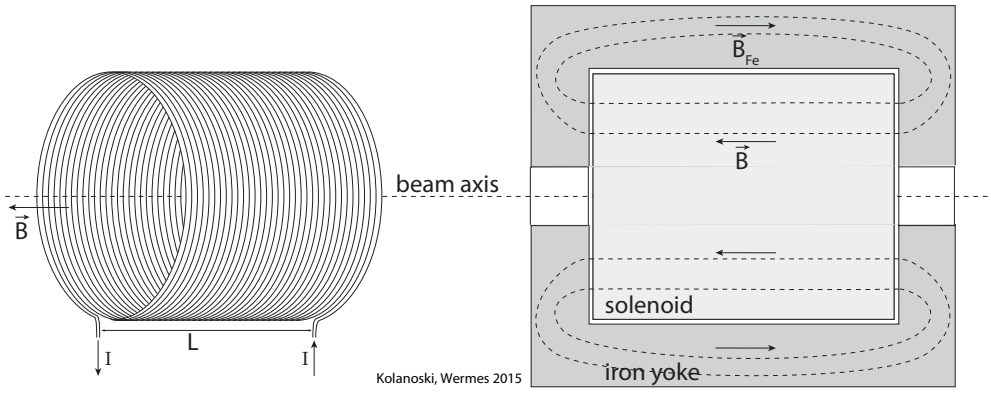


Fig. 9.3 Solenoid magnet. Left: A current I flowing through a coil with N windings (over the length L) generates a magnetic field $\vec{B}$ inside the coil. Right: Cross section of a solenoid with an iron yoke. Apart from an opening for feeding the beam through the solenoid can be completely surrounded by an iron yoke. By means of the iron yoke with high permeability the field strength is increased and becomes more homogeneous in the border regions. In storage ring experiments most of the particle detection system is inside the coil, with some extension into the beam directions. The iron is usually also used as hadron absorber for the identification of muons. In this case tracking chambers are additionally installed outside the iron. Often the iron yoke is built up from several layers interleaved with tracking chambers which provide muon detection and measurements of the track curvature in the magnetised iron.

Solenoids are often chosen when particles are produced inside the magnetic field. This is mainly the case in collider ring experiments where usually the interaction point (IP) lies in the centre of the solenoid; the solenoid axis and the magnetic field point in the beam direction. Inside the solenoid tracking detectors are installed, often consisting of cylindrical drift chambers or cylindrically arranged semiconductor detectors and vertex detectors (see chapters 7 and 8 and section 14.6.2).

Characteristic data of some solenoid magnets are given in table 9.1. The dimensions of a solenoid magnet are typically some metres in diameter and length. While large solenoid magnets with normal conducting coils reach field strengths below 1 T, limited by the saturation of magnetisation of the iron and the necessary power for operation, superconducting coils reach typically 2 T. Regarding volume and field strength, the largest solenoid magnet up to now is employed in the CMS experiment with a field of 4 T in a volume of about 360 m^3 (table 9.1) [298].

For the safety of superconducting magnets the stored energy and the energy per mass (E/M) are important quantities, which are also listed in table 9.1. The energy per mass is proportional to the temperature increase in the case of an abrupt transition from superconductivity to normal conductivity, a so-called 'quench'. In such a case the total magnetic energy will be abruptly converted into heat.

9.2.3 Toroid magnet

If the coil of a solenoid magnet is bent in a way that both coil ends meet each other, the field lines will be closed and a toroidal magnetic field arises (fig. 9.4(a)). If the current coils are bent circularly, for symmetry reasons the field lines will also be circles. The integral of the magnetic field along a closed path, which we choose here along a

Table 9.1 Data of different solenoid magnets (experiments with a * are no longer operational). For about 30 years solenoid magnets for experiments at storage rings were almost exclusively built with superconducting coils; the respective examples are taken from [746]. If the electromagnetic calorimeter (section 15.5) is placed outside the coil, the coil thickness x in units of the radiation length X_0 is an important specification (left blank where the electromagnetic calorimeter is inside the coil). For superconducting magnets the stored energy and the energy per mass (E/M) are also listed. For comparison, data of several normal conducting solenoids are also shown [245,32,73,63]. As the critical quantity, instead of the stored energy, the operation power P is given.

Experiment	Laboratory	B (T)	Radius (m)	Length (m)	x/X_0	Energy (MJ)	E/M (kJ/kg)	P (MW)
			superconducting solenoid magnets					
CDF*	Tsukuba/Fermi	1.5	1.5	5.07	0.84	30	5.4	
CLEO-II*	Cornell	1.5	1.55	3.8	2.5	25	3.7	
ALEPH*	Saclay/CERN	1.5	2.75	7	2	130	5.5	
DELPHI*	RAL/CERN	1.2	2.8	7.4	1.7	109	4.2	
ZEUS*	INFN/DESY	1.8	1.5	2.85	0.9	11	5.5	
H1*	RAL/DESY	1.2	2.8	5.75	1.8	120	4.8	
BELLE*	KEK	1.5	1.8	4		42	5.3	
ATLAS	CERN	2	1.25	5.3	0.66	38	7	
CMS	CERN	4	3	12.5		2600	12	
			normal conducting solenoid magnets					
TASSO*	DESY	0.5	1.35	4.4	1.2			2.8
MARK II*	SLAC	0.5	1.56	4.05	1.3			1.8
ARGUS*	DESY	0.8	1.4	2.8				2
OPAL*	CERN	0.435	2.18	6.3	1.7			5

field line with radius r, equals the current which flows through the area encircled by the path:

$$N\,I = \frac{1}{\mu_0\,\mu}\oint \vec{B}\,d\vec{l} = \frac{1}{\mu_0\mu}\,B(r)\,2\pi r\,. \tag{9.14}$$

It follows that the field inside the torus has a $1/r$ dependence:

$$B(r) = \frac{\mu_0\,\mu\,N\,I}{r}\,. \tag{9.15}$$

Here μ is the permeability of the material in the magnetic field, usually $\mu \approx 1$ for air and up to about 10 000 for iron.

Toroid magnets are employed, for example, supplementing solenoid magnets to cover the forward and backward directions. In such a case particles coming from the interaction point (assumed to be in the centre of the solenoid) pass the forward/backward magnetic field about perpendicularly, independent of the azimuth angle around the beam axis. A dipole magnet in the forward/backward direction would also have the field perpendicular to the beam axis, but would lack the azimuthal symmetry.

The deployment of toroid magnets in the ATLAS experiment is rather unusual. Figure 9.4(b) shows the geometry of the magnet coils of the ATLAS detector which generate a solenoidal field inside and toroidal fields outside. The central toroid magnet

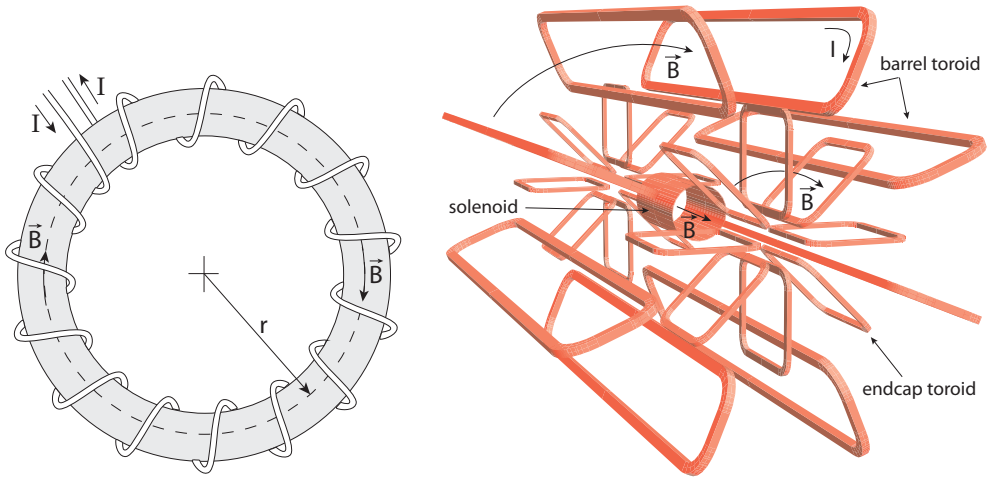


Fig. 9.4 Toroid magnet. (a) Principle of a toroidal magnetic field. The magnetic field lines run inside the coil whose ends are brought together such that the magnetic field is completely enclosed within the coil. (b) Magnet coils of the ATLAS detector: a solenoid coil in the innermost volume is surrounded by toroid coils [4]. The toroid coils cover the central region (*barrel toroid*) and the regions in the directions of the beams (*endcap toroids*), each with eight separate coils where each coil corresponds to one winding in the schematic drawing (a). The currents flowing in the coils build up the magnetic field inside the torus. While the solenoid coil generates a magnetic field pointing in the beam direction, the toroid coils generate magnetic fields which run concentrically around the beam axis. In the toroid magnetic fields a precise measurement of the muon momenta can be achieved with only small disturbances by multiple scattering. Source: ATLAS Experiment, CERN.

(barrel toroid) has eight separate coils, about 25 m long parallel to the beam axis and radially about 5 m wide. A magnetic field of on average 0.5 T (0.15–2.5 T) passes through the coils along a circular path around the beam axis. The magnetic field volume is equipped with drift chambers and RPCs (see section 7.7.3) for triggering on and tracking of muons. The toroid does not contain a hadron-absorbing iron core as is customary in devices used for muon identification. Therefore muons undergo less multiple scattering, resulting in more precise momentum measurements. In this case the function of a hadron absorber, otherwise mostly assumed by an iron yoke, is taken over by the calorimeters placed inside the toroid magnet.

9.2.4 Magnetised iron

Muons are identified by their ability to penetrate dense materials which absorb most other particles, in particular electrons and hadrons (see section 14.3). Hadron absorbers often consist of iron. If the iron is magnetised and tracking detectors are placed in gaps between iron layers, the muon momenta can be measured at the same time. The field strengths lie typically in the region of 1 T, limited by the magnetic saturation polarisation which could range up to about 2 T in soft iron. Examples for the application of magnetised iron are the iron yokes of the solenoid magnets at collider rings, as shown in the right panel of fig. 9.3 and for the CMS experiment in fig. 2.11 (the red layers

are magnetised iron). A fixed-target experiment employing magnetised iron for muon momentum measurements is the neutrino experiment described in [532].

9.2.5 Mapping of magnetic fields

Only in few cases can magnetic fields be assumed to be homogeneous. Since in the boundary area of a field inhomogeneities cannot be avoided, a homogeneous field approximation is best justified where particles are produced in the centre of a magnetic field, as for example in the magnetic field around the interaction point of a collider. But even in the solenoid magnets of such experiments inhomogeneities on the per cent level occur, while the requirements for momentum and direction reconstruction lie rather in the region of one-tenth of a per cent. As a rule, therefore, magnetic fields have to be surveyed, mapped and parametrised for their use in reconstruction algorithms. The absolute field strength is usually measured by a nuclear magnetic resonance (NMR) probe. The resonance frequency of the mostly used ^{1}H isotope ($\nu = 42.5759\,\mathrm{MHz/T}$ [511]) can be determined with a precision of better than 10^{-6}. In order to fully exploit the potential precision several conditions have to be met, for example, the homogeneity of the field in the probe volume ($O(\mathrm{cm}^3)$) has to be of the same order of magnitude as the desired precision of the field measurement.

For the measurement of all field components, also in the inhomogeneous regions, Hall probes are used which can be oriented in three orthogonal directions [846]. Usually the Hall data are used for the determination of the parameters of a field model which by construction accounts for the material and current distributions and which fulfils the conditions imposed by Maxwell's equations. The field models are computed using appropriate programs, for example the program MAFIA [975]. One can find for many magnetic spectrometers descriptions of the practical procedure of field measurements, for example in [733] for the CDF detector.

The parameters of a field model can also be determined using the so-called *floating-wire* method (see e.g. [95]). This method exploits the fact that a magnetic field exerts a force on a current. In order to use this effect, a wire through which a current flows is strung under a certain tension through a magnetic field. The deflection of the wire can be determined with a measuring microscope and compared with the field model predictions.

9.3 Particle trajectories in magnetic fields

9.3.1 Parametrisation of particle trajectories

In a field-free space a particle trajectory is described by a straight line and in a homogeneous field by a helix. A straight line in space is given by four parameters, for example offset and slope for two orthogonal projections, taken here in the xy and xz planes:

$$\begin{aligned} y &= t_{yx}x + y_0\,, \\ z &= t_{zx}x + z_0\,. \end{aligned} \tag{9.16}$$

Therefore at least two measurements for each projection are necessary to determine the parameters in (9.16).

For the parametrisation of the helix in a homogeneous magnetic field we resort to the representation (9.4) using the reference frame as defined in section 9.1 with the

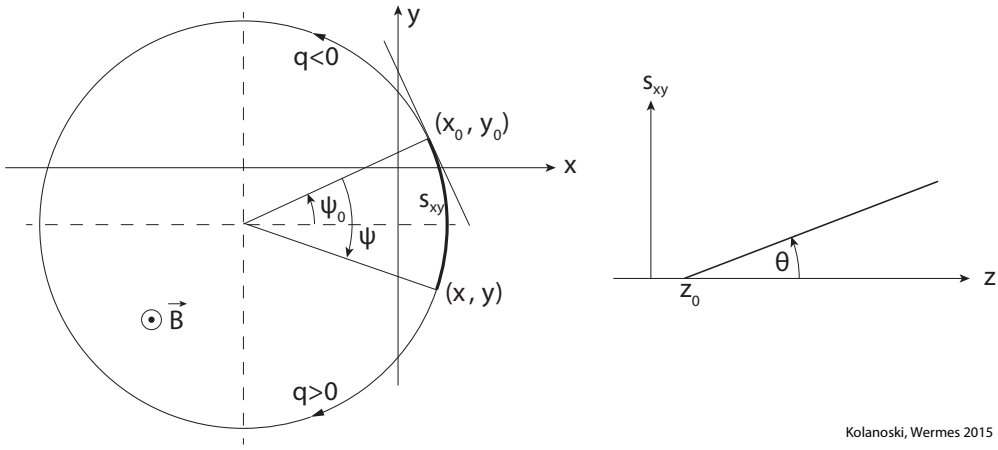


Fig. 9.5 Display of the parameters of a particle trajectory in a homogeneous magnetic field according to (9.17). See also the explanations in the text.

magnetic field in z direction. Since the time evolution of the helix is not measured we replace the phase $\omega_B t$ by a positive angle ψ which has the value $\psi = 0$ at the starting point (x_0, y_0, z_0) increasing in the direction of rotation of a positive particle (fig. 9.5). The starting point in the xy plane is turned by an angle ψ_0 against the x-axis in the mathematically positive sense. Thus the equation of a helix has the form (fig. 9.5):

$$\begin{aligned} x &= x_0 + R\,(\cos(\psi_0 - \eta\psi) - \cos\psi_0)\ , \\ y &= y_0 + R\,(\sin(\psi_0 - \eta\psi) - \sin\psi_0)\ , \\ z &= z_0 + \frac{\psi\, R}{\tan\theta}\,. \end{aligned} \tag{9.17}$$

The projection of the track length onto the xy plane is $s_{xy} = \psi\, R$. The direction of rotation is given by $\psi \geq 0$ and the sign of the charge, $\eta = q/|q|$. Positively charged particles rotate anticlockwise if viewed in the direction of the B field and negative particles rotate clockwise (see also fig. 9.1). In the representation (9.17) the point (x_0, y_0, z_0) is the starting point on the helix and the circle centre of the xy projection lies at $(x = x_0 - R\cos\psi_0,\ y = y_0 - R\sin\psi_0)$. The helix has six parameters: the start coordinates x_0, y_0, z_0, the two angles ψ_0, θ and the curvature in the xy projection,

$$\kappa = -\eta/R \qquad (\eta = q/|q| = \pm 1)\,, \tag{9.18}$$

with the sign depending on the charge as given above.

The projection of the helix onto the plane perpendicular to the magnetic field describes a circle determined by the three parameters x_0, y_0 and R. The representation

$$y = y_0 + \sqrt{R^2 - (x - x_0)^2} \tag{9.19}$$

depends nonlinearly on the parameters x_0 and R. The expansion of (9.19) for large radii R, corresponding to large momenta,

$$y = y_0 + R\left(1 - \frac{(x - x_0)^2}{2R^2} + \ldots\right) = \left(y_0 + R - \frac{x_0}{2R^2}\right) + \frac{x_0}{R}\,x - \frac{1}{2R}\,x^2 + \ldots \tag{9.20}$$

leads to a parabolic approximation of the trajectory which can be written as a linear function of three parameters:

$$y = a + bx + \frac{1}{2}cx^2 \,. \tag{9.21}$$

The new parameters a, b, c are functions of the parameters x_0, y_0, R corresponding to the first three coefficients in the expansion (9.20). Due to the linearity of the problem a matrix formalism, as described in appendix F, can be employed in order to find solutions for the parameters a, b, c when a set of measurements y_i at the positions x_i is given.

Particle trajectories in inhomogeneous magnetic fields are often described by piecewise concatenation of helices assuming constant averaged field strength and direction in each section.

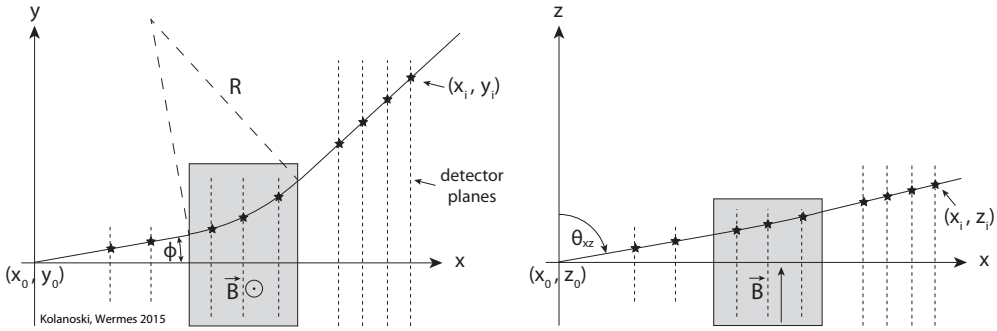
9.3.2 Detector arrangements in magnet spectrometers

An arrangement of one or more magnets with tracking detectors which is used for momentum determination is called a magnet spectrometer or just spectrometer. Figure 9.6 shows magnet spectrometers with typical detector arrangements for the measurement of position coordinates.

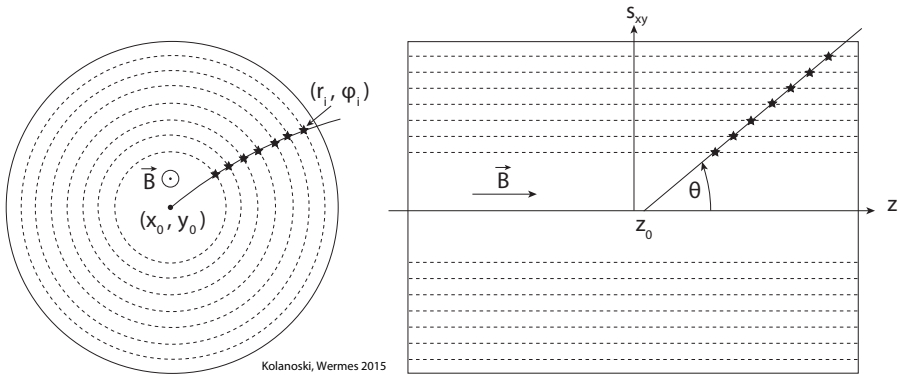
Forward spectrometers. The arrangement in fig. 9.6(a) with a dipole magnet is typical for a fixed-target experiment. It is called a 'forward spectrometer' because the particles which are produced by a beam hitting a fixed target move preferentially into the forward direction within a narrow solid angle. Therefore the detectors are installed preferentially perpendicular to the beam direction (here the x-axis) and deliver measurement points y_i and z_i at fixed x_i. The interaction point can be exploited as a measurement point as well if it is sufficiently well localised. In a homogeneous magnetic field the particle trajectories are helices, described by (9.17); in front and behind the magnet they are straight lines, described by (9.16).

The minimal number of measurements required for a full reconstruction are three points in the plane of momentum deflection and two points in the perpendicular plane. For example, in the xy plane of the arrangement in fig. 9.6(a) one point can be in front of the magnet and two behind. The line through the two points behind the magnet determines where the trajectory exits the magnet and is a tangent to the circle within the magnet in this point. Correspondingly, the line which passes the measurement point before the magnet and the, yet to be determined, entrance point into the magnet is a tangent to the circle at the entrance point. These constraints allow for an analytical determination of the entrance point. In the xz projection the trajectory slope is invariant (in a homogeneous magnetic field) and thus the angle at the start point can be determined. With the measured coordinates z_1, z_2 the slope of the line is $t_{xz} = (z_2 - z_1)/s_{xy}$, where s_{xy} is the trajectory length projected onto the xy plane between the detectors at x_1 and x_2.

Solenoid spectrometers. The arrangement in fig. 9.6(b) with a solenoid magnet for momentum analysis is typical for a collider experiment where the beams usually cross in the centre of the solenoid. The beam size in the deflection plane xy is typically about 100 µm or less so that the centre of the beam cross section in the xy plane can often be considered as a measurement point. In general this *beam constraint* will not



(a) Trajectory in a dipole spectrometer. The detectors are positioned in front, inside and behind the magnet. For the sake of simplicity in this example all detectors are in parallel to the front and back magnet surfaces. In the xz plane, which contains the magnetic field, the trajectory is nearly straight. The polar angle relative to the z-axis is shown as its xz projection θ_{xz}.



(b) Trajectory in a solenoidal magnetic field. In a solenoid field the measured quantities are usually the azimuthal coordinates on fixed radii ($(r\phi)$ measurements in cylindrical coordinates). The dashed lines present detector layers, for example the wire layers of a drift chamber. In order to obtain a precise momentum measurement usually the $(r\phi)$ coordinates have a better resolution than the z coordinates (see section 7.10.4).

Fig. 9.6 Particle trajectories in (a) dipole and (b) solenoid magnet spectrometers. In each case the magnetic field points into the direction of the z-axis; in both (a) and (b), the plane perpendicular to the magnetic field is shown on the left and a plane which contains the magnetic field is shown on the right. The dashed lines depict the detector layers. The measurement points along a track are marked by asterisks (*).

.

be required, however, in order to maintain the possibility to reconstruct tracks from decays which do not come from the primary vertex (see e.g. fig. 3.33).

In many detector arrangements the tracks are measured inside the solenoid by cylindrical track detectors, the measurement coordinates being (r_i, z_i, ϕ_i). In cylindrical drift chambers (section 7.10) or correspondingly arranged semiconductor detectors (section 8.4) the detector cells lie on a cylinder surface with fixed radii r_i on which the azimuthal angle ϕ_i and possibly also z_i are measured. For the different electrode configuration in TPCs we refer to section 7.10.10.

9.3.3 Fitting to a track model

The parametrised description of a particle trajectory is called a 'track model'. Such track models could be a combination of simple straight lines and helices, as given in (9.16) and (9.17), or in the case of inhomogeneous fields more generally a piecewise concatenation of helices or approximations, as for example the parabolic expansion in (9.21). Often direction and momentum changes by multiple scattering and bremsstrahlung along the trajectory are also taken into account. If the resolution is sufficient, kinks in the track caused by a strong scatter or by an energetic radiation might be observable, which can be accounted for by the track fitting. Such kinks are also observed when a charged pion decays, $\pi \to \mu\nu$ (particularly pronounced at low energies, see fig. 16.2). The decay of a long-living particle, like $K_s^0 \to \pi^+\pi^-$, where the pions do not come from the primary interaction point, should also be reconstructable (see e.g. fig. 14.29). For the reconstruction of decays of charm and bottom hadrons (fig. 9.14) as well as τ leptons, specialised vertex detectors (section 14.6.2) are necessary.

The track model depends on a set of m parameters $\theta = (\theta_1, \ldots, \theta_m)$ which are fitted to the N measurement points by the track reconstruction program. Usually the fitting proceeds according to the *least square method* by minimising the expression:

$$S = \sum_{i=1}^{N} \frac{(\xi_i^{\text{meas}} - \xi_i^{fit}(\theta))^2}{\sigma_i^2} \,. \tag{9.22}$$

The minimum of S with respect to the parameters θ follows a χ^2 distribution if the measurements ξ_i^{meas} are normally distributed around their expectation value $\hat{\xi}_i^{\text{meas}}$ with standard deviations σ_i. The expectation values $\hat{\xi}_i^{\text{meas}}$ are estimated by the $\xi_i^{fit}(\theta)$ which result from the optimal fitting of the parameters θ. Details of the method with references to the literature can be found in appendix F.

9.3.4 Particle trajectories in forward spectrometers

In the simplest case, that is, if the field can be assumed to be homogeneous within a box, the tracking model for a forward spectrometer, as shown in fig. 9.6(a), consists of the two straight lines in front and behind the magnet and a helix curve within the magnet. The track pieces have to fulfil the constraint that they connect continuously at the field borders. The constraints can be analytically included in the trajectory formulation or they can be accounted for by adding a 'Lagrange multiplier' term[1] to the χ^2 expression in (9.22).

In the general case of an inhomogeneous magnetic field, a precise magnetic field map (section 9.2.5) is required such that a track can be followed through the field (usually in stepwise approximations) and the set of parameters of the track model can be determined. In most cases the chosen parameters are the slopes and offsets as well as the inverse rigidity q/p at the interaction point, $(y_0, z_0, t_{yx}(y_0), t_{zx}(z_0), q/p)$. In each cell i the local parameters $(y_i, z_i, t_{yx}(y_i), t_{zx}(z_i), q/p)$ are used in order to follow the track into the next cell $(i+1)$ and to evaluate there the local parameters again. The set of global parameters is iteratively fitted to the measurement points (see e.g. [692]).

[1] See the corresponding literature on classical mechanics, for example [467], and on statistical methods of data analysis, for example [246, 224].

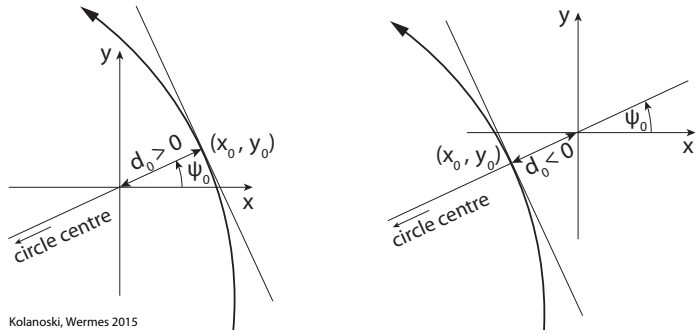


Fig. 9.7 Distance of closest approach of the track from the interaction point, here placed on the origin of the coordinate system. The given sign convention for the cases that the origin lies inside (left) or outside (right) of the circle, respectively, corresponds to the calculation of the point x_0, y_0 according to (9.23).

9.3.5 Particle trajectories in cylinder-symmetric spectrometers

We now consider the reconstruction of tracks in a solenoid spectrometer, as depicted in fig. 9.6(b), with the magnetic field parallel to the solenoid axis and with the origin of the coordinate system in the centre of the magnet. The centre of the interaction region should coincide with the centre of the magnet, i.e. the origin. Usually the detectors are arranged cylindrically providing measurement points $(r_i,\, \phi_i)$ in the plane perpendicular to the magnetic field ((r, ϕ) plane) and also space points $(r_i,\, \phi_i,\, z_i)$; for the measurement of the third coordinate see section 7.10.4. The particle's production point will usually not be used in the reconstruction as a constraint. Inside a solenoid magnet the magnetic field is relatively homogeneous so that a helix as a track model is a good approximation. If inhomogeneities, energy losses and scattering have to be taken into account the model has to be refined, usually leading to a piecewise, iterative reconstruction.

Usually the production point of a particle is not known, only that it lies on the helix trajectory in the region of the interaction zone. Since therefore the starting point of the helix is not reconstructible, only five of the six parameters in (9.17) can be determined. A possible choice of parameters is as follows:

κ curvature in the $r\phi$ plane corresponding to (9.18): $\kappa = -\frac{\eta}{R}$ $(\eta = q/|q|)$;

ψ_0 angle in the $r\phi$ plane between the x-axis and the vector from the circle centre to the position of the closest distance to the origin;

d_0 closest distance of the helix to the origin in the $r\phi$ plane; $d_0 > 0$ ($d_0 < 0$) if the origin lies inside (outside) of the circle (fig. 9.7);

θ slope angle of the track against the z-axis at the position of the closest distance to the origin; this angle is invariant as you follow the track around the cylinder (in a homogeneous field and without disturbance of the track) or, in other words, the track is a straight line on the unrolled cylinder surface where the helix is localised;

z_0 coordinate on the z-axis in the rz projection of the track.

In this way the start of the helix in the xy projection is placed at the point of closest approach to the origin in this projection. If as usual the origin coincides with the nominal interaction point then the starting point defined in this way is a good

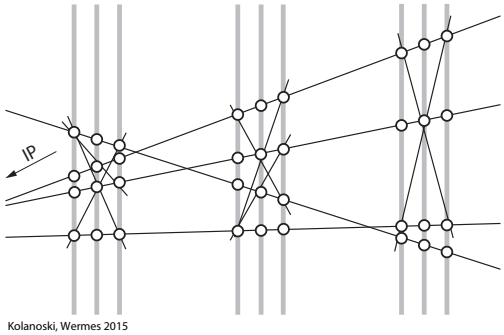


Fig. 9.8 Triplet combinations of hits in nine detector layers, of which three are always more closely arranged in so–called super-layers. In each super-layer hit triplets can be formed which should point into the direction of the interaction point. Obviously many possible combinations can be excluded based on this criterion.

approximation of the particle's production point (in case of primary production). The corresponding coordinates x_0, y_0 can be determined with the help of the angle ψ_0 (fig. 9.7):

$$x_0 = d_0 \cos\psi_0 \,, \qquad y_0 = d_0 \sin\psi_0 \,. \tag{9.23}$$

Usually one starts with fitting the first three parameters using the measured points in the $r\phi$ projection. In the second step the points in the rz projection, measured usually with worse resolution, are used to determine the remaining parameters. In most cases this procedure facilitates *pattern recognition*, that is, the assignment of measurement points to tracks (see next section). A simultaneous fitting of the track to the measurements in both projections is also possible and is preferentially employed after the pattern recognition step.

9.3.6 Pattern recognition

The measurement points registered by a detector for a certain event have to be assigned to individual tracks which are to be reconstructed. This procedure belongs to the general subject of *pattern recognition*, which summarises methods allowing the recognition of regularities or similarities in data which might be incomplete and noisy. Methods of pattern recognition are applied in a variety of fields, like image, voice and character recognition, radar survey and also in event recognition in particle physics [485, 691].

Pattern recognition is the more difficult—but at the same time the more important—the higher the particle density in a detector and correspondingly the higher the number of hit points. The simplest approach is to examine each combination of hits in different layers for consistency with the track model. However, with an increasing particle number the number of combinations becomes rapidly very large so that the procedure becomes inefficient or is even no longer achievable for realistic computing resources. For example, the number of combinations for 10 particles in four detector layers is 10^4 and for 100 particles it would already be $100^4 = 10^8$.

Therefore it is important to reduce the number of possible combinations of hits as early in the reconstruction sequence as possible. One way is to consider only local combinations and impose certain constraints on them. For example, the straight line connecting two or three hits could be required to point to the interaction point within some allowance (fig. 9.8). Such track elements (doublets or triplets) could then be combined in the next step.

An important property of a pattern recognition algorithm is its fault tolerance in case of incomplete and noisy data. For example, it must be possible that measurement points are missing in one or more layers. The algorithm can be made more robust if

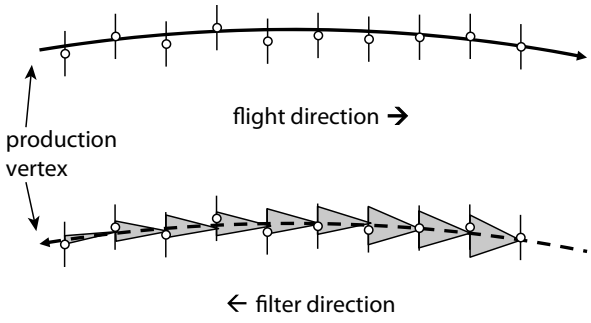


Fig. 9.9 Representation of a track search using the Kalman filter method. Moving towards the production point from the most distant measurement point, the next measurement point will be predicted within some tolerances. With each found point the predictions for subsequent layers will be refined.

outliers, that is, points which are by several (e. g. three) standard deviations away from the fitted curve, can be discarded. The algorithm also has to resolve ambiguities, like the left-right ambiguity in drift cells (section 7.10.4), and should be able to connect the measurements in different projections.

A systematic method of pattern recognition, which provides at the same time a track fitting, is the *Kalman filter* algorithm [429]. Starting from a certain point, for example on a layer of a tracking detector, called the *seed*, a search region for a point in the next layer will be predicted. Usually the measured point which is furthest away from the production vertex of a possible track is chosen as the seed (fig. 9.9). The search region will be restricted by certain conditions, for example that the track should originate from the target within some allowance. If a point in this layer has been found, the prediction for the next layer will be refined. This will be repeated until the last layer before the interaction point is reached and the whole information of all points assigned to the track is available for a determination of the track parameters and their corresponding errors.

The methods discussed so far are local in the sense that only hit connections to next neighbours are considered. Global methods of pattern recognition attempt to assign all measured points such that simultaneously the image of all tracks appears (see e.g. [691]). To these methods belong, amongst others, *neural nets*, *adaptive templates* or transformations from the position space into the parameter space (*Hough transforms*).

9.4 Track parameter resolution

9.4.1 Track models and other ingredients

For magnet spectrometers the momentum and direction resolutions are the most important quality criteria. In general the fitting of the measurements to a track model (section 9.3.3), for example by minimising the quadratic deviations according to expression (9.22), provides the full covariance matrix of the parameters of the track model. The covariance matrix contains the parameter errors and correlations from which the resolutions of momentum, direction and extrapolated positions can be calculated. For the detailed treatment of the track fitting using the matrix formalism and the discussion of the covariance matrix of the parameters we refer to the respective literature, for example [691, 217, 134, 363]. Appendix F contains a compendium of the *least squares method* with a description of the matrix formalism for solving problems linear in the fitted parameters together with applications to straight lines and parabolic functions. In this section we restrict ourselves to linear problems, which for our applications implies for example that the magnetic deflections are small and that

the angle under which the particles cross a detector plane can be assumed to be the same for all measurement planes.

In the following we discuss for various typical configurations of tracking detectors and magnetic fields the resolution effects both due to position measurement errors and due to multiple scattering. All the presented formulae are approximations; more precise estimates usually have to be obtained by measurements or by simulations. We will show with the help of some examples what the resolutions of momentum, direction and vertex position depend on and point out what should be considered in the design phase for the optimisation of a detector.

Uncertainties due to position measurements. When deriving formulae for the case of position uncertainties we will restrict ourselves to the minimally necessary number of measurements for the parameter determination. In this way the resolutions can be analytically calculated without the need for a fitting procedure. The generalisation to many measurement points is described in appendix F. For many cases one finds these formulae in the standard article of Gluckstern [465], which we will frequently refer to in the following.

Uncertainties due to multiple scattering. The errors caused by multiple scattering (see section 3.4) in the material along the trajectory come in addition to the errors due to the position resolution of the detectors. Since in the case of multiple scattering the derivation of the resolution effects via fits leads to rather complex expressions, we retreat to employing the formulae for average scattering angles and offsets, as presented in section 3.4 by (3.106) and the corresponding fig. 3.32. Our results on direction and curvature resolutions mostly agree on the level of a few per cent with those derived by fits including the full covariance matrix, for example as in [363].

The characteristic quantity for multiple scattering which we will use in the following is the standard deviation of the distribution of the projected scattering angle, as defined in section 3.4 (this implies that we assume a Gaussian distribution, ignoring the long tails of the Molière distribution). Here we employ formula (3.102) without the logarithmic correction term,[2]

$$\theta_{ms} = \frac{0.0136\,\mathrm{GeV}/c}{p\,\beta}|z|\sqrt{\frac{x}{X_0}}\,, \tag{9.24}$$

where p, β and z are momentum, normalised velocity and charge (in units of the elementary charge) of the particle, respectively, and x/X_0 is the path length in the scattering medium in units of the radiation length X_0. If the measurement planes are mainly perpendicular to the considered particle trajectory (assuming small deflection angles) and perpendicular to the deflection plane, the length in the scatterer becomes

$$x = d' = d/\sin\theta \qquad \text{(under the conditions specified above)}\,, \tag{9.25}$$

where d is the thickness of the detector plane and θ is the angle between the trajectory and the direction of the magnetic field. In general, if the above assumptions are not applicable, one can replace the terms like $d/\sin\theta$ in the formulae below by the actual path length in the scatterer.

[2]This term has only been dropped in order to simplify the formulae and can easily be re-included if higher precision is required.

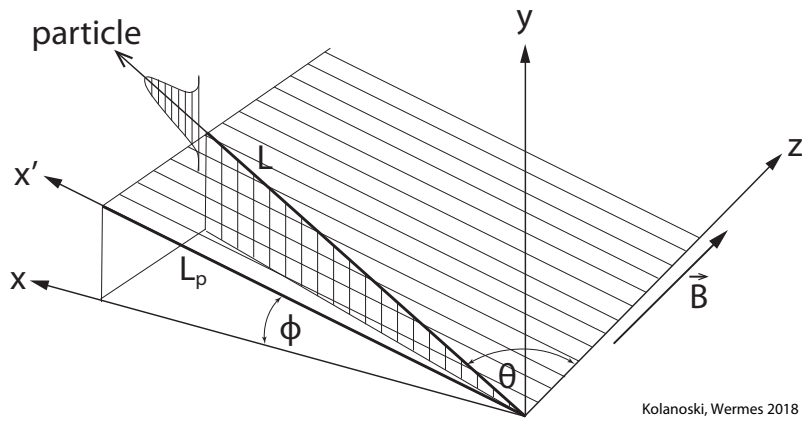


Fig. 9.10 Perspective view of a particle trajectory and its projection onto the deflection plane (xy plane). For simplicity the trajectory is assumed to be approximately straight. The path length within the detector and the path length projected onto the xy plane are L and L_p, respectively. The magnetic field points in the direction of the z-axis and hence the deflection plane is the xy plane. The angle θ is the particle's polar angle defined in the $x'z$ plane spanned by the trajectory and the z-axis (rotated by an angle ϕ against the xz plane) which is perpendicular to the deflection plane and contains also the track projection onto the deflection plane. The distribution of the scattering angle projected onto the vertically hatched plane is indicated, the orthogonal distribution would be in the $x'z$ plane (omitted for clarity).

The scattering angle is defined relative to the particle direction. In the following we often use the projection of the scattering angle onto the deflection plane, that is, the plane perpendicular to the magnetic field, in fig. 9.10 the xy plane. This projection is obtained by first decomposing the spatial scattering angle into its components in the $x'z$ plane (in fig. 9.10 the diagonally hatched plane) and the corresponding perpendicular plane which contains the particle trajectory (vertically hatched). The scattering angle distributions in both projections have standard deviations θ_{ms}. In the next step the scattering angle projection onto the vertically hatched plane has to be further projected onto the deflection plane according to

$$\theta'_{ms} = \theta_{ms}/\sin\theta\,. \tag{9.26}$$

Note that in this formula the $\sin\theta$ term is independent on the actual path through the scatterer, whereas θ_{ms} depends on that path, which only under the conditions formulated above for (9.25) is given by $d' = d/\sin\theta$. For the discussion of momentum resolution, θ'_{ms} is the most essential scattering angle projection. There is a corresponding projection onto the xz plane which, however, we are not referring to in the following because we are mainly interested in the scattering effect on the particle deflection.

9.4.2 Direction resolution in field-free space

First we consider the simplest case of a charged particle passing through a field-free space whose position in y is measured in N detector planes perpendicular to the x-axis at fixed x_i $(i = 1, \ldots, N)$ (situation as in fig. 9.6(a) in front and behind the magnet). The particle trajectory is a straight line, whose projection onto the xy plane

is $y = a + bx$ (we consider only this projection since in field-free space all projections can be treated in the same way). Since we are interested in the direction resolution we consider the error of the slope parameter b.

9.4.2.1 Error due to position measurement

Starting with the minimum number of two measurements we can evaluate the error analytically. Then b is given by

$$b = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_2 - y_1}{D}\,, \tag{9.27}$$

where $D = x_2 - x_1$ is the length over which the track is measured in the xy plane. The y coordinates are measured with resolution σ_{meas}. Under the assumption that σ_{meas} is the same for all measurements the error of the slope is

$$\sigma_b = \sqrt{2}\,\frac{\sigma_{\text{meas}}}{D}\,. \tag{9.28}$$

For N uniformly spaced measurements the error of the slope is given by (F.11) in appendix F:

$$\sigma_b = \frac{\sigma_{\text{meas}}}{D}\sqrt{\frac{12(N-1)}{N(N+1)}}\,. \tag{9.29}$$

The resolution improves as $1/D$ with the lever arm of the measurements and about as $1/\sqrt{N}$ with the number N of measurement points.

9.4.2.2 Error due to multiple scattering

For the case of a continuous scatterer, fig. 3.32 shows the average change of the slope of a straight line by an angle $\langle\psi_{plane}\rangle = \frac{1}{\sqrt{3}}\,\theta_{ms}$ (see eq. (3.106)) which is given as the offset $\langle y_{plane}\rangle$ divided by the scatterer thickness x. In the case of continuous scattering along a path length x we therefore assign as slope error

$$\sigma_b = \langle\psi_{plane}\rangle = \frac{1}{\sqrt{3}}\,\theta_{ms} = \frac{0.0136\,\text{GeV}/c}{\sqrt{3}\,p\,\beta}|z|\,\sqrt{\frac{x}{X_0}}\,. \tag{9.30}$$

On the right-hand side the expression for θ_{ms} given in (9.24) has been used. A comparison of (9.29) and (9.30) shows that the slope error due to multiple scattering decreases with momentum while the error due to position measurement is momentum independent.

Now we want to consider the case that multiple scattering occurs in N detector planes at positions $x_i,\ i = 1, \ldots, N$ distributed over the length D (that means $x_N - x_1 = D$). In each plane scattering leads to an average offset $\langle y_i^{ms}\rangle$ in the last plane at x_N

$$\langle y_i^{ms}\rangle = \theta'_{ms,i}\,(x_N - x_i) = \frac{\theta_{ms,i}}{\sin\theta}\,(x_N - x_i) = \frac{0.0136\,\text{GeV}/c}{p\,\beta\sin\theta}|z|\,\sqrt{\frac{d'_i}{X_0}}\,(x_N - x_i)\,. \tag{9.31}$$

For each layer i the projected scattering angle $\theta'_{ms,i}$ is given by (9.26) and (9.24) with $x = d'_i$ where d'_i is the layer thickness along the particle path. The total average offset

$\langle y^{ms} \rangle$ is determined by quadratic summation of all $N-1$ contributing offsets (scattering in the last plane does not contribute) as given by (9.31). This finally yields an estimator of the slope error (corresponding to $\langle \psi_{plane} \rangle$ in the continuous case (9.30)):[3]

$$\sigma_b = \frac{\langle y^{ms} \rangle}{D} = \frac{1}{D} \sqrt{\sum_{i=1}^{N-1} \theta'^2_{ms,i} \left(x_N - x_i\right)^2} = \frac{1}{D \sin\theta} \sqrt{\sum_{i=1}^{N-1} \theta^2_{ms,i} \left(x_N - x_i\right)^2} \,. \tag{9.32}$$

In the special case of equally distributed detector layers with spacing $D/(N-1)$ and for equal scattering angles $\theta_{ms,i} = \theta_{ms,sl}$ (sl = single layer) and equal thickness $d_i = d$, (9.32) can be readily evaluated yielding

$$\sigma_b = \frac{\theta_{ms,N}}{\sin\theta} \sqrt{\frac{N(2N-1)}{6(N-1)^2}} := \frac{\theta_{ms,N}}{\sin\theta} \sqrt{E_N} \tag{9.33}$$

$$\text{with} \quad \theta^2_{ms,N} = \sum_{i=1}^{N-1} \theta^2_{ms,i} = (N-1)\, \theta^2_{ms,sl} \quad \text{and} \quad E_N = \frac{N(2N-1)}{6(N-1)^2} \,. \tag{9.34}$$

The introduction of E_N also allows us to apply the formula to other detector configurations. Using (9.34) and (9.24) the direction resolution due to multiple scattering for the equal spacing case then becomes

$$\sigma_b = \frac{\theta_{ms,sl}}{\sin\theta} \sqrt{N-1} \sqrt{E_N} = \frac{0.0136\,\text{GeV}/c}{p\,\beta \sin\theta} |z| \sqrt{\frac{d/\sin\theta}{X_0}} \sqrt{\frac{N(2N-1)}{6(N-1)}} \,. \tag{9.35}$$

Here we have also used $d' = d/\sin\theta$ assuming the conditions specified for (9.26) (the track projection onto the deflection plan passes perpendicular through the detector planes).

Evaluating (9.33) for $N = 2$ and $\sin\theta = 1$ yields, as expected, $\sigma_b = \theta_{ms}$ which is in this case equal to $\theta_{ms,sl}$. If N becomes very large we expect to recover the continuum case as given in (3.106) for $\langle \psi_{plane} \rangle$ and indeed, the large N limit yields $\sigma_b = \theta_{ms,N}/\sqrt{3}$ (as in (9.30)).

Obviously the resolution becomes worse with increasing number of detector layers. Therefore if only multiple scattering would contribute to the error one would always choose the minimal number of layers, in this case $N = 2$. In practice one has to optimise the number of layers with respect to other requirements, like spatial resolution and pattern recognition. Corresponding considerations hold for the discussions of different multi-layer detectors and different parameters in the following.

9.4.3 Momentum resolution: measurements outside the magnetic field

We consider a spectrometer as in fig. 9.6(a), but without the detectors inside the magnet, such that the particle trajectories are only measured in front and behind the magnet. In this case the momentum is determined by the change of the direction in front and behind the magnet. The deflection angle α can be obtained from fig. 9.11:

[3] In this simplified approach we consider the scattering contributions of each plane as independent, thus avoiding correlation terms which arise when fitting to the coordinates measured in each plane.

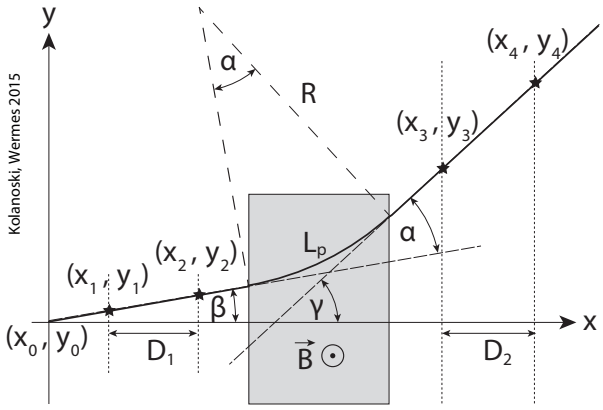


Fig. 9.11 Measurement of a particle trajectory in front and behind a magnet. The momentum is determined by the deflection angle α, the angle between the two straight lines.

$$\alpha = \frac{L_p}{R} = \frac{|q|}{p_T} L_p \, B \,, \tag{9.36}$$

where L_p is the path length inside the magnet (we only consider the projection onto the plane perpendicular to the magnetic field). The right-hand side of (9.36) follows from (9.5). The *bending power* of the magnet is proportional to L_pB or in an inhomogeneous field proportional to the path integral over the momentum-deflecting field component:

$$\alpha \propto \int_{L_p} B \, dl \,. \tag{9.37}$$

9.4.3.1 Error due to position measurements

According to fig. 9.11 the angle α is given by the difference of the slope angles of the line behind and before the magnet, $\alpha = \gamma - \beta$. Let the straight lines in front and behind the magnet both be determined by the minimal number of two measurement points. The slope angles are then given by (designations as in fig. 9.11)

$$\tan\beta = \frac{y_2 - y_1}{x_2 - x_1} = \frac{y_2 - y_1}{D_1} \,, \qquad \tan\gamma = \frac{y_4 - y_3}{x_4 - x_3} = \frac{y_4 - y_3}{D_2} \,. \tag{9.38}$$

Here $D_1 = x_2 - x_1$ and $D_2 = x_4 - x_3$ are the fixed distances of the measurement planes in which at fixed x coordinates the y coordinates are measured with the resolution σ_{meas}. Under the assumption that σ_{meas} is the same for all measurements the errors of the slopes are

$$\sigma_{\tan\beta} = \sqrt{2}\,\frac{\sigma_{\text{meas}}}{D_1} \,, \qquad \sigma_{\tan\gamma} = \sqrt{2}\,\frac{\sigma_{\text{meas}}}{D_2} \,. \tag{9.39}$$

In the approximation of small angles one obtains for the deflection angle α:

$$\alpha = \gamma - \beta \approx \frac{y_4 - y_3}{D_2} - \frac{y_2 - y_1}{D_1} \,. \tag{9.40}$$

The error of α follows from error propagation:

$$\sigma_\alpha = \sqrt{2}\,\sigma_{\text{meas}} \sqrt{\frac{1}{D_1^2} + \frac{1}{D_2^2}} \tag{9.41}$$

or for the same distances of the measurement planes ($D_1 = D_2 = D$)

$$\sigma_\alpha = \frac{2\,\sigma_{\text{meas}}}{D}\,. \tag{9.42}$$

Hence the resolution is proportional to the position resolution σ_{meas} of a measurement and inversely proportional to the lever arm D over which the straight lines are measured. However, the resolution cannot be arbitrarily improved by enlarging D because there are always conditions constraining the size of the detector. For example, for a larger distance of a magnet from the interaction point its angular acceptance becomes smaller. At larger distances of the measurement planes the alignment of the detectors may become less precise and may compensate the gain provided by the lever arm.

For N equally spaced measurement points we take the resolution of the slope of one of the lines as in (9.29). For the same number of measurement points N in front and behind the magnet the error for the deflection then becomes

$$\sigma_\alpha = \sqrt{\frac{24(N-1)}{N(N+1)}}\;\frac{\sigma_{\text{meas}}}{D}\,. \tag{9.43}$$

For $N = 2$ this corresponds to (9.42).

Using (9.36) the resulting momentum resolution is obtained by error propagation:

$$p_T = \frac{|q|}{\alpha}\,L_p\,B \qquad \Rightarrow \qquad dp_T = \frac{p_T^2}{|q|L_pB}\,d\alpha\,. \tag{9.44}$$

Replacing $d\alpha$ by σ_α from (9.43) this yields for the relative momentum error due to the position measurements:

$$\left(\frac{\sigma_{p_T}}{p_T}\right)_{meas} = \frac{\sigma_{\text{meas}}\,p_T}{|q|\,L_p\,B\,D}\sqrt{\frac{24(N-1)}{N(N+1)}} = \frac{\sigma_{\text{meas}}\,p_T}{0.3\,|z|\,L_p\,B\,D}\sqrt{\frac{24(N-1)}{N(N+1)}}\,, \tag{9.45}$$

where on the right-hand side the momentum is given in units GeV/c, all lengths in metres and the magnetic field in tesla (see (9.10)). The equation holds for N measurements distributed equally over the length D, both in front and behind the deflecting magnet (i.e. there is a total of $2\,N$ measurements).

9.4.3.2 Error due to multiple scattering

In the following we consider two specific cases of multiple scattering (both for measurements outside the magnetic field). In the first case multiple scattering occurs only in the detector planes (discrete scatterers); in the second case scattering occurs continuously inside the magnetic field.

Error due to multiple scattering outside the magnetic field. If multiple scattering occurs only in the detector planes and other scattering can be neglected, we can use the result for the slope error of straight lines given in (9.33), which holds for N detector layers each of thickness d and equally spaced over the length D.

As before the error on the deflection angle is obtained by quadratically adding the errors of the track slopes in front and behind the magnet. Assuming that both are measured with equal precision one obtains from (9.33):

$$\sigma_\alpha^{scat} = \frac{\theta_{ms,N}}{\sin\theta}\sqrt{\frac{N(2N-1)}{3(N-1)^2}} = \frac{\theta_{ms,sl}}{\sin\theta}\sqrt{\frac{N(2N-1)}{3}}\,. \tag{9.46}$$

The average scattering angle $\theta_{ms,N}$ is defined as in (9.34) by quadratically summing the scattering contributions of the detector planes except for the last plane.

Substituting $d\alpha$ in (9.44) by σ_α^{scat} in (9.46) we obtain the momentum error due to multiple scattering for scattering in the detector planes only:

$$\begin{aligned}\left(\frac{\sigma_{p_T}}{p_T}\right)_{scat} &= \frac{p_T\,\theta_{ms,N}}{|q|L_p\,B\,\sin\theta}\sqrt{\frac{N(2N-1)}{3(N-1)^2}} \\ &= \frac{0.0136}{0.3\,\beta\,L_p\,B}\sqrt{\frac{(N-1)\,d'}{X_0}}\sqrt{\frac{N(2N-1)}{3(N-1)^2}} \qquad [L_p] = \mathrm{m},\ [B] = \mathrm{T}\,.\end{aligned} \tag{9.47}$$

In the second line we have inserted the scattering angle from (9.24) together with the unit conversion as given in (9.10). As before $d' = d/\sin\theta$ if the detector planes are perpendicular to the track projection onto the deflection plane.

Example. Let $N = 5$ measurement points be distributed over a length of $D = 1\,\mathrm{m}$ both in front and behind the magnet (fig. 9.11). For all points the position resolution is $\sigma_{\mathrm{meas}} = 100\,\mu\mathrm{m}$, the magnetic field strength and length are $B = 1.5\,\mathrm{T}$ and $L_p = 2\,\mathrm{m}$, respectively, and the thickness of a detector plane along the particle trajectory is 2% of a radiation length. For a particle with charge $q = e$ this yields the resolution due to the measurement errors

$$\frac{\sigma_{p_T}}{p_T} = 0.2 \times 10^{-3}\,p_T/(\mathrm{GeV}/c)\,, \tag{9.48}$$

which is 0.2% for momenta of $10\,\mathrm{GeV}/c$ and 2% for $100\,\mathrm{GeV}/c$.

In order to compare this resolution due to position measurement to the resolution due to scattering, we evaluate (9.47) with the assumptions $p = p_T,\ \beta \approx 1,\ z = 1$. The average scattering angle is computed using (9.33):

$$\theta_{ms,N} \approx \frac{0.0136\,\mathrm{GeV}/c}{p_T}\,2\sqrt{0.02} = \frac{0.00385\,\mathrm{GeV}/c}{p_T}\,. \tag{9.49}$$

Inserting this into (9.47) one finds for the resolution contribution due to scattering for the given parameters:

$$\left(\frac{\sigma_{p_T}}{p_T}\right)_{scat} = 0.41\%\,, \tag{9.50}$$

This means that the momentum resolution is dominated by multiple scattering for momenta up to about $20\,\mathrm{GeV}/c$. The influence of multiple scattering can be decreased by employing a smaller number of layers (at the expense of position resolution; see the corresponding remark at the end of section 9.4.2 on page 391).

Error due to multiple scattering inside the magnetic field. We now consider the case that the particles are only scattered by multiple scattering in the magnetic field while the scattering in the detectors should be negligible. An important example for this situation is a muon spectrometer with magnetised iron which serves at the same time both as hadron absorber and as momentum analyser (see section 9.2.4).

The deflection angle α acquires an additional uncertainty $\sigma_\alpha^{scat} = \theta_{ms}/\sin\theta$, where the scattering angle is given by (9.24) and (9.26). With (9.44) the contribution of multiple scattering to the momentum resolution becomes

$$dp_T = \frac{p_T^2}{|q|L_p\,B}\frac{\theta_{ms}}{\sin\theta} = \frac{p_T^2}{|q|L_p\,B}\,\frac{0.0136\,\mathrm{GeV}/c}{p\,\beta\,\sin\theta}\,|z|\sqrt{\frac{L_p/\sin\theta}{X_0}}\,, \tag{9.51}$$

where θ is the angle between the particle trajectory and the magnetic field, $p_T = p\sin\theta$ (see (9.6)) and X_0 is the radiation length of the scattering material. The term $L_p/(\sin\theta\,X_0)$ is the full path length within the iron in units of the radiation length. The contribution of multiple scattering to the momentum resolution is therefore (we use again (9.10)):

$$\left(\frac{\sigma_{p_T}}{p_T}\right)_{scat} = \frac{0.0136}{0.3\,\beta\,L_p\,B}\sqrt{\frac{L_p/\sin\theta}{X_0}}, \qquad [L_p] = \mathrm{m},\ [B] = \mathrm{T}\,. \tag{9.52}$$

This contribution depends on the momentum only via β and therefore dominates at small momenta over the contribution due to position measurements, which is small at low p_T and increases proportional to p_T. A larger field length L_p improves the resolution as $1/\sqrt{L_p}$. However, L_p cannot be arbitrarily long because the absorber thickness determines the momentum threshold.

Example. We consider a muon spectrometer which consists of magnetised iron ($X_0 = 1.76\,\mathrm{cm}$) of length $L_p = 2\,\mathrm{m}$ (correspondig to the range of 2.6-GeV muons [762]) and a field of $B = 1.5\,\mathrm{T}$. The particles move approximately in the plane perpendicular to the magnetic field ($\sin\theta \approx 1$). Inserting this into (9.52) yields a momentum resolution due to multiple scattering of

$$\frac{\sigma_{p_T}}{p_T} = 16\%\,. \tag{9.53}$$

Up to some $100\,\mathrm{GeV}/c$ this error is typically much larger than the error due to the position measurement. Detectors with only mediocre position resolution and/or with fewer measurement points would not significantly deteriorate the total momentum resolution.

9.4.4 Momentum resolution: measurements in the magnetic field

We consider now the case where the curvature of a track is measured inside the magnetic field. The connection between the transverse momentum p_T and the curvature κ, as defined in (9.18), is

$$p_T = |q|\,B\,R = \frac{|q|\,B}{|\kappa|}\,. \tag{9.54}$$

9.4.4.1 Error due to position measurements

The measurement of a momentum in a homogeneous magnetic field boils down to the determination of the sagitta, that is, the largest perpendicular distance of the trajectory from the connecting line between the particle's entrance into and exit from the magnetic field volume (fig. 9.12). From the figure we obtain the following relations for the sagitta:

$$\frac{R-s}{R} = \cos\frac{\alpha}{2} \approx 1 - \frac{\alpha^2}{8} \qquad \text{and} \qquad \frac{L_p}{2\,R} = \sin\frac{\alpha}{2} \approx \frac{\alpha}{2}\,. \tag{9.55}$$

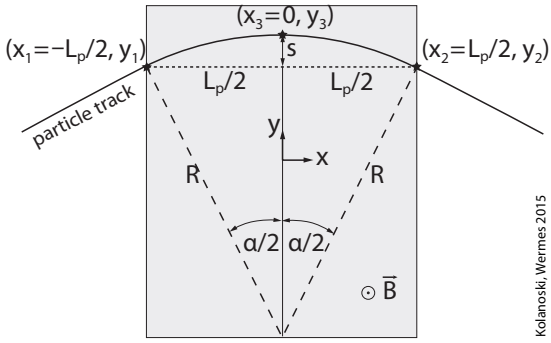


Fig. 9.12 Geometric relations for the deflection of a charged particle in a homogeneous magnetic field of length L_p. The drawing delineates the sagitta s of the circular arc in the magnetic field and the deflection angle α of the particle.

The approximations are valid for small deflection angles, respectively large momenta. This yields for the sagitta

$$s \approx \frac{R\,\alpha^2}{8} = \frac{1}{8}\frac{L_p^2}{R} = \frac{1}{8}\,L_p^2\,|\kappa|\,. \tag{9.56}$$

Hence the sagitta s is proportional to the curvature κ and therefore the error of the sagitta determination is also proportional to the error of the curvature from which the momentum is determined:

$$\sigma_\kappa = \frac{8}{L_p^2}\,\sigma_s\,. \tag{9.57}$$

For the determination of a circular trajectory at least three measurement points are required which are optimally placed at the beginning, the end and in the centre of the magnetic field, as depicted in fig. 9.12 (the points are given by their coordinates in the deflection plane). The three points may be measured by three detector planes positioned perpendicular to the mean direction of the particle trajectories (x direction) at $x = -L_p/2,\, 0,\, L_p/2$. The detectors deliver the y coordinates with errors σ_{meas}, assumed to be the same for all three points. From the measured points the sagitta is derived according to

$$s = y_3 - \frac{y_1 + y_2}{2} \qquad \Rightarrow \qquad \sigma_s = \sqrt{\sigma_{\text{meas}}^2 + \frac{1}{4}\,2\,\sigma_{\text{meas}}^2} = \sqrt{\frac{3}{2}}\,\sigma_{\text{meas}}\,. \tag{9.58}$$

With (9.57) the error of the curvature then follows as

$$\sigma_\kappa = \sqrt{\frac{3}{2}}\,\frac{8}{L_p^2}\,\sigma_{\text{meas}} = \frac{\sqrt{96}}{L_p^2}\,\sigma_{\text{meas}}\,. \tag{9.59}$$

The factor $\sqrt{96}$ is valid for $N = 3$ measurements. The general formula for N equally spaced measurements reads[4] (see the related eq. (F.16) in appendix F and [465])

$$\sigma_\kappa = \frac{\sigma_{\text{meas}}}{L_p^2}\sqrt{\frac{720(N-1)^3}{(N-2)N(N+1)(N+2)}}\,. \tag{9.60}$$

[4]The formula traces back to Gluckstern [465]. It should be noted that in [465] the measurements are numbered 0 to N so that there the number of measurements is $N + 1$ while in this book the number of measurements is N.

For many measurements ($N \gtrsim 10$) one obtains

$$\sigma_\kappa \approx \frac{\sigma_{\text{meas}}}{L_p^2}\sqrt{\frac{720}{N+4}}\,. \tag{9.61}$$

The resolution of the transverse momentum can be calculated by error propagation using the relation between transverse momentum and curvature in (9.54):

$$\sigma_{p_T} = \frac{p_T^2}{|q|\,B}\sigma_\kappa\,. \tag{9.62}$$

When deriving the p_T resolution by Gaussian error propagation one has to keep in mind that the error of the transverse momentum is not normally distributed (even if the error of the curvature is), in particular at very large momenta (see also the corresponding discussion of the p_T dependence in the listing below). For N equally spaced points and in the limit of large N the resolution of the transverse momentum is then obtained using (9.61) together with (9.10):

$$\left(\frac{\sigma_{p_T}}{p_T}\right)_{\text{meas}} = \frac{p_T}{0.3|z|}\frac{\sigma_{\text{meas}}}{L_p^2 B}\sqrt{\frac{720}{N+4}}\,, \qquad [p_T] = \text{GeV}/c,\ [L_p],\ [\sigma_{\text{meas}}] = \text{m},\ [B] = \text{T}\,. \tag{9.63}$$

This relation, called the 'Gluckstern formula', describes the momentum resolution at high momenta where multiple scattering can be neglected [465].

The momentum resolution σ_{p_T} depends on the position resolution σ_{meas} of the points in the plane of momentum deflection, the number N of measurements and the length L_p of the track in the projection perpendicular to the magnetic field. The dependences of the equation on each of the parameters provide indications on how the momentum resolution can be optimised:

$\propto \sigma_{\text{meas}}$*:* The position resolution of the detector should be optimised up to the point where other resolution contributions, for example multiple scattering, dominate. Other precision limitations come, for example, from mechanical tolerances and temperature dependences which can introduce larger uncertainties than the intrinsic detector precision, in particular in large detector arrangements.

$\propto 1/L_p^2$*:* A large gain in momentum resolution comes from the length of the measured part of the track, that is, from the size of the detector. However, the costs for a detector, for example at a collider, scale at least as strongly as L_p^2.

$\propto 1/B$*:* For the same momentum a stronger magnetic field yields a larger curvature of the track leading to an improved resolution. However, too large curvatures hinder efficient pattern recognition, for example by curling low-energy tracks generating high hit density ('curlers').

$\propto 1/\sqrt{N}$*:* A large number of measurements improves the resolution, but merely by the square root of the number of measurements. However, having many measurement points is also beneficial for pattern recognition and for measurements of the specific ionisation (dE/dx measurements, see section 14.2.2).

$\propto p_T$*:* For high momenta the tracks are less curved and eventually can no longer be distinguished from straight lines. Therefore the resolution function is not symmetric in p_T. However, the resolution function of the curvature, $|\kappa| = 1/R \propto 1/p_T$, is

approximately normally distributed. In simulations, therefore, κ is usually generated according to a normal distribution and then converted into p_T.

The highest resolvable momenta can be defined by the condition that the sign of the curvature can still be determined within n standard deviations (e.g. $n = 2$):

$$|\kappa| = \frac{0.3\,|z|\,B}{p_T} > n\sigma_\kappa \quad \Rightarrow \quad p_T < \frac{0.3\,|z|\,B}{n\sigma_\kappa}\,. \tag{9.64}$$

Here the units are chosen as in (9.9). From (9.61), which holds for a large number of measurement points, it follows that $n\,\sigma$ separation is possible for momenta

$$p_T < \frac{0.3\,|z|\,B\,L_p^2}{n\sigma_{\text{meas}}}\sqrt{\frac{N+4}{720}}\,, \qquad [p_T] = \text{GeV}/c,\ [L_p],\ [\sigma_{\text{meas}}] = \text{m},\ [B] = \text{T}\,. \tag{9.65}$$

Example. For a tracking detector with $N = 15$, $\sigma_{\text{meas}} = 100\,\mu\text{m}$, $B = 1\,\text{T}$, $L_p = 1\,\text{m}$ and for a particle with $|z| = 1$ the momentum limit is about $250\,\text{GeV}/c$ if 2-σ separation is requested. See also the discussion of tracking versus calorimetry in section 15.1.

9.4.4.2 Error due to multiple scattering

The scattering changes the sagitta on average by an amount which corresponds to the mean offset at half of the scatterer thickness $\langle s_{plane}\rangle$ in fig. 3.32 on page 68 which is given by (3.106):

$$\sigma_s = \langle s_{plane}\rangle = \frac{1}{4\sqrt{3}}\,L_p\,\frac{\theta_{ms}}{\sin\theta}\,. \tag{9.66}$$

For continuous scattering over the path length L the scattering angle θ_{ms} is defined in (9.24) with $x = L = L_p/\sin\theta$ (as before θ is the angle between the trajectory and the magnetic field direction). For scattering in discrete scattering layers (the detector planes) we proceed as in section 9.4.2 and relate $\langle s_{plane}\rangle$ to the average offset $\langle y^{ms}\rangle$ in the last plane (at x_N) as defined in (9.31):

$$\sigma_s = \langle s_{plane}\rangle = \frac{1}{4}\langle y^{ms}\rangle = \frac{1}{4\sin\theta}\sqrt{\sum_{i=1}^{N-1}\theta_{ms,i}\,(x_N - x_i)}\,. \tag{9.67}$$

Here we used (3.106) and assumed a correspondence between $\langle y_{plane}\rangle$ in the continuous case and $\langle y^{ms}\rangle$ in the discrete case. The $\theta_{ms,i}$ with $i = 1, \ldots, N-1$ are the average scattering angles in the individual detector layers. In analogy to (9.33) we obtain for the case of detector planes that are uniformly distributed over the length L_p the expression

$$\sigma_s = \frac{1}{4}\,\frac{\theta_{ms,N}}{\sin\theta}\,L_p\sqrt{\frac{N(2N-1)}{6(N-1)^2}}\,, \tag{9.68}$$

with $\theta_{ms,N}$ being the average scattering angle as defined in (9.33). According to (9.57) the error of the sagitta relates to the error of the curvature:

$$\sigma_\kappa = \frac{8}{L_p^2}\sigma_s = \frac{\theta_{ms,N}}{L_p\,\sin\theta}\sqrt{\frac{2\,N(2N-1)}{3(N-1)^2}} := \frac{\theta_{ms,N}}{L_p}\,\sqrt{C_N}\,. \tag{9.69}$$

Introducing C_N on the right-hand side and assuming that in general σ_κ depends linearly on $\theta_{ms,N}/L_p$ the expression can be applied to different detector distributions.[5] For uniformly distributed detector layers we have from (9.69):

$$C_N = \frac{2N(2N-1)}{3(N-1)^2}\,. \tag{9.70}$$

For the minimally necessary number of detector planes $N = 3$ one obtains $C_N = 2.5$ and for $N \to \infty$ the continuous scattering case in (9.66) is approached with $C_N \to 1.33$.

Using (9.69) and (9.62) the relative momentum resolution can be calculated to be

$$\left(\frac{\sigma_{p_T}}{p_T}\right)_{\text{scat}} = \frac{p_T}{|q|\,B}\,\frac{\theta_{ms,N}}{L_p \sin\theta}\,\sqrt{C_N}\,. \tag{9.71}$$

For uniformly distributed detector layers and with $\theta_{ms,N} = \sqrt{N-1}\,\theta_{ms,sl}$ as given by (9.34) the momentum resolution due to scattering is (using again (9.24) together with (9.10))

$$\left(\frac{\sigma_{p_T}}{p_T}\right)_{\text{scat}} = \frac{0.0136}{0.3\,\beta\,L_p\,B}\sqrt{\frac{(N-1)\,d/\sin\theta}{X_0}}\,\sqrt{C_N}\,, \quad [p_T] = \text{GeV}/c,\ [L_p] = \text{m},\ [B] = \text{T}\,. \tag{9.72}$$

Here the length x in (9.24) was set to $x = (N-1)\,d/\sin\theta$, which is the total path length in the scattering detector planes, each of thickness d (again, θ is the angle between the track and the magnetic field).

It is interesting to note that here the resolution is also proportional to $1/B$, the interpretation being that for stronger curvatures the influence of scattering is relatively smaller.

9.4.4.3 Total momentum resolution

The full transverse momentum resolution is obtained by quadratically adding the two different contributions:

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{\left(\frac{\sigma_{p_T}}{p_T}\right)^2_{\text{meas}} + \left(\frac{\sigma_{p_T}}{p_T}\right)^2_{\text{scat}}} := \sqrt{(a\,p_T)^2 + b^2}\,. \tag{9.73}$$

The right-hand side is a common parametrisation of track resolutions (for $\beta \approx 1$) with the coefficients a and b describing the contributions from the measurement error and from multiple scattering, respectively. This is graphically displayed in fig. 9.13. For small momenta the resolution reaches a saturation value given by the multiple scattering term, while at high momenta the measurement term dominates. By changing the number of detector planes both terms can be adjusted in order to optimise for a particular momentum range (see the corresponding remark at the end of section 9.4.2 on page 391). Central tracking systems of modern collider experiments typically reach at high momenta: $\sigma_{p_T}/p_T \approx (0.1\text{–}0.2\%)\,p_T$.

Example. As an example we quote resolutions obtained with the central jet chamber of the OPAL experiment. The relevant detector parameters are

[5] The C_N are similarly defined as the corresponding coefficients in [465]. Note, however, that in [465] the scattering is continuous while here the scattering occurs in discrete planes (where also the measurement points are located).

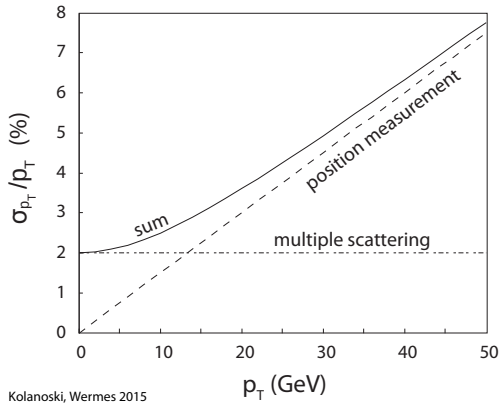


Fig. 9.13 Sketch of the momentum resolution of a magnet spectrometer according to (9.73), combining contributions due to the measurement accuracy of the detector and due to multiple scattering. The numerical values on the axes correspond to the resolution of the OPAL jet chamber according to (9.74) [202].

$$L_p = 1.6\,\mathrm{m},\ B = 0.435\,\mathrm{T},\ N = 159,\ \sigma_{\mathrm{meas}} = 135\,\mu\mathrm{m}.$$

Then the calculated contribution from the position measurement is

$$\sigma_{p_T}/p_T = 8.5 \times 10^{-4}\, p_T\,.$$

The actual resolution measured by OPAL is [202]:

$$\frac{\sigma_{p_T}}{p_T} = \sqrt{(0.0015\, p_T)^2 + (0.02)^2} \qquad (p_T \text{ in GeV}/c)\,. \tag{9.74}$$

The measurement and scattering terms become equal at about 13 GeV.

9.4.5 Direction resolution in a magnetic field

To completely reconstruct the momentum vector (usually at the production point of the track) in addition to p_T two independent angles also have to be measured. The case where the angles are reconstructed in a field-free space through straight lines is treated in section 9.4.2. Here we discuss the direction resolution for the case that the trajectory is measured in a magnetic field and then extrapolated to the interaction point or the entrance point into the magnetic field, respectively.

9.4.5.1 Error due to position measurement

For measurements in a magnetic field we are referring to the configuration in fig. 9.6(a), but with measurements inside the magnet only. In order to determine the direction resolution we use the linearised form (9.21) of a circle trajectory:

$$y = a + bx + \frac{1}{2}cx^2\,. \tag{9.75}$$

The parameter c is the curvature which is the same as κ used in other formulae. In appendix F the errors of a, b, c ($\sigma_a,\ \sigma_b,\ \sigma_c$) and their covariances ($\sigma_{ab},\ \sigma_{bc},\ \sigma_{ac}$) are determined for the case that the origin of the x-axis is the centre of gravity of N measurements which are equally spaced in x. In a more general case, one wants to determine the slope at an arbitrary x coordinate x_0, corresponding to the derivative of the function (9.75) at $x = x_0$:

$$\left.\frac{dy}{dx}\right|_{x_0} = b + c\,x_0 = b' \qquad \Rightarrow \qquad \sigma_{b'}\big|_{x_0} = \sqrt{\sigma_b^2 + \sigma_c^2\, x_0^2 + 2\,\sigma_{bc}\, x_0}\,. \tag{9.76}$$

In the case of equally spaced measurements with their centre of gravity in the origin of x the error of b' is calculated from the standard deviations of b and c and their correlation which are given in appendix F by the equations (F.15) to (F.17):

$$\left(\sigma_{b'}\big|_{x_0}\right)_{meas} = \sqrt{\sigma_b^2 + \sigma_c^2\, x_0^2} = \frac{\sigma_{\text{meas}}}{L_p}\sqrt{\frac{12(N-1)}{N(N+1)} + \frac{x_0^2}{L_p^2}\frac{720(N-1)^3}{N(N-2)(N+1)(N+2)}}\,. \tag{9.77}$$

Note that for the chosen distribution of measurements (detector planes) the correlation term in (9.76) vanishes, $\sigma_{bc} = 0$.

In the considered configuration (fig. 9.6(a) with measurements only inside the magnetic field) the direction of the particle at its generation point is given by the track direction at the entrance into the magnetic field at $x = x_1$ (fig. 9.12). Therefore the direction error has to be determined at $x_0 = x_1 = -L_p/2$:

$$\left(\sigma_{b'}\big|_{x_0=x_1}\right)_{meas} = \frac{\sigma_{\text{meas}}}{L_p}\sqrt{\frac{12(2N-1)(8N-11)(N-1)}{N(N-2)(N+1)(N+2)}} \overset{N \text{ large}}{\longrightarrow} \frac{\sigma_{\text{meas}}}{L_p}\sqrt{\frac{192}{N+3.9}}\,. \tag{9.78}$$

The numerical value 3.9 on the right-hand side is obtained asymptotically; for moderate values of N the value 4 gives a better approximation. Equation (9.78) shows that the direction resolution in the plane of magnetic deflection is proportional to σ_{meas}/L_p, which is the position uncertainty relative to the length over which the measurements are taken.

9.4.5.2 Error due to multiple scattering

For the calculation of the direction uncertainty of a curved track due to multiple scattering the covariance elements σ_b, σ_c and $\sigma_{bc}\,(=0)$ are given in (9.33) and (9.69), respectively:

$$\sigma_b^2 = \frac{\theta_{ms,N}^2}{\sin^2\theta}\,E_N\,, \qquad \sigma_c^2 = \frac{\theta_{ms,N}^2}{L_p^2\,\sin^2\theta}\,C_N\,, \qquad \sigma_{bc} = 0\,. \tag{9.79}$$

The factors E_N and C_N are given for equal spacing of N detector planes by (9.33) and (9.70), respectively. Inserting (9.79) into (9.76) we obtain the contribution of multiple scattering to the direction resolution:

$$\left(\sigma_{b'}\big|_{x_0}\right)_{scat} = \frac{\theta_{ms,N}}{\sin\theta}\sqrt{E_N + \frac{x_0^2}{L_p^2}C_N}\,. \tag{9.80}$$

Extrapolation to the entrance of the track into the magnet at $x_0 = x_1 = -L_p/2$ yields with (9.33):

$$\left(\sigma_{b'}\big|_{x_0=x_1}\right)_{scat} = \frac{0.0136\,\text{GeV}/c}{p\,\beta\,\sin\theta}|z|\sqrt{\frac{(N-1)\,d/\sin\theta}{X_0}}\sqrt{\frac{N(2N-1)}{3(N-1)^2}}\,. \tag{9.81}$$

9.4.6 Impact parameter resolution

In high energy experiments secondary vertices signal weak decays of heavy quarks and leptons (see section 14.6.2). As an example fig. 9.14 shows an event at a collider

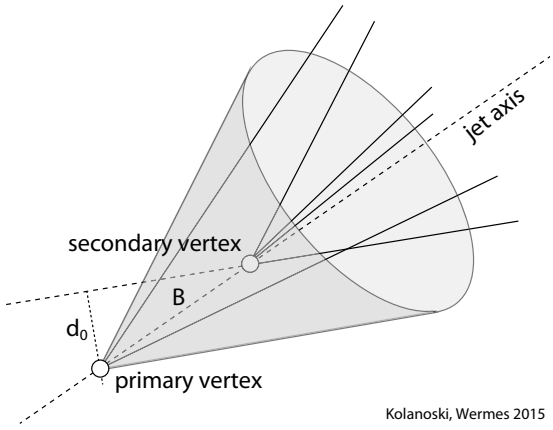


Fig. 9.14 Sketch of a b-quark jet in a collider experiment. In the jet a bottom hadron (B) decays at a secondary vertex which is drawn in the plane perpendicular to the beam. In this projection the primary vertex usually can be reconstructed with a precision of some micrometres using the remaining tracks. For each track the distance of closest approach from the primary vertex, the impact parameter d_0, can be used to estimate its probability to come from a secondary vertex.

which contains a jet in which a bottom hadron decays (see also fig. 14.31 on page 578). The presence of a secondary vertex is suggested if the extrapolations of one or more reconstructed tracks miss the primary interaction point (often given by the nominal beam position). The failure to extrapolate to the primary vertex is quantified by the *impact parameter* d_0, the distance of closest approach to the interaction point in the momentum deflection plane (see fig. 9.14 and also fig. 9.7 on page 385). The significance of the impact parameter determination is given by the ratio of the determined value to its resolution, d_0/σ_{d_0} (see also section 14.6.2). This significance is used as a selection criterion for tracks which are candidates for being produced at a secondary vertex (e.g. requiring $d_0/\sigma_{d_0} > 2, 3, \ldots$).

The impact parameter resolution is a measure for the ability of a detector to reconstruct vertices. Since the 1980s collider experiments have been equipped with dedicated vertex detectors (see chapters 8 and 14) which are positioned as close as possible to the interaction point. With position resolutions per layer of around 10 µm they are able to resolve vertex separations of some 100 µm.

9.4.6.1 Error due to position measurements

We first consider a linear approximation of the trajectory's projection onto the deflection plane given by the straight line $y = a + bx$. At high momenta and small distances to the interaction point this approximation is also usually sufficient in magnetic fields to estimate the vertex resolution. Extrapolating to an x coordinate x_0 (e.g. to the nominal interaction point) one can determine the y coordinate y_0 and its distance from the nominal value y_{pv} of the primary vertex. Without loss of generality we choose a coordinate system where the track projection points in the x direction. Then the impact parameter is given by $d_0 \approx y_0 - y_{pv}$ and its error can be evaluated according to (F.12):

$$\sigma_{d_0} = \sigma_y\big|_{x_0} = \sqrt{\sigma_a^2 + x_0^2\,\sigma_b^2} = \frac{\sigma_{\text{meas}}}{\sqrt{N}}\sqrt{1 + \frac{12(N-1)}{(N+1)}\,r^2} = \frac{\sigma_{\text{meas}}}{\sqrt{N}}\,Z_{lin}(r,N)\,. \quad (9.82)$$

The parameter $r = x_0/L_p$ is the ratio of the extrapolation lever arm to the length over which the measurement points are distributed. The measurements should again be equally distributed in x with the centre of gravity in the origin such that the

covariance vanishes, $\sigma_{ab} = 0$ (see eq. (F.11)). Then x_0 is the distance in x from the centre of the measurements to the extrapolation point. While $\sigma_{\text{meas}}/\sqrt{N}$ represents the intrinsic position resolution, $Z_{lin}(r, N)$ characterises the geometry of the arrangement (the subscript '*lin*' indicates that the extrapolation is linear). Equation (9.82) conveys that for a given L_p the point to which one wants to extrapolate should be as close as possible to the first detector layer. In general the minimal distance is limited by the beam pipe, the radius of which, on the other side, should not be too small because of the radiation background near the beam.

Taking the curvature into account as well, the extrapolation error is given in appendix F by (F.19):

$$\begin{aligned}\sigma_{d_0} = \sigma_y\big|_{x_0} &= \sqrt{\sigma_a^2 + x_0^2\,\sigma_b^2 + \frac{1}{4}x_0^4\,\sigma_c^2 + x_0^2\,\sigma_{ac}} \\ &= \frac{\sigma_{\text{meas}}}{\sqrt{N}}\sqrt{1 + r^2\frac{12(N-1)}{(N+1)} + r^4\frac{180(N-1)^3}{(N-2)(N+1)(N+2)} + r^2\frac{30N^2}{(N-2)(N+2)}} \\ &= \frac{\sigma_{\text{meas}}}{\sqrt{N}}\,Z_{par}(r, N)\,.\end{aligned} \tag{9.83}$$

The factor Z_{par} characterises the geometry of the arrangement for parabolic extrapolation. The last two terms under the square-root additionally arise from the uncertainty of the curvature leading to a deterioration of the resolution with respect to the linear extrapolation.

Example. With the pixel detector of the ATLAS, experiment primary and secondary vertices are determined according to fig. 9.14. The specifications of the pixel detector are [3]:

$$\begin{gathered} N = 3,\ \sigma = 10\,\mu\text{m}, \\ x_1 = 5.05\,\text{cm},\ x_2 = 8.85\,\text{cm},\ x_3 = 12.25\,\text{cm} \\ \Rightarrow L_p = 7.20\,\text{cm},\ x_0 = 8.72\,\text{cm},\ r = x_0/L_p = 1.21,\ Z_{lin}(r, N) = 3.13\,. \end{gathered} \tag{9.84}$$

Note that the layers are only approximately equidistant which, however, has only a minor effect on our estimation. The numbers above yield for the impact parameter resolution in linear approximation:

$$\sigma_{d_0} = 18.1\,\mu\text{m}\,. \tag{9.85}$$

If a curvature measurement of the pixel detector alone would be used the parabolic approximation yields $Z_{par}(r, N) = 15.61$ and thus a deterioration by about a factor 5. However, usually the curvature is measured independently by a larger detector such that the d_0 resolution becomes similar to the field-free case (9.82).

Moving the detector layers closer to the interaction point or adding an additional layer at a smaller radius has a large impact on the vertex resolution. For example, ATLAS has later added a fourth layer at a radius of 3.3 cm [775]. Approximating this arrangement by assuming equidistant spacing of the four layers between 3.3 cm and 12.25 cm, the predicted impact parameter resolution σ_{d_0} (without multiple scattering) improves from 18.1 µm calculated above for three layers to 12.1 µm for four layers.

9.4.6.2 Error due to multiple scattering

The influence of multiple scattering on the impact parameter resolution was discussed in section 3.4 for two detector layers, the minimum number to determine a direction. In the general case of a detector with N layers contributing to multiple scattering we refer to the slope error derived in (9.32). As for the example in section 3.4 we assume that the multiple scattering in the beam pipe is negligible or that the first detector layer is so close to the beam pipe that both together can be treated as one scattering layer (situation as in figs. 3.33 and 8.3(b)). If we further assume that the first layer at x_1 has no measurement error (because here we only consider the contribution due to multiple scattering) the extrapolation to the primary vertex can be done from x_1 with the slope error σ_b only, with vanishing offset error σ_a. With x_{pv} being the distance from the first layer to the primary vertex, (9.32) yields the error of the impact parameter as

$$(\sigma_{d_0})_{scat} = \sigma_b\, x_{pv} = \frac{x_{pv}}{L_p}\sqrt{\sum_{i=1}^{N-1}\frac{\theta_{ms,i}^2}{\sin^2\theta}\,(x_N - x_i)^2}\,, \tag{9.86}$$

With equidistant detector layers (separated by $L_p/(N-1)$) this becomes according to (9.33):

$$(\sigma_{d_0})_{scat} = \frac{\theta_{ms,N}}{\sin\theta}\, x_{pv}\sqrt{E_N} = \frac{\theta_{ms,N}}{\sin\theta}\, x_{pv}\sqrt{\frac{N(2N-1)}{6(N-1)^2}}\,. \tag{9.87}$$

For $N = 2$ the square root factor in (9.86) becomes 1 and for $\sin\theta = 1$ we recover the result (3.107) obtained in section 3.4:

$$(\sigma_{d_0})_{scat} = \theta_{ms,2}\, x_{pv}\,. \tag{9.88}$$

With an increasing number of detectors the square root factor decreases, for example by 0.79 and 0.72 for $N = 3$ and 4, respectively, reaching for very large N the limit $1/\sqrt{3} \approx 0.577$. However, if each additional layer adds the same scattering material, $\theta_{ms,N}$ is proportional to $\sqrt{N-1}$ and the combined term $\theta_{ms,N}\sqrt{E_N}$ deteriorates with increasing N, for example by factors 1.12 and 1.25 for $N = 3$ and 4, respectively. Therefore, concerning multiple scattering a minimal number of detector layers is optimal. However, if position resolution at higher momenta becomes dominant, more layers may be required (see the corresponding remark at the end of section 9.4.2 on page 391).

Example. In the case of the ATLAS pixel detector, with specification (9.84) each of the three layers is $0.035\,X_0$ thick. The beam pipe is made of 0.8-mm beryllium corresponding to $0.0023\,X_0$ and can thus be neglected for an estimate of the influence of multiple scattering. For tracks perpendicular to the beam line (9.86) yields for the impact parameter resolution due to multiple scattering:

$$(\sigma_{d_0})_{scat} = \frac{142\,\mu\mathrm{m\,GeV}/c}{p} \quad (3\text{ layers})\,. \tag{9.89}$$

Here the dependence of the multiple scattering on the inverse momentum p is made explicit. The contribution from multiple scattering becomes smaller than that from the position resolution for momenta above about $8\,\mathrm{GeV}/c$, reasonably matching the requirements for the detection of heavy-flavour decays in jets.

Adding an innermost layer at a radius of 3.3 cm with a thickness of only $0.0155\,X_0$ yields a significant reduction of the multiple scattering contribution to the impact parameter resolution:

$$(\sigma_{d_0})_{scat} = \frac{93\,\mu\mathrm{m\,GeV}/c}{p} \quad (4 \text{ layers})\,. \tag{9.90}$$

The improvement is due to the shorter extrapolation distance to the primary vertex while the scattering in the thin additional layer deteriorates the resolution by only about 5%. Again, above about $8\,\mathrm{GeV}/c$ the multiple scattering influence becomes smaller than the effect of the position measurements.

10
Photodetectors

By the name *photodetector* we describe in this book detectors for photons with wavelengths in the UV and optical regime (about 200–700 nm). In principle, in this definition all instruments are included which—by absorbing photons—generate electrical signals or cause chemical mutations (e.g. in a film), including cameras and photocells. Also, the fact that our eye is a (chemical) photon detector was exploited in early subatomic experiments, for example by spotting scintillation light flashes on a fluorescent screen in the famous Geiger–Marsden experiment [453] carried out under the direction of Rutherford to establish the atomic nucleus. The human eye can in fact reach single photon sensitivity.

Here, however, we restrict ourselves to photodetectors which are used in detector systems for particle or astroparticle physics experiments or in similar systems for medical physics. The main applications in this context are found for scintillation light or Cherenkov light detection. Even in this limited field of applications there is a large variety of criteria to select a particular type of photodetector. Often a large number of photons enter the light-detecting device (e.g. 10 000 scintillation photons in an area of several cm^2); in other applications very few photons must be recorded over a large area in a spatially resolved way, for example individual photons on a Cherenkov ring. For the latter single photon detection capability is therefore important.

10.1 Physics of light detection

The detection of optical photons (or light) employs a large variety of detection principles, many of which are described elsewhere in this book. To a large extent photodetection today lies in the domain of two different types of 'photo-converters': (a) photocathode-based devices (described for example in [375] or [497]) and (b) semiconductor devices, most notably photodiodes. Both types appear in many variants and are also combined in so-called 'hybrid' devices. Photodetectors not classified under these categories, that nevertheless are or have been very important for particle detectors, are gas-filled chambers, in particular MWPCs (see MWPC applications in section 7.8.5 on page 213) and drift chambers with photon absorbing vapour additions like TMAE or TEA to detect Cherenkov photons (see section 11.6.1) as well as photoemulsions described in section 6.3.3.

Photodetectors generate detectable (and further processable) charge carriers (electrons or electron–hole pairs) in response to incoming light. In the case of photocathodes the electron is emitted and leaves the material, whereas in semiconductors the generated charge carriers remain inside. We discuss the light absorption and signal generation mechanisms in the following subsections.

10.1.1 Photoelectric effect at high and low photon energies

The physics process relevant for the detection of optical photons is the photoelectric effect, if one interprets the term in a broader sense than introduced for photons with energies larger than keVs in section 3.5.3. More generally, the term photoeffect comprises the absorption of photons by atoms thereby generating one (or more) charge carriers. At low (optical) photon energies, rather than being emitted, electrons are excited from a lower to a higher atomic level. In a semiconductor they are excited from the valence band to the conduction band, hence generating movable charge carriers that remain in the medium and cause a detectable signal (*internal photoeffect*). At higher photon energies an electron can be released from an atomic shell into the 'continuum'. This occurs with highest probability from the shell with the largest binding energy that is still smaller than the photon energy. The atom is ionised by the process. The emitted electron can either remain in the medium or can leave it. The latter is often called *external photoeffect* if it needs to be distinguished from the former.

Figure 10.1 shows the energy dependence of the photon attenuation cross section from eV (optical regime) to some 100 keV (X-ray regime) energies with silicon as the absorbing material. The attenuation cross section $\sigma_{att} = \sigma_{abs} + \sigma_{scat}$ contains absorption and scattering processes. The photoeffect (internal and external) represents the dominant contribution to σ_{att} from the gap energy up to several 10 keV. At the high energy end of fig. 10.1 the Compton effect is the dominant process. At 100 keV the photoeffect still contributes with 11%. In the optical regime, relevant for photodetectors, σ_{att} has contributions from the internal photoeffect and from phonon excitation. The corresponding absorption length ($\lambda_{abs} = 1/n\sigma_{abs}$) for photons follows opposite trends for X-ray and optical photons. While the absorption depth for X-ray photons decreases rapidly with decreasing energy, from 2.3 cm at 100 keV to 100 nm at 0.1 keV, optical photons at low energy (long wavelength) can deeply enter into silicon (e.g. tens to hundreds of microns for $\lambda > 750$ nm), whereas UV photons only penetrate fractions of microns (0.1 µm at 400 nm, see also figure 10.4).

10.1.2 Electron emission in photocathodes

The photoemission process in photocathodes can be described as a three-step process (Spicer model [904]):

(1) The absorption of a photon by photoeffect in a certain depth of the cathode material with kinetic energy transfer to an electron,
(2) The (successful) migration of this electron to the surface of the material (characterised by the mean escape depth).
(3) The escape of the electron from the surface of the cathode into the vacuum.

The efficiency with which an incident photon results in an emitted electron is called the *quantum efficiency*:

$$QE = \frac{\text{number of emitted photoelectrons}}{\text{number of incident photons}}. \tag{10.1}$$

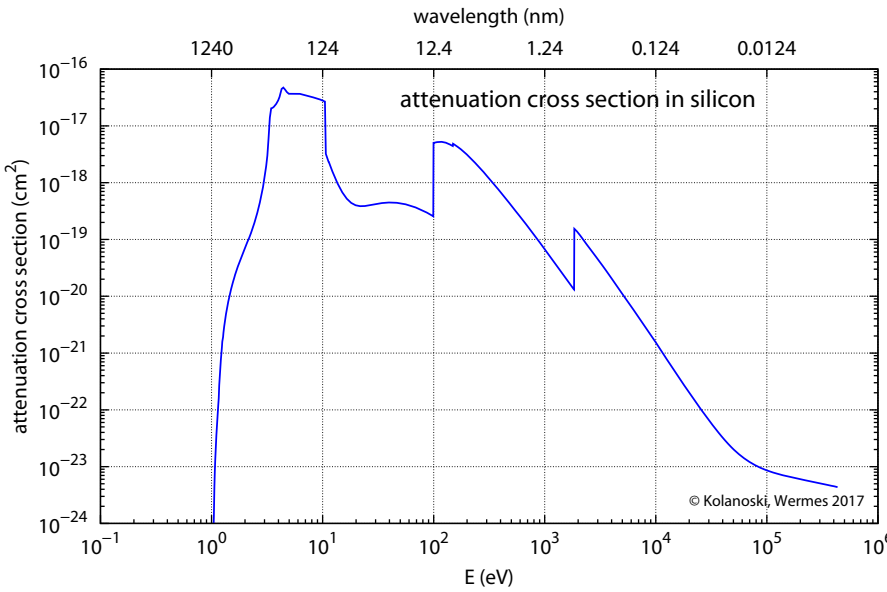


Fig. 10.1 Photon attenuation cross section σ_{att} in silicon as a function of the photon energy from about 1 eV (optical wavelength region) to 400 keV (X-rays), where the term 'attenuation' includes scattering processes besides absorption. The graph is composed of three parts from different sources: up to about 5 eV [536], 5–10.5 eV [252], and from 10.69 eV [289]. In the X-ray regime (100 eV–100 keV) the cross section falls with increasing energy by six orders of magnitude exhibiting characteristic shell absorption edges. In the optical (eV) regime σ_{att} steeply rises with energy up to about 3.4 eV. Below this energy only indirect band transitions (VB→CB) are possible in Si, which are strongly suppressed (see text in section 10.1.3).

In all three steps light intensity or energy needed to emit the electron from the cathode material can be lost. In step (1) only the absorbed part of the incident light contributes to the quantum efficiency. The reflected or transmitted fraction is lost. In step (2) photoelectrons can lose energy by electron–electron or electron–lattice (phonon) collisions. In step (3), finally, to escape the surface a potential barrier must be passed (see fig. 10.2(a)). Such a barrier $e\Phi_A$ always exists at an interface between a material and the vacuum. To pass it, the photoelectron at the surface must have kinetic energy:

$$E^e_{kin} = h\nu - e\Phi_A - E_{loss} \geq 0\,, \tag{10.2}$$

with $h\nu = hc/\lambda$ being the incident photon energy and E_{loss} the energy lost on the electron's path to the surface.

The quantum efficiency can thus be expressed as a product of individual probabilities

$$QE(\lambda) = (1-\eta_R)\, P_\lambda \left(\frac{1}{1+\frac{1}{\mu d}}\right) P_{esc}\,, \tag{10.3}$$

where η_R is the reflection probability entering the cathode, μ the absorption coefficient, and d the mean escape depth of the material. P_λ is the probability that a photon with wavelength λ excites an electron to a level larger than the vacuum level (see below) and P_{esc} is the probability that electrons reaching the surface are released into the vacuum.

The term in parentheses arises from the probability that the photon is absorbed within a certain depth x multiplied with the probability that the created electron reaches the surface. The latter is determined by the *mean escape depth* d and the assumption that the probability for the electron to be lost is proportional to its travel distance. Integrating over all depths x yields:

$$\int_0^\infty \mu\, \mathrm{e}^{-\mu x}\, \mathrm{e}^{-x/d}\, dx = \frac{1}{1+\frac{1}{\mu d}}\,. \tag{10.4}$$

Note that the product of absorption and escape depth enters here and in (10.3).

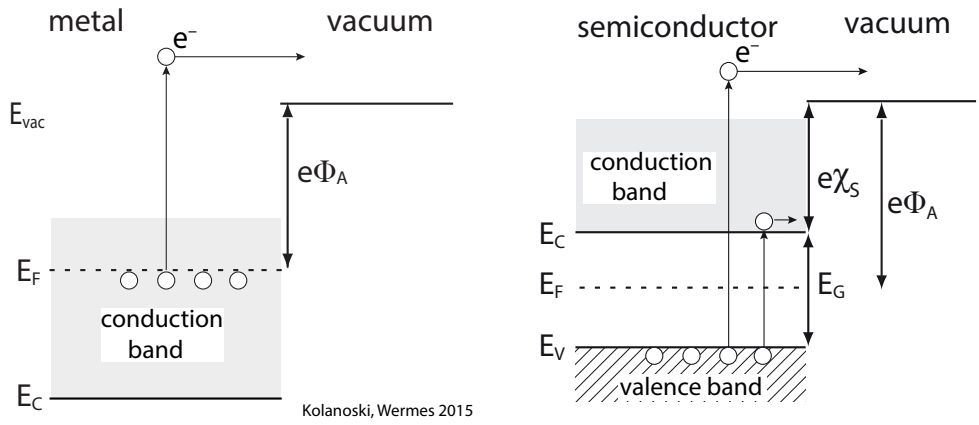
The situation in metals is sketched by the simplified band diagram in fig. 10.2(a) assuming zero temperature. At $T = 0$, the Fermi level E_F characterises the energy up to which all energy levels in the conduction band are filled with electrons, thus being the highest energy that electrons can have. At $T > 0$ the population probability is smeared about E_F according to the Fermi–Dirac distribution function (see also section 8.2.3 on page 267). The energy difference $e\Phi_A = E_{vac} - E_F$ is defined as the *work function*.

The energy losses in the three steps are very material dependent. In metals a large fraction of the incident light is reflected and due to the large density of free electrons also the electron–electron collision losses are high, resulting in short (relative to $1/\mu = \lambda$) photoelectron escape depths d to reach the surface (typically only a few nanometres for d compared to micrometres for $1/\mu$). The corresponding probability term (10.4) is small. Also the work function is relatively large for metals ($e\Phi_A \gtrsim 3$ eV, violet), making them sensitive only to ultraviolet photons, with the exception of alkali metals with $e\Phi_A$ reaching down to about 2 eV (caesium, orange). For visible light the resulting quantum efficiencies are very small (typically $< 0.1\%$).

Therefore photocathodes are usually made of compound semiconductors as photoemitters (most often one of the compound partners is an alkali metal), as they are more efficient in all of the three process steps mentioned at the beginning of this subsection. A simplified semiconductor band diagram is shown in fig. 10.2(b) for comparison. Some common photocathode materials are described in section 10.2.1.

The semiconductor band diagram is described in detail in chapter 8. In addition to the Fermi energy E_F as a characterising quantity, which for semiconductors usually lies inside the band gap, the electron affinity $e\chi_s$, the energy distance of the conduction band edge to the vacuum,[1] is the relevant quantity to characterise the work needed to remove an electron from the lattice binding into the vacuum (see section 8.3.4). For photoemission to take place the photoelectron (excited from the valence band) must overcome the band gap energy E_G plus the electron affinity $e\chi_s$. In terms of the three-step process semiconductors are much superior to metals as photocathodes. In step (1) light absorption is effective already at much lower energies than in metals. The losses in step (2) are reduced to essentially only phonon scattering losses because the electron density in the conduction band is low compared to metals (no electron collision losses) and thus result in a much larger escape depth (typically 25 nm) as well as absorption to escape depth ratio (μd). Finally, for step (3), the energy $E_G + e\chi_s$ to be surmounted for emission from the photocathode can be synthesised with tailored semiconductor

[1] In solid state physics the vacuum level is often defined near the surface, acting as a reference for energy level changes along material boundaries. It is typically not a constant depending also on energy contributions from electric fields generated by charges at the surface and/or in the solid's bulk.



(a) Electron emission from a metal. (b) Electron emission from a semiconductor.

Fig. 10.2 Simplified band model diagrams of photoemission: (a) metal–vacuum boundary, (b) semiconductor–vacuum boundary. E_G, E_F, E_C, E_V, E_{vac} are energies of band gap, Fermi level, conduction band lower edge, valance band upper edge, and vacuum level, respectively. $e\Phi_A$ is the work function, representing in metals the energy barrier for electron emission; $e\chi_S$ is the electron affinity that electrons must surmount in semiconductors. Photoelectrons must reach energy levels higher than E_{vac} to be emitted from the absorber material into the vacuum.

materials to below 2 eV, that is, lower than the work functions of metals. Still, to escape, an electron excited to some energy beyond the conduction band lower edge must diffuse to the surface of the photocathode keeping sufficient energy to remain above the work function barrier. Phonon scattering, however, will bring the electron down in energy to the bottom of the conduction band on a time-scale of the order below picoseconds, after which escape is no longer possible.

By a special surface treatment of the photocathode the electron affinity can still be lowered, even assuming negative values (hence being referred to as 'negative electron affinity' [859, 903]) and causing a much lower energy threshold for photoemission, as explained in fig. 10.3. If the semiconductor (e.g. GaAs, GaP or Na_2KSb) is heavily p-doped, for example GaP by Zn atoms ($\simeq 10^{19}\,\mathrm{cm}^{-3}$), the Fermi level E_F is fixed close to or even below the acceptor level. By depositing a very thin film (a few atom layers) of an electropositive material (e.g. Cs) on the surface, then denoted for example GaP(Cs), a special situation for the energy levels at the surface is created [538, 974] (fig. 10.3). To a depth of about 10 nm into the cathode material, Cs electrons create surface states by filling the acceptor levels and leaving behind positive Cs ions at the boundary. As a result all energy levels bend downwards while the Fermi level remains constant (see also section 8.3); near the boundary E_A is below the Fermi level. In addition, the strong dipole field, created by the electrostatic attraction of the positive Cs ions and the (negatively) filled acceptors (Zn), is accompanied by Cs atoms that do not lose electrons to the bulk, but rather become polarised, hence forming another dipole layer (double dipole). This leads to a change in the energy balance that effectively lowers the vacuum level E_{vac} near the surface boundary. Co-deposition of oxygen enhances the effect (see [974] and references therein).

With E_{vac} now being lower than the bottom of the conduction band in the bulk,

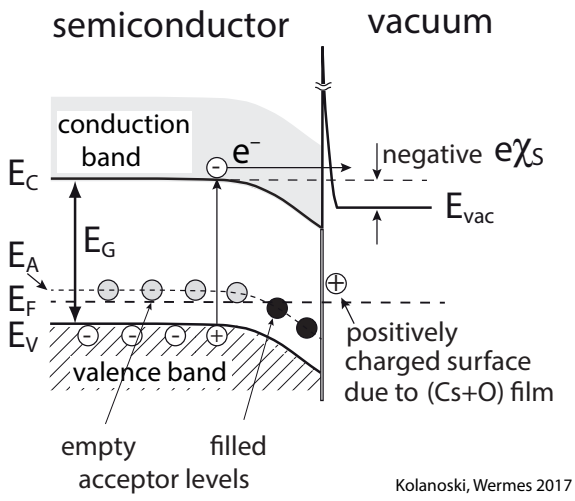


Fig. 10.3 Negative electron affinity obtained by a thin electropositive film (Cs+O) bending the energy bands down by the potential caused by positive surface charges and negative filled acceptor states (see also text).

escaping of conduction band electrons in the bulk by tunnelling through a thin residual surface barrier is energetically enabled. Compared to 'normal' (positive) electron affinity, where transitions are only possible for electrons energetically well above the bottom of the conduction band, two orders of magnitude in time for tunnelling are gained, because recombination into the valence band occurs on a time-scale of $\mathcal{O}$(100 ps) whereas relaxation within the conduction band is of the order of picoseconds. The average escape depth for materials treated this way can reach 10–100 nm. Quantum efficiencies of so-designed semiconductor photocathodes reach maximum values around 400 nm of 25–30% and can generally be extended to red and infrared wavelengths (see also figs. 10.6 and 10.10).

10.1.3 Direct photon detection in semiconductors

Photoemission is not required in order to obtain an electrical signal, for example in a silicon photodiode. Instead, the signal is produced by an electron–hole pair (e/h) created by the (internal) photoeffect and separated inside the diode by an electric field (see also chapter 8). The movement of the charge carriers by drift in a depleted, carrier-free region causes an electrical signal induced at the readout electrodes.

The minimal photon energy needed to generate an e/h pair is the band gap energy E_G (1.12 eV in Si). The corresponding maximum wavelength for absorption in silicon is:

$$\lambda_{max} = \frac{hc}{E_G} \approx 1100\,\text{nm}\,. \tag{10.5}$$

The absorption cross section therefore depends on the probability for a band gap transition, which in turn depends on the details of the band configuration and the semiconductor lattice.

Because silicon is an indirect semiconductor (see section 8.2.2 on page 264) the energy distance between valence and conduction band at the same crystal momentum is much larger than the minimum gap energy of 1.12 eV. A direct transition without additional momentum transfer therefore requires an energy of about 3.4 eV (fig. 8.8 on page 265). This explains the sharp change in the absorption cross section in fig. 10.1

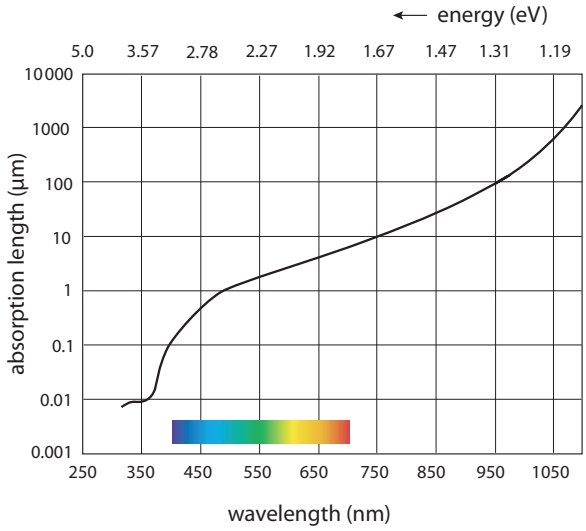


Fig. 10.4 Absorption depth of photons with optical wave lengths in silicon. In the blue wavelength region (470 nm) the absorption depth is about 0.6 μm, in the red wavelength region (625 nm) about 2.9 μm.

at this energy, because for all lower energies indirect transitions govern the absorption process.

For indirect transitions to happen, the minimum energy of 1.12 eV plus an additional momentum change of order h/a ($a =$ lattice constant) is needed, as described in section 8.2.3. This momentum change cannot be provided by low energy optical photons because the photon momentum h/λ approaches h/a only at energies of about 100 keV. Phonons, as the quanta of lattice vibrations, however, occurring with 'acoustical' (~ 4 THz) to 'optical' (~ 15 THz) frequencies[2] at comparatively slow velocities, can more easily provide large momentum transfers for a given energy $\hbar\omega$ [927] (the dispersion relation yields $\hbar k \approx \hbar\omega/v_{ph}$ for $\lambda \gg a$, v_{ph} = phase velocity).

The phonon energy distribution follows the Bose–Einstein distribution function $(\exp(E_p/kT) - 1)^{-1}$ with E_p being the phonon energy, T the temperature and k the Boltzmann constant. This in turn determines the probability that the necessary momentum transfer to enable indirect band-gap transitions can be provided at a given temperature and explains the steep rise from 1.12 eV to 3.4 eV (direct transition) in fig. 10.1.

Figure 10.4 shows the absorption length as a function of wavelength for the optical wavelength region, corresponding to the region below 5 eV in fig. 10.1. At 470 nm wavelength the penetration depth is approximately 0.6 μm, whereas for 1050 nm it is almost 1 mm, that is, larger than typical thicknesses of silicon detectors (200–500 μm). Above $\lambda \approx 1100$ nm silicon is transparent to electromagnetic radiation (fig. 10.4).

10.2 Systems with photocathode and electron amplification

10.2.1 Photomultiplier

The strong motivation from many fields to detect low light intensities down to single (optical) photons produced the *photomultiplier tube* (PMT). It has been successfully used for decades in many applications and still continues to be constantly improved.

[2]The terms 'acoustical' and 'optical' phonons correspond to different transverse vibration modes (in phase and out of phase) of diatomic chains in lattices [120, 627].

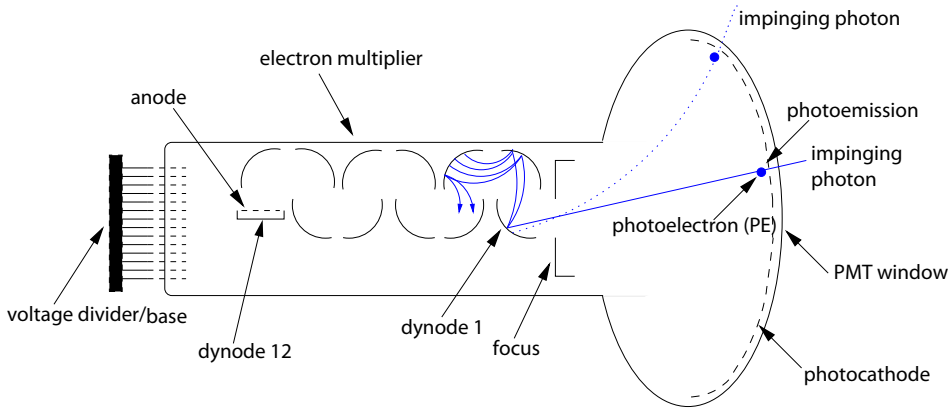


Fig. 10.5 Function principle of a photomultiplier. Voltages for the electrodes (cathode, anode, dynodes) are supplied via a voltage divider chain, the 'base'. Outputs are available from the anode and in addition from some dynodes (adapted from [865]).

PMTs are still the best instrument for many low intensity applications due to their low noise performance and easy handling.

Figure 10.5 displays the PMT functional principle. It consists of a photocathode, an amplification system containing the so-called *dynodes*, and an anode from where the signal is extracted. Usually the signal can also be extracted from one or more dynodes at the end of the amplification chain. The dynodes are also coated with a photosensitive layer, reducing the work function.

Light entering through the entrance window of the vacuum tube hits the photocathode material which is vapour-deposited at the inside of the tube. Photoelectrons are emitted via photoeffect (section 10.1.2) into the tube vacuum (photoemission) and are focused onto the first dynode where they are amplified by the emission of secondary electrons. This process is repeated at all subsequent dynodes and the sum of secondary electrons is finally collected at the anode. This way photoelectrons are amplified by up to a factor of about 10^9 (typical rather is 10^5–10^7). The output can then be processed further as an electrical signal (current pulse).

10.2.1.1 Photocathode

For the reasons explained above, commonly used photocathode materials are monoalkali (e.g. Cs–I, Cs–Te, K–Br, Sb–Cs[3]), bialkali (e.g. Sb–Rb–Cs[3]) or multialkali materials, (e.g. Sb–Na–K–Cs[3]), but also III–V semiconductors, 'activated' (doped) with caesium, like GaAs(Cs), GaP(Cs), or InGaAs(Cs).

Monoalkali cathodes are usually sensitive to vacuum-ultraviolet and ultraviolet wavelengths only and are therefore also called 'solar blind'. Bialkali compounds show—compared to monoalkali—high quantum efficiency for photoemission from ultraviolet wavelengths up to and beyond $\gtrsim 650$ nm (see fig. 10.6). The spectral range of multialkali photocathodes even reaches into the infrared, at the expense of higher noise rates however, because sensitivity to longer wavelengths corresponds to smaller energy

[3]The chemical formulae are Cs_3Sb, Rb_2CsSb, and $Na_2KSb(Cs)$, respectively.

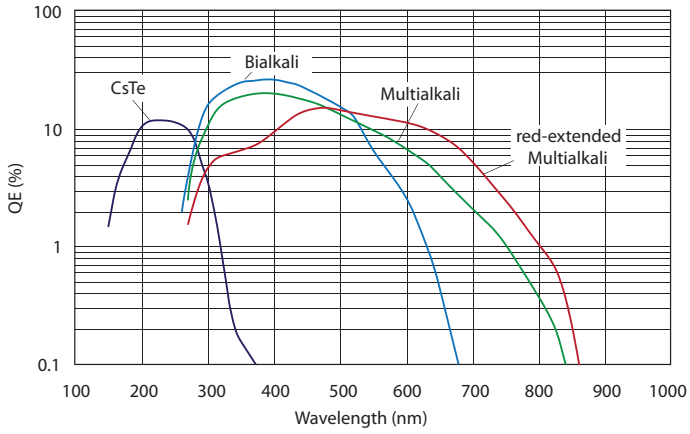


Fig. 10.6 Quantum efficiencies for different typical monoalkali, bialkali and multialkali photocathode materials as a function of wavelength (adapted from [558]).

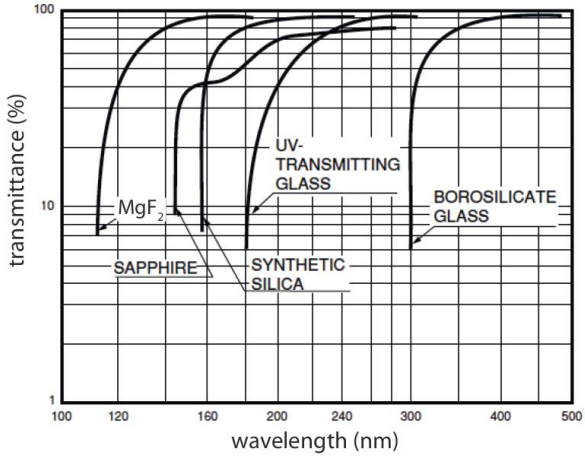


Fig. 10.7 Transmittance of typical PMT entrance windows: MgF_2, sapphire, quartz (synthetic silica), UV borosilicate, and standard borosilcate glass. Source: Hamamatsu [497].

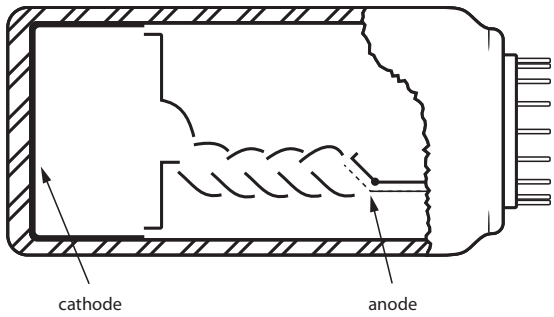
level gaps employed used for the emission, such that also thermal electron emission becomes more likely.

10.2.1.2 Tube glasses

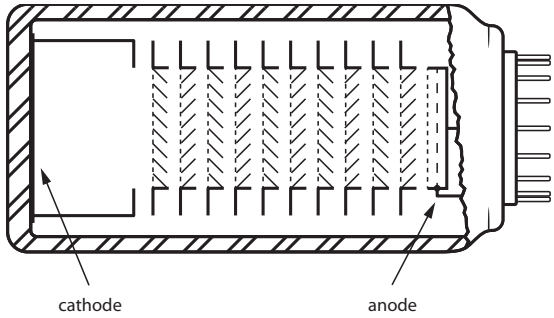
For the light entrance window of the tube the most commonly used material is borosilicate glass, whose transmission however cuts off at 300 nm. Borosilicate is easy to produce and form and is also cheap, the main reasons for its common usage. Figure 10.7 illustrates the transparency curves of some PMT glasses. To extend the range to shorter wavelengths glasses more transparent for UV light must be used, for example MgF_2 (wavelength cut-off at about 115 nm), sapphire (140 nm), synthetic silica (fused quartz, $\sim$150 nm), or UV-borosilicate ($\sim$185 nm) [497].

10.2.1.3 Electron multiplier

The electron multiplier, that is, the dynode system plus the anode, is an amplifier characteristic for its fast and high amplification at low noise. Figure 10.8(a) shows a commonly used type of electron multiplier with a dynode geometry optimised for maximum gain and best linearity. In fig. 10.8(b) a multiplier with slit-like dynodes



(a) Linearly focusing dynodes.



(b) Photomultiplier with slit-like dynodes.

Fig. 10.8 Different electron multiplier types. Source: EMI Electronics Ltd [373].

(*venetian blinds*) is shown which easily adapts to the tube form with a head-on electric field configuration allowing large photocathodes and featuring small 'after-pulsing'. The electron multiplier is very sensitive to magnetic fields which affect the dynode amplification, especially in the first stage with a large distance from the photocathode to the first dynode. To shield against not too large magnetic fields one can cover the PMT by a μ-metal housing (see page 503 in section 13.1).

10.2.1.4 Voltage divider (base)

To focus and accelerate the electrons in the electron multiplier, increasingly higher potentials are applied on the cathode, the dynodes, and the anode. In practice these voltages are derived from the operation voltage V_A by means of a voltage divider, called the 'base'. A typical base, using a commonly used resistor chain to subdivide the applied voltage, is shown in fig. 10.9. In 'active' dividers the dynode potentials are fixed by transistors controlling the current flow. They are smaller in size and can be used for pulsed operation with not too high currents.

For applications with pulsed light sources the choice of the total resistance of the voltage divider $R_B = \sum R_i$ has to consider the average pulse current at the anode $\langle I_A \rangle$. On the one hand, the current I_B must remain small in order not to represent a too large load; on the other hand, it must be large in relation to $\langle I_A \rangle$ in order to keep the potentials at the dynodes stable (e.g. $I_B \gtrsim 100 \langle I_A \rangle$). Problematic are very bright light pulses for which the last dynodes need currents larger than I_B. For this reason the potentials at these dynodes are short-time supported by currents from additional buffer capacitors. The choice of the capacitance depends on the maximum and the duration of the current peaks and on the question how large a loss in PMT gain can

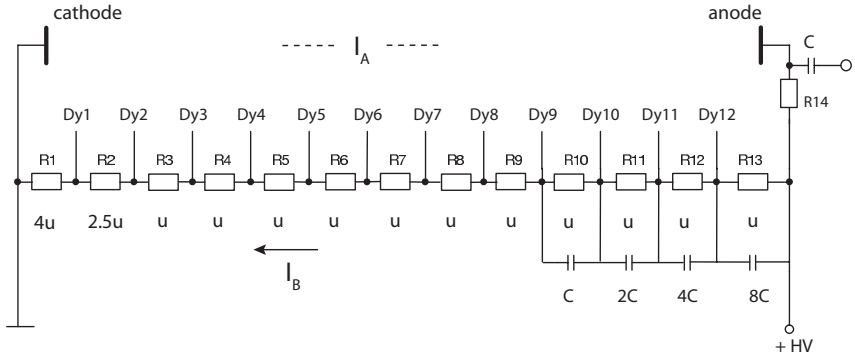


Fig. 10.9 A typical 'base' of an electron multiplier generating a potential drop at the dynodes. By means of the resistors the high voltage (typical 1.8 kV) is subdivided into stepwise increasing potentials to be applied at the different dynodes. I_A denotes the tube current and I_B the current through the resistor chain. Typical voltage steps in units of 'u' volts are indicated below the resistors. By means of the capacitors applied over the last dynodes the applied voltage is buffered when high currents (from bright signals) pass the dynode system.

be tolerated. Assuming a k-fold amplification per dynode stage (e.g. $k = 2$) and the safety factor 100 from above, the capacitances should typically be dimensioned as $C_i \approx 100/k^{(n+1)-i}\, Q_A/V_i$, where n is the number of dynodes, Q_A the maximum pulse charge at the anode and V_i the voltage over the respective capacitor [865].

Usually PMTs are operated with negative high voltage (HV) on the cathode relative to the anode (being on ground potential) in order to avoid potential differences between the anode and the external electronics. For large PMTs in particular, however, the opposite arrangement is often chosen to avoid the cathode being on a high potential with respect to the grounded environment (housing, μ-metal cage, see page 503). Otherwise electrons could become deflected from the cathode to the housing, which may lead to scintillations in the PMT glass. This not only increases the noise rate but also damages the PMT in the long run. It is then preferred to put the cathode to ground potential and the anode on positive HV, as is shown in fig. 10.9. To avoid also putting the readout electronics on HV the output signals must then be extracted via AC coupling.

10.2.1.5 PMT efficiency

The total PMT efficiency is determined by the product of three contributions (probabilities): (1) the transmission probability (transmittance) through the PMT tube glass; (2) the photocathode quantum efficiency; (3) the electron collection efficiency at the (effective) active surface of the first dynode. The product is dominated by (2) for wavelengths above the glass transmission cut-off where typical transmittance values exceed 90% for all used glasses (see fig. 10.7). Furthermore, the electron collection efficiency (3) can be made to approach values near 100% provided that the voltage between cathode and first dynode is more than about 100 V [497] and if magnetic field effects can be eliminated. Losses in later stages of the dynode chain can be neglected. At wavelengths lower than 400 nm special PMT entrance windows with larger UV trans-

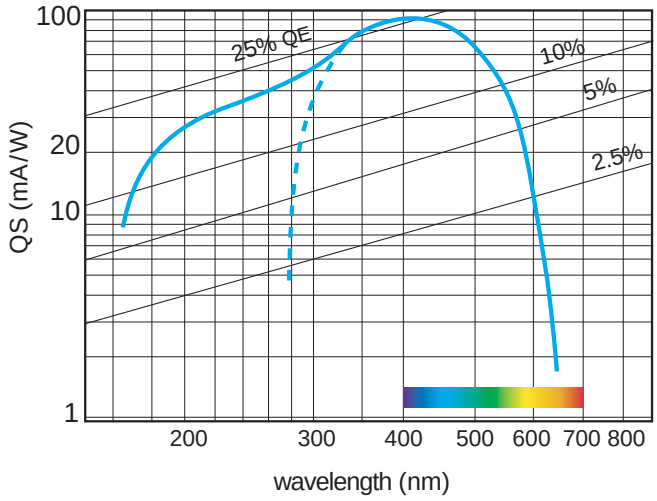


Fig. 10.10 Typical radiant sensitivity QS (mA/W) of a photomultiplier with (solid) and without (dashed) UV-transparent entrance window. Also shown are lines of the quantum efficiency QE, measured in %.

mittance can be chosen, combined with UV sensitive photocathodes (e.g. bialkali), as explained in sections 10.1.2 and 10.2.1.

The overall PMT spectral response specifying the signal efficiency per incident photon as a function of wavelength is usually also simply termed *quantum efficiency* QE, defined analogous to (10.1) (with the number of emitted photoelectrons replaced by the number of output signals) measured in %. An alternative measure is the *radiant sensitivity* QS measured in A/W. The incoming total photon energy $E = N_{ph} \times h\nu$ (unit W s) of N_{ph} photons with energies $h\nu$ is transformed into N_{pe} photoelectrons with total charge $N_{pe} \times e$ (unit As):

$$QS = \frac{N_{pe} \times e}{N_{ph} \times h\nu} \ . \tag{10.6}$$

Quantum efficiency as well as radiant sensitivity are wavelength dependent. Figure 10.10 shows the spectral dependence of both quantities for a typical photomultiplier [375, 497]. The spectral range is limited at large wavelengths by the cathode material and at small wavelengths by the transmittance of the entrance window.

Photomultipliers have typical quantum efficiency maxima at some wavelength around 25% up to at most about 40% at small noise rates (see e.g. [945] and [496]). The QE maximum is usually found at wavelengths between 300 nm and 600 nm. The spectral sensitivity range of a PMT must match the photon input spectrum, as shown in fig. 10.11. The light emission of most scintillator crystals peaks between 400 nm and 550 nm; plastic scintillator between 390 nm and 450 nm. The examples shown in fig. 10.11 are for CsI(Tl) and CsI(Na); see also tables 13.1 and 13.3 in section 13.3.2. A few scintillation crystals, however, emit in the UV range, like BaF_2 which is important for fast scintillation measurements.

It is a great technical challenge to fabricate large photomultipliers with homogeneous response characteristics. The so far the largest PMTs (fig. 10.12) have been produced for the experiment Super-Kamiokande [432], which has discovered neutrino oscillations (see section 16.6.1 on page 694).

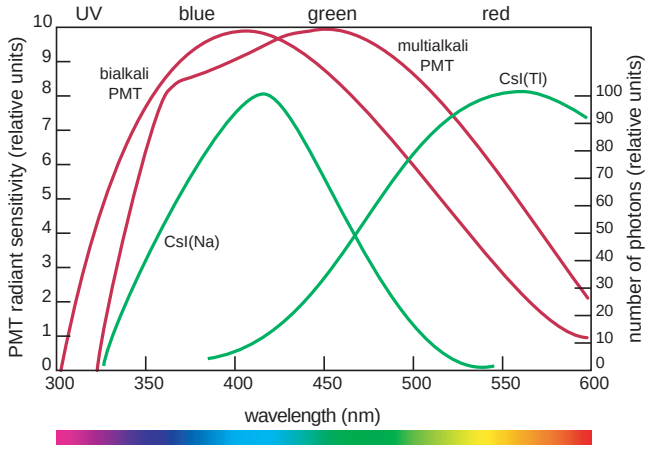


Fig. 10.11 Photoemission spectra of CsI(Tl) and CsI(Na) scintillator crystals (green lines, right-hand y-axis) in comparison to the spectral sensitivities of two types of PMT with bialkali, resp. multialkali photocathodes (red lines, left hand y-axis); adapted from [938].

Fig. 10.12 Large photomultiplier tube (Hamamatsu R3600) with a cathode diameter of 51 cm for the Super-Kamiokande experiment. With kind permission of Hamamatsu Inc.

10.2.1.6 PMT characteristics summary

In conclusion, the following characteristic properties of photomultipliers can be pointed out:

- high gain ($\gtrsim 10^6$),
- very low noise,
- single photon detection capability,
- small quantum efficiency by comparison (typically about 25%),
- good time resolution in the order of $\lesssim 200$ ps,
- commercially available in many variants for a range of application scenarios,

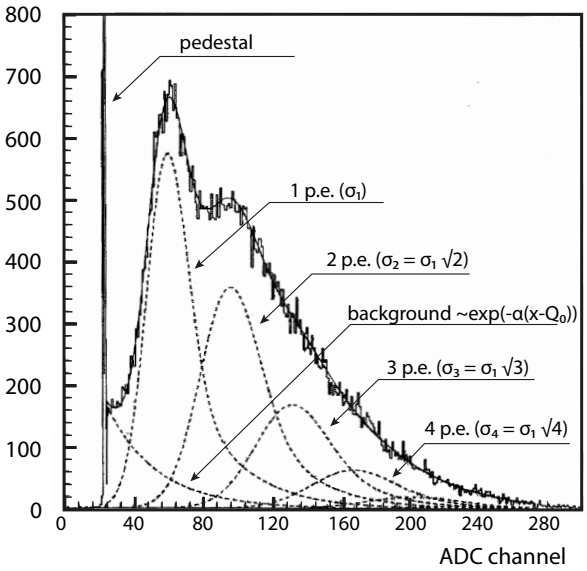


Fig. 10.13 Typical PMT spectrum with single photon sensitivity showing the pedestal and contributions from 1–4 photoelectron peaks fitted by a sum of Gaussians plus a background contribution. Adapted from [174] with kind permission of Elsevier.

– difficult to handle regarding size and shape in comparison to alternatives (e.g. photodiodes).

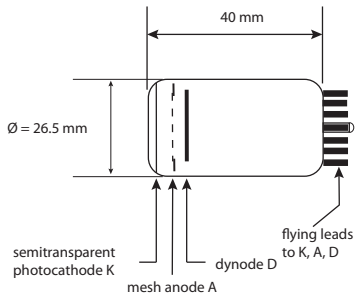
Photomultipliers are particularly interesting if the light intensity to be detected is very faint or if even individual photons are to be detected (see also section 10.6). A typical PMT spectrum with individual photon sensitivity is shown in fig. 10.13.

10.2.2 Vacuum phototriode

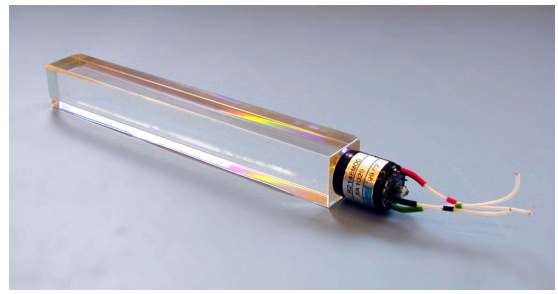
Vacuum phototriodes (VPT) are used when the usage of PMTs with their dynode system is not possible in strong magnetic fields. VPTs are single stage photomultipliers consisting of a photocathode behind the tube glass (often UV transparent) and a fine-meshed anode grid [173] sitting in front of a dynode (fig. 10.14(a)). A certain fraction of photoelectrons (typically 50%) reach through the anode grid and create secondary electrons at the dynode (typically $m \approx 10$–20 per photoelectron) which are largely collected at the anode. A fraction, however, passes through the anode's mesh, is stopped and brought back by the electric field producing tertiary electrons in the dynode, some of which are collected. The tertiary emission factor is much smaller than m (typically 0.5 per photoelectron [172]) and further emissions are negligible [172].

Due to the smaller total multiplication factor of VPTs compared to PMTs, the former are not suited for single photon detection (as would be needed for example to measure Cherenkov rings) but rather for reading out higher light intensities, for example calorimeter crystals for the detection of high energy photons ($\gtrsim$ GeV) that generate sufficient light in the crystals (e.g. $\gtrsim$100 photons per MeV), in particular when operated in magnetic fields which forbid the use of PMTs. The overall quantum efficiency of VPTs typically is about 20–22% (see e.g. [173]) at wavelengths around 400–430 nm.

The endcap regions of the electromagnetic calorimeters of the experiments OPAL at LEP (lead glass) [63] and CMS at LHC ($PbWO_4$) [173, 298] being influenced by



(a) VPT schematic drawing.



(b) $PbWO_4$ crystal with VPT readout.

Fig. 10.14 Vacuum phototriode: (a) schematic drawing showing photocathode, meshed anode, and dynode (adapted from [173]), (b) $PbWO_4$ scintillation crystal with VPT readout (source: CMS/CERN [315]).

the magnetic fields of the tracking detectors are therefore read out by VPTs. Figure 10.14(b) shows a crystal used in the endcap calorimeter of CMS [315].

10.2.3 Microchannel plate

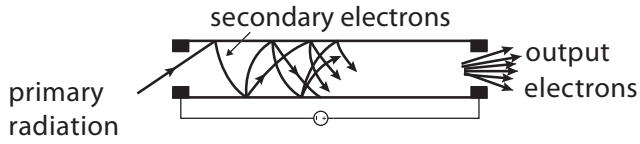
The microchannel plate (MCP, fig. 10.15) [1004] is a photomultiplier with additional spatial sensitivity, usually operating with somewhat smaller multiplication gains (10^4–10^7) than PMTs. The MCP consists of 0.4–2 mm thick glass plates containing microchannels of 10 µm diameter separated by pitches of micrometres. The channels are coated with photocathode material. A typical two-stage arrangement is given in fig. 10.15(c). The microchannels (fig. 10.15(a)) allow spatially resolved two-dimensional light detection and act as continuous dynodes when high voltage is applied. Often they are arranged at an angle to the incoming light direction, as shown in fig. 10.15(c), to avoid direct passage of photoelectrons.

Microchannel plates have been developed for sensitive light detection (also single photons) and possess quantum efficiencies of up to 25% with spatial resolutions in the millimetre range. Due to the short transition time of the photoelectrons and the small resulting time jitter, MCPs feature excellent time resolution, for one-stage devices in the range of 20 ps [405, 746].

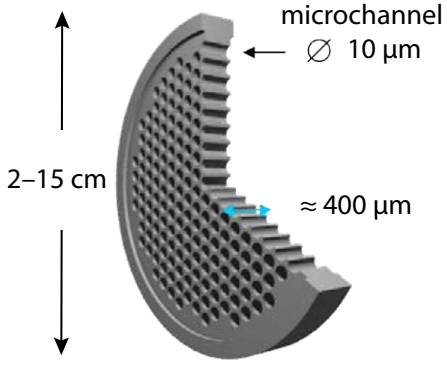
MCPs also find applications as image intensifiers stepping up weakly illuminated exposures in optical systems (e.g. in devices for night vision). In particle detectors MCPs are found as alternatives to PMTs when space and/or very good time resolutions are required, for example for the readout of Cherenkov photons from quartz bars of a DIRC Cherenkov detector [406] (see chapter 11). MCPs are also much less sensitive to magnetic fields than PMTs, which renders them attractive for such applications. A disadvantage is the relatively long recovery time per channel in the range 10–20 ms, which is due to the large effective RC time constant imposed by the channel geometry and resistivity in the recharging process [1004], as well as the much shorter lifetime in comparison to other photodetectors.

10.3 Semiconductor-based photodetectors

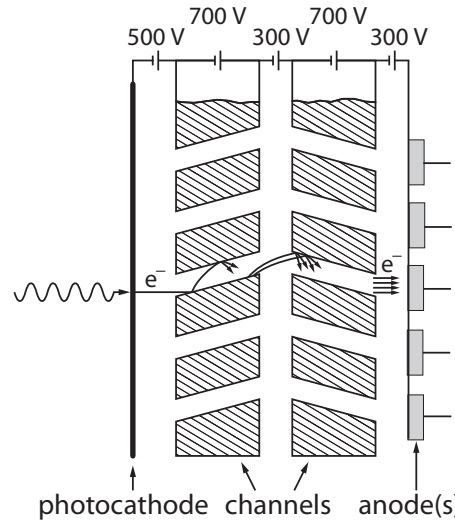
As in the field of electronics, semiconductors have revolutionised the field of light generation and detection since the 1950–1960s. The electron tube technique as in PMTs, used almost exclusively until then for light detection, received strong competition



(a) MCP principle.



(b) MPC cutout.



(c) Chevron-channel structure.

Fig. 10.15 Microchannel plate photomultiplier (MCP). (Figure (c) adapted from [351] with kind permission, © 1977 IEEE).

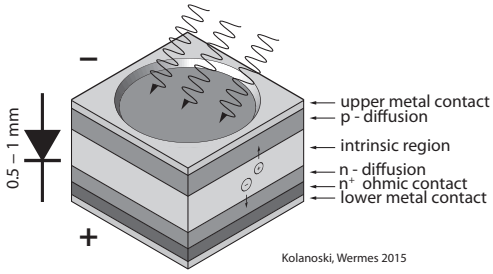
from semiconductor photodiodes. Besides the different detection characteristics of direct semiconductor photodetectors the smaller geometrical dimensions and the price also play an important role for many applications. In what follows we restrict ourselves to silicon as the base semiconductor material for photodetectors.

10.3.1 Photodiode

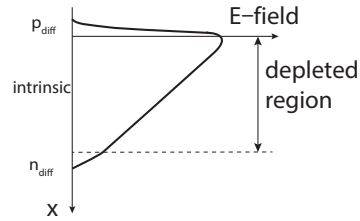
Concept-wise the photodiode is the simplest semiconductor-based photodetector. It transforms light into electric current by the (internal) photoeffect (see section 10.1.1). Very common is the so-called PIN photodiode whose form and function are very similar to those of large area semiconductor detectors described in chapter 8. Figure 10.16(a) shows its principle composition. Between a p- and an n-doped region there is an undoped zone whose free charge carrier density is determined by the impurity density rather than by controlled doping. This zone is therefore called 'intrinsic' zone and refers to the 'I' in PIN.

When applying a reverse bias voltage the quasi carrier-free depletion zone extends deeply into this intrinsic zone so that light quanta absorbed in the zone create e/h pairs that induce current signals on the electrodes when separated by the electric field in the diode (fig. 10.16(b)). Figures 10.16(c) and 10.16(d) show a photograph of a large area photodiode example and its equivalent circuit, respectively.

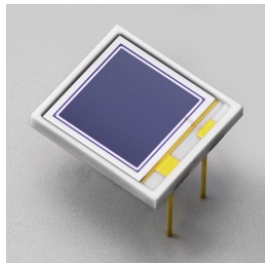
Compared to PMTs silicon photodiodes have a much larger quantum efficiency over a large wavelength range from 190 nm for special UV-sensitive diodes up to 1100 nm.



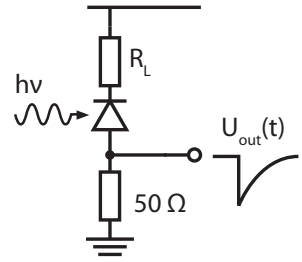
(a) Principle.



(b) Electric field.



(c) Photograph.



(d) Operation circuit.

Fig. 10.16 PIN photodiode: (a) principle composition (thicknesses of layers not to scale); (b) electric field; (c) photograph of a large area photodiode (Hamamatsu Inc. with kind permission); (d) operation circuit.

Quantum efficiencies of more than 70% are reached, whereas typical QEs for PMTs are around 25%. The QE coverage of the large wavelength range is shown in fig. 10.17 where the efficiency of a PMT and the emission spectrum of a CsI (Tl) scintillation crystal are also shown for comparison.

Besides price and compactness the advantages compared to PMTs include the low bias voltage needed for operation and the fact that photodiodes are largely insensitive to external magnetic fields. Disadvantages are the smaller sensitivity at low light intensities because photodiodes have no intrinsic amplification, a worse time resolution (∼2 ns compared to ∼200 ps for PMTs) as well as a comparatively small sensitive area (see also section 10.6).

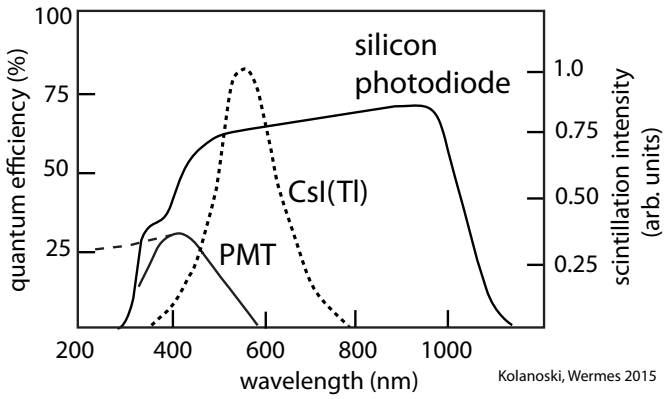


Fig. 10.17 Typical quantum efficiencies (left y-axis) of photodiodes in comparison to photomultipliers (PMT) (solid lines, dashed line for UV-sensitive photocathode). For comparison, a typical emission spectrum of a CsI (Tl) scintillation crystal (dotted line and right-hand y-axis).

Different to PMTs the output signals of photodiodes need to be amplified to achieve sufficiently large signals. The noise characteristics of photodiode systems are usually dominated by the amplification electronics, where the diode's capacitance and its dark current are responsible for the dominant noise contributions (see sections 13.4.2 and 17.10.3.3) and exclude single photon sensitivity.

10.3.2 Avalanche photodiode

Avalanche photodiodes (APDs) are photodiodes possessing an additional buried pn-junction called a *metallurgical junction*[4] with high p and n doping as shown in fig. 10.18. The high doping causes a high electric field at the junction (see also section 8.3.1) whose magnitude is proportional to the doping concentration. This way an intrinsic amplification is obtained. Different to the multistage amplification in the PMT's dynode system, however, in an APD avalanche amplification is obtained in the high field region of the metallurgical junction—similar to the gas amplification in gas-filled detectors (chapter 7). The amplification gain of commercially available APDs is in the order of between ten and several hundreds. APDs especially developed for high gain reach gains larger than 1000.

The APD is sensitive to photon absorption in the thicker, more weakly doped and depleted n region. Generated signal electrons drift towards the readout electrode, thereby passing the high-field region (typically about 10 kV/cm) in which avalanche amplification takes place. The amplified charge induces a signal on the readout electrode which usually is further amplified electronically. Most APDs are so-called 'reach-through' structures in which light reaches from the n^+ side (positive side) through the amplification zone to the absorption region. Such structures are light sensitive over a wavelength range from typically 450 nm to 1000 nm (absorption depths from 1 µm to 1 mm, cf. fig. 10.4). Structures with sensitivity to shorter wavelengths down to the UV range ($\lambda \lesssim 100$ nm) feature light entrance on the p side (smaller entrance depth needed) and are typically sensitive between 200 nm and 800 nm. APDs with diameters from 100 µm to about 1.5 cm are commercially available.

Since APDs—contrary to photodiodes—deliver an already amplified signal to the subsequent electronics the signal-to-noise ratio (SNR) is usually determined by the APD itself and not by the electronics amplifier. In addition to the thermally generated leakage current contribution I_0 which generates shot noise (see section 17.10.3.1) and

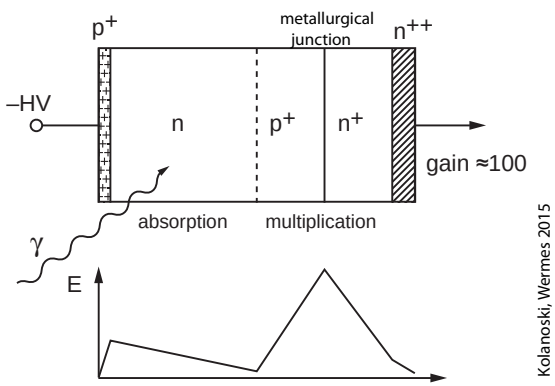


Fig. 10.18 Doping regions and field relations in an avalanche photodiode (schematic). Photons are dominantly absorbed in the larger low-field region. Avalanches are formed in the high-field section of the 'metallurgical junction'. The high-field region is drawn larger compared to the low-field region than in reality.

[4]Historical definition of a pn junction with equally high p and n doping ($N_A = N_D$).

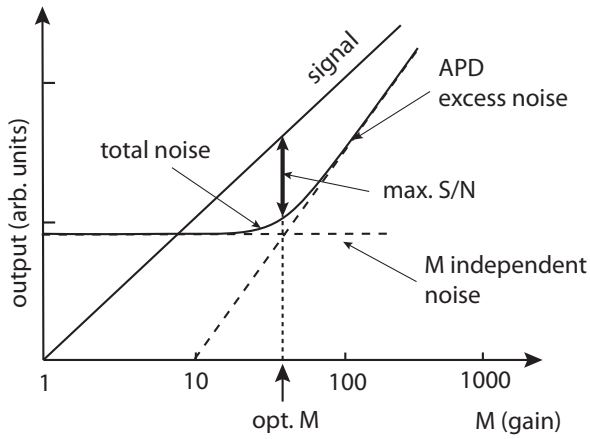


Fig. 10.19 Signal and dominant noise contributions of an avalanche photodiode (APD) as a function of the amplification gain. The noise contributions (dashed) shown are the gain independent contributions (thermal noise of the subsequent electronic amplifier and surface contribution of (10.7)) as well as shot noise (leakage current plus 'excess'-noise) of the APD itself (adapted from [915] with kind permission of Hamamatsu Inc.).

is present in all semiconductor structures, the avalanche process causes its own shot noise contribution called 'excess noise' originating from the statistical fluctuations in the avalanche amplification process. This contribution is described by an 'excess-noise factor' F, by which the leakage current is multiplied. Excess noise deteriorates the noise performance of APDs in comparison to PIN photodiodes without amplification. While the surface contribution of the leakage current does not reach the amplification zone, leakage current originating from the depleted region (absorption region) is amplified like the signal itself.

Altogether we obtain the following shot-noise contribution [915] (see also eq. (17.90) on page 787):

$$d\langle i^2\rangle_{\text{shot}} = 2e\,\langle\, I_{0S} + I_{0B}\,M^2\,F\,\rangle\,df\,, \tag{10.7}$$

where I_{0S} and I_{0B} are the surface and bulk contributions of the leakage current, respectively, and M is the amplification factor (gain). The shot-noise contribution caused by the volume leakage current increases with the gain. Therefore an optimum gain M can be determined for maximal SNR, as detailed in fig. 10.19.

Therefore, although the APD delivers an already amplified signal, the SNR is often not better than that of a photodiode without amplification. At low light intensities, however, the situation can change in favour of APDs. This is because, detection with a photodiode would require to increase the gain $g \approx Z_f/Z_{in}$ of the subsequent amplification circuit, where Z_f and Z_{in} are feedback and input impedance of the amplifier. If a resistive feedback amplifier is used, increasing $Z_f = R_f$ has unwanted effects, namely that (a) the signal response becomes slower[5] and (b) the thermal noise contribution to the output voltage ($d\langle v^2\rangle_{\text{therm}} = 4kT\,R_f$, see section 17.10.3.1) becomes larger. In this situation, i.e. low light intensity, APDs with small excess noise can outperform photodiodes in terms of the total SNR.

G-APDs. Figure 10.20 schematically shows the amplification characteristics of an avalanche photodiode over a large voltage range. Depending on the bias voltage, APDs can be operated in linear or in Geiger mode. For Geiger mode a bias voltage is set

[5] The bandwidth is reduced with higher gain, since the gain × bandwidth product is constant (see eq. (17.18) in section 17.2.3).

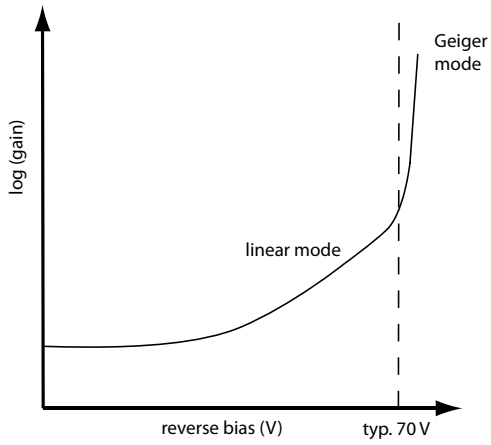


Fig. 10.20 Schematic APD gain curve with indication of the two operation modes 'linear' and 'Geiger'.

approximately 10–20% above the nominal breakdown voltage[6]. The APD then breaks down upon occurrence of an ionising photon signal, that is, delivers a large discharge current, independent of the magnitude of the primary ionisation. In this mode the APD is a counting device sensitive also for single photons (*single photon avalanche diode*, SPAD). After the breakdown the discharge is quenched by passive (via a resistor) or active (via a transistor circuit) quenching methods, which usually act such that the bias voltage decreases. Besides the (always almost constantly) large output signal the advantages of the Geiger operation mode include the high light sensitivity even for single photons (starting from gains of about $M > 10^5$) with excellent timing resolution ($\ll 1$ ns). Also the G-APD's insensitivity to magnetic fields (compared to linear mode APDs) is advantageous for their use in many particle physics experiments.

SPADs are often arranged in arrays of many pads ($\sim$mm^2) or pixels ($\sim$0.01 mm^2) to detect light over a larger area ($>$ cm^2); see also sections 10.4.1, 10.4.2 and 10.5. The photon detection efficiency P_{det} for a SPAD array is approximately composed by the product of the quantum efficiency QE for e/h pair creation, the probability P_G for the Geiger process (not every photon initiates a Geiger breakdown) and the photoactive fraction (*fill factor*) of the SPAD area f_A:

$$P_{\text{det}} = QE \times f_A \times P_G \,. \tag{10.8}$$

Reflections at the SPAD surface and recovery times are here neglected, assuming that the cell occupancy is sufficiently low. Note that also for single cell APDs f_A is usually not 100%, because the active area is usually less than the electrode area and the photon entrance window, but is rather given by the implant area of the amplification region. Typical values for P_{det} of 60% (50%) in the blue (green) wavelength region are reached.

Following the substantially improved technical quality of APDs since the late 1990s their usage for the readout of particle detectors has much increased (see also sections 10.4 and 10.5). The $PbWO_4$ crystals in the barrel part of the CMS electromagnetic calorimeter [298] at the LHC for example are read out by (linear mode) APDs (see section 15.5.2.2).

[6]Note that for semiconductors Geiger operation is not distinguished from break down operation, different to gas-filled detectors (see section 7.4.2).

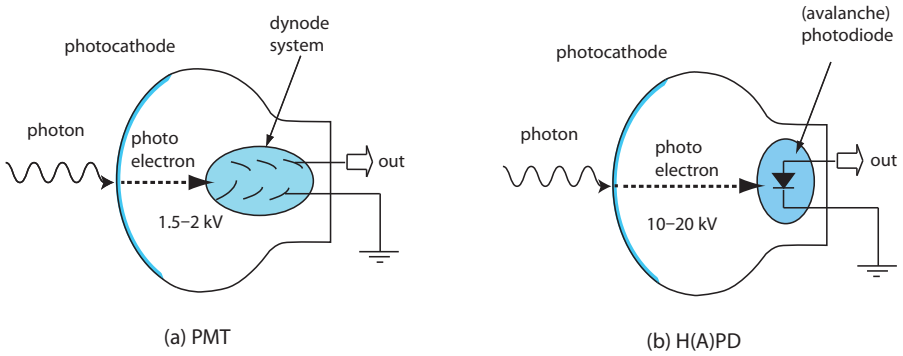


Fig. 10.21 The PMT principle (a) compared to the principle of a hybrid arrangement (b) of photocathode + vacuum + photodiode or APD.

10.4 Hybrid photodetectors

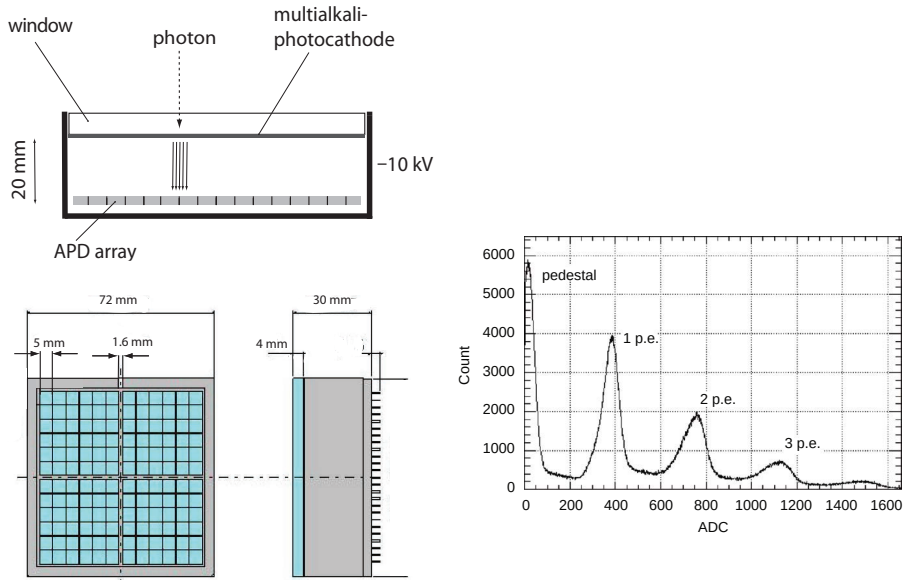
The rapid development of semiconductor detectors for charged particles as well as for light detection has initiated further development of the standard PMT which have led to various 'hybrid' photodetector devices in which PMT and semiconductor features are combined. By means of these hybrid developments certain disadvantages of the individual system parts for some application can be removed. For example, the wavelength range of photodiodes can be extended into the short wavelength regime by employing a bialkali photocathode as the photoconversion element. Or, as another example, replacing the PMT dynode system by a semiconductor detector drastically reduces the sensitivity of the device to magnetic fields. Furthermore, one can exploit the possibility that semiconductor diodes can easily be structured, to create anodes with spatially resolving pad or pixel structures. Better SNR performance can be obtained due to very small capacitances (pixel structures) and/or when individual cells are read out by single step amplification (small cell APDs, SPADs).

If one replaces the dynode system of a PMT by a photodiode with pixel- or pad-structured electrodes or by an avalanche photodiode (fig. 10.21), a hybrid photodiode (HPD) or a hybrid avalanche photodiode (HAPD) is obtained, respectively [922,586]. Hereby one exploits the fact that electrons created in a photocathode and accelerated onto a silicon detector create e/h pairs in the silicon by ionisation. This corresponds to an amplification of the photoelectron signal.

Following the argumentation line replacing more and more PMT elements by semiconductor components, we first describe in section 10.4.1 the HAPD, for which the PMT amplification system is replaced by a relatively coarse (pads) APD array. The HPD, by contrast, not having intrinsic amplification but featuring a finer array structure (pixels) is described in section 10.4.2.

10.4.1 Hybrid avalanche photodiode (HAPD)

A HAPD consists of a photosensitive cathode and an array (typically 12×12 cells) of avalanche photodiodes with dimensions in the range of $5 \times 5\,\mathrm{mm}^2$ per cell. Figure 10.22(a) shows the device geometry and principle. Photoelectrons generated in the photocathode are accelerated by a high electric field over a short distance (2 cm) and deposit their kinetic energy (some keV) in the semiconductor diode by ionisation (e/h pair creation). The number of produced e/h pairs corresponds to an amplifica-



(a) Composition. (b) Photoelectron spectrum.

Fig. 10.22 Hybrid avalanche photodiode (HAPD): (a) composition, (b) signal spectrum (see text). From [40] with kind permission of Elsevier.

tion of the initial photon by a factor of about 10^3. The APD, operating in linear mode (see section 10.3.2 and fig. 10.20) adds another amplification factor of 40–50, together about 5×10^4.

The characteristic properties of the HAPD can be described as follows. The APD array has an active area of about 70% of the total area of the entrance window. As for PMTs, the total quantum efficiency is limited mainly by the QE of the photocathode to about 25%. While the total gain is lower than for PMTs, the high first stage gain determines the signal-to-noise ratio and is thus responsible for a better energy resolution for single photons. In HAPDs, one photoelectron produces over 1000 electrons even before avalanche amplification, whereas the first dynode gain of a PMT is typically only 5 or 6 electrons. A photon (photoelectron) spectrum is shown in fig. 10.22(b). The individual peaks are well separated. The spectrum can be compared to that of a PMT at very low intensities shown in fig. 10.13. A common measure to characterise the single photon performance of photodetectors is the *peak-to-valley ratio* (P/V) defined by the height of single photon peak divided by the valley between background and single photon peak. The P/V value for the HAPD read-off from fig. 10.22(b) is about 13.

Due to the high electric field at the entrance stage the sensitivity of HAPDs to external magnetic fields is reduced compared to PMTs. They can be used in fields above 1 T. HAPDs (and also HPDs, see section 10.4.2) also have a fast time response and excellent timing resolution due to the comparatively small capacitances of the array cells.

For the Hyper-Kamiokande project [722], Hamamatsu developed an HAPD with a diameter of 33 cm. It can be compared with a conventional photomultiplier (R8055) of the same size [947]. The HAPD has about the same photocathode quantum efficiency

Table 10.1 Comparison of HAPD and PMT (both Hamamatsu) of the same size (13'') [24].

	13'' HAPD	13'' PMT (R8055)
time resolution (single photon)	190 ps	1400 ps
energy resolution (single photon)	24%	70%
photoelectron collection efficiency	97%	70%
quantum efficiency (photocathode)	$\approx 20\%$	$\approx 20\%$
gain	$\approx 10^5$	$\approx 5\times 10^7$
power	$\ll 700\,\mathrm{mW}$	700 mW

but provides better time and energy resolution as well as higher collection efficiency of the photoelectrons (table 10.1). The total gain, however, is much lower than the gain of the R8055 and further external amplification is needed.

A choice between HAPD and PMT in an application cannot only be based on technical criteria. Often space limitations, device lifetime and reliability, as well as price play a significant role (see also table 10.2 on page 437).

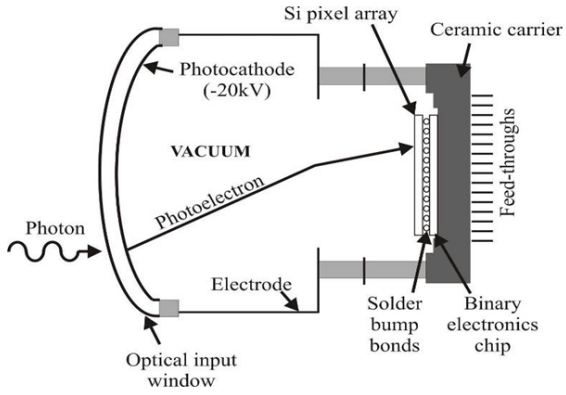
10.4.2 Hybrid photodiode (HPD)

Another hybrid variant of the photomultiplier is the *hybrid photodiode* (HPD) or *pixel hybrid photodiode*. Rather than exploiting the intrinsic amplification in an array with relatively large (mm) pixel/pad sizes, as in the HAPD concept, small, non-amplifying pixel electrodes are used in the HPD concept. The pixel sizes are typically about $100 \times 100\,\mu\mathrm{m}^2$ resulting in small capacitances and hence low noise and good SNR. The dark current, being proportional to the sensitive volume underneath the electrode, is also correspondingly reduced. Amplification is obtained by accelerating the created photoelectron on its way to the pixel detector by a high electric field (typically 5–10 kV/cm). The photoelectron's kinetic energy is converted into 275 e/h pairs in the pixel detector per deposited keV of energy. A typical total signal corresponds to about 5000 e/h pairs [87]. Using a quartz entrance window the HPD can cover a large spectral range from deep UV (200 nm) to red, which is particularly interesting for the detection of Cherenkov radiation.

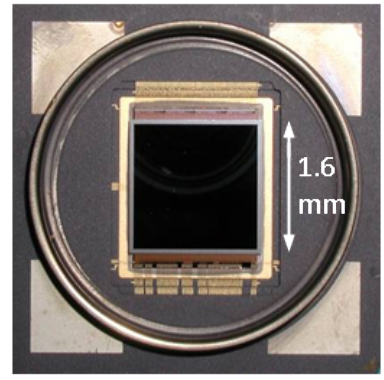
The readout of the HPD pixel detector is done by a separate chip, bonded to the pixel array by the 'bump-bond' and 'flip-chip' technique (see section 8.7). HPD photodetectors are used for example to readout the RICH detectors in the LHCb experiment [417]. Figure 10.23 shows a schematic drawing, the pixel sensor, and a photograph of the assembled HPD [87]. In fig. 10.23(d) the performance of the HPD for single photon detection is shown. The peak-to-valley ratio is about 10–15, to be compared with figs. 10.22(b) and 10.13.

10.5 SiPM: silicon photomultiplier

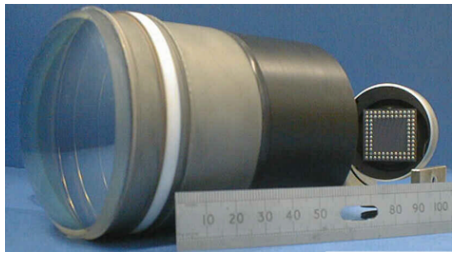
With the technical improvement of Geiger mode APD pixel arrays featuring lower dark count rates, photodetectors with single photon sensitivity were developed that completely relinquish the use of a photocathode. Operating as photon counters these G-APD arrays are called MRS-APD (*metal resistive layer semiconductor* APD), MPPC (*multi-pixel photon counter*) or SiPM (*silicon photomultiplier*). The name silicon photomultiplier has become widely accepted even though the main characteristics of conventional PMTs like photocathode and vacuum tube are not present in SiPMs.



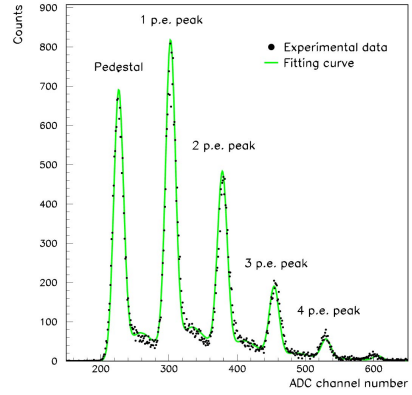
(a) HPD schematic.



(b) Pixel sensor.



(c) Photo: HPD of the LHCb-RICH detector.

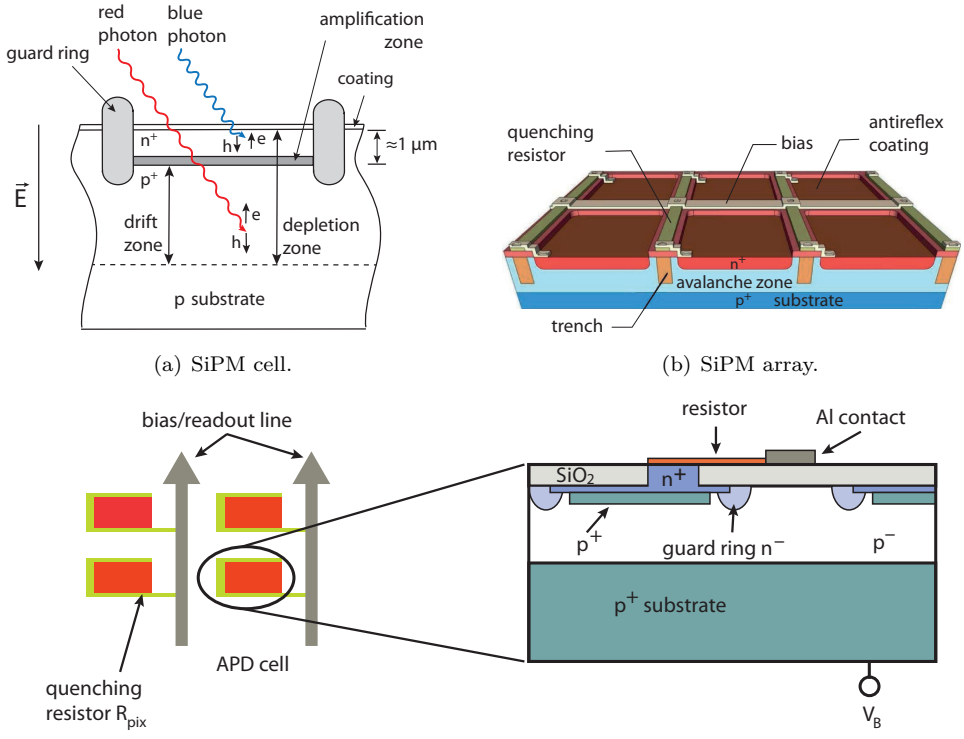


(d) Photoelectron spectrum.

Fig. 10.23 (a) Schematic drawing of a hybrid photodiode (HPD); (b) pixel sensor with 1024 pixels, $500 \times 500\ \mu m^2$ in area; (c) HPD photo, used for readout of the RICH detector of the LHCb experiment [87, 417]; (d) single photon spectrum [657]. Source: LHCb Collaboration and CERN.

The APD pixel matrices of an SiPM are operated in (limited) Geiger mode at an amplification gain of about 10^6 [233, 266]. In this mode each APD cell is a binary (yes/no) operating photodetector. The pixel cells are small with typical dimensions in the range 15–70 µm [266]. The operation principle exploits the high density of the cells. For a photon flux low enough that a given cell is likely only hit by at most one photon (during a time window determined by the time resolution of the device) the number of hit pixels is proportional to the light intensity (number of photons).

The SiPM principle and device technology is explained in fig. 10.24. Figure 10.24(a) shows the cell schematics. Depending on the wavelength different absorption depths apply (see fig. 10.4), for example between 0.6 µm for blue (470 nm) and 2.9 µm for red (625 nm). Charge carriers created in the depletion zone drift in the field towards the amplification zone. After avalanche amplification the charge carriers induce a sufficiently large signal at a readout electrode. Figure 10.24(b) shows a commercial struc-



(a) SiPM cell.

(b) SiPM array.

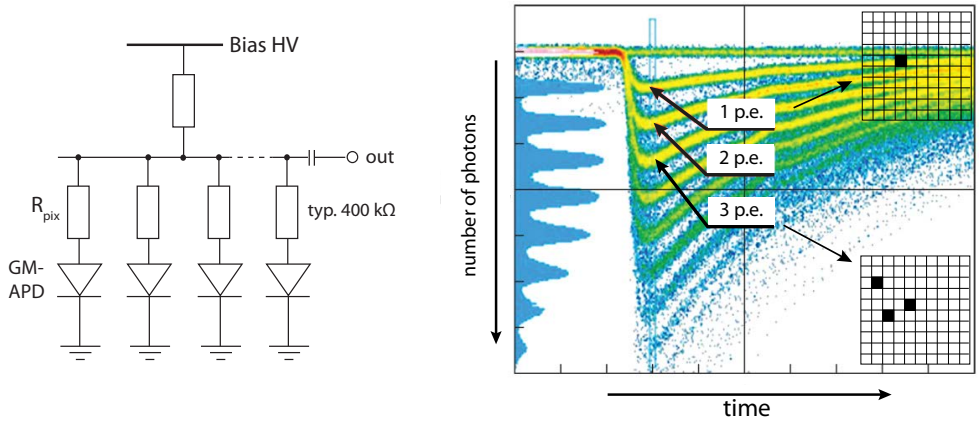
(c) APD cells with quenching structure.

Fig. 10.24 Silicon photomultiplier (SiPM). (a) Schematic of a SiPM cell. Photons of different wavelengths are (on average) absorbed at different depths. The generated charge carriers are amplified in the amplification zone of the APD. (b) SiPM pixel cell (Source: Ketek GmbH [600]): The cells are optically decoupled by trenches. (c) APD cells with quenching structure; detail on the right (adapted from [266]).

ture with antireflex coating and deep trench isolations[7] between the APD cells to suppress cross talk by photons entering a neighbour cell [600].

The operating voltage of the APD array (typically < 80 V) is set about 10–15% higher than the breakdown voltage. The discharge is quenched by using polysilicon strips as quench resistors ($R_{\text{pix}} \approx 400$ kΩ) in every pixel (Fig. 10.24(c)). The discharge current over R_{pix} causes the operating voltage to drop below the breakdown voltage. The resistor also decouples the individual pixels from each other and determines the decay time of the pulse (typical is $C_{\text{pix}} R_{\text{pix}} \approx 30$–$100$ ns). The rise time, by contrast, is given by the rapid avalanche process such that the resulting pulse shape has a fast peaking time (≈ 0.5 ns) and a longer fall time (typically 30 ns to about 100 ns). The recovery time after a signal has occurred ($\lesssim 100$ ns) is short in comparison to other quenched detectors, for instance gaseous detectors.

[7] Trenches are isolation structures filled with dielectrics (e.g. SiO_2). 'Deep' trenches, reaching deeper into the bulk than the surface structures, are produced by special processing techniques (*deep resistive ion etching*, DRIE).



(a) Analog readout. (b) Output pulses.

Fig. 10.25 Analog SiPM. (a) Analog merging of the readout signal in a SiPM. The APD cells are joined via the quench resistors R_{pix}. (b) Readout pulses for different numbers of hit SiPM pixels; left hand side: projected pulse height spectrum (Source: Hamamatsu Inc. with kind permission). The rise and fall times of the pulses are in the range of nanoseconds and tens of nanoseconds, respectively.

10.5.1 Analog SiPMs

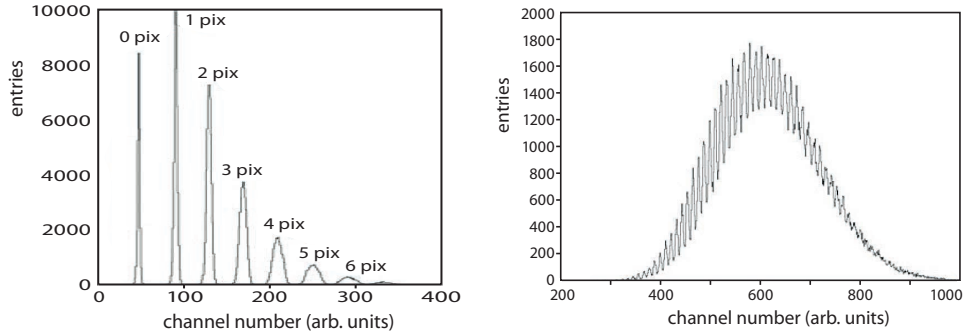
The outputs of the SiPM's cells usually are joined via the quench resistors R_{pix} to a single output node, as shown in fig. 10.25(a). The output signal hence is proportional to the number of hit cells (fig. 10.25(b)). The dynamic range begins with detecting single photons and is limited by the total number of pixels N_{pix} of the SiPM sensor array. In [266] the upper end of the dynamic range is approximately quoted as $N_{ph} \approx 0.6\, N_{pix}$ photons in a specified time window. At higher intensities than this, the dependence of the summed output signal on the incoming photon flux becomes non linear since the probability for multiple hits per pixel is no longer negligible.

The photon detection efficiency of an SiPM detector follows the same criteria as for a single cell APD given by (10.8). The fill factor f_A is in the range 20–80%, depending on the design.[8] Together with the high quantum efficiency of semiconductor photodetectors (70–80%) and a Geiger probability P_G of about 90% photon detection efficiencies P_{det} of up to 60% are reached.

The capability to detect single photons by hit pixel counting is demonstrated in fig. 10.26(a). The output spectrum in fig. 10.26(b) is the SiPM response to a light pulse with on average 46 photoelectrons in the detector [267]. It shows a maximum corresponding to this number. The quantisation of the individual photon contributions is still clearly visible, dropping to zero at $n_{p.e.} \approx 25$ (channel number ≈ 300) following Poissonian statistics.

The SiPM pixel cells have capacitances of typically 100 fF. The single pixel gain (charge per electron) is determined by the pixel capacitance $Q_{\text{pix}} = C_{\text{pix}}\, \Delta V$, where $\Delta V = V_{\text{bias}} - V_{\text{breakdown}}$ is the difference between operating voltage and Geiger breakdown voltage which typically is in the range of a few volts [266]. The signal therefore has a typical charge of some 100 fC corresponding to a gain of 10^6. SiPMs have very

[8] Smaller fill factors have smaller signals but feature smaller capacitances and hence lower noise.



(a) SiPM: single photon spectrum. (b) SiPM: signal of a light pules.

Fig. 10.26 Silicon photomultiplier (SiPM). (a) single photon spectrum, (b) signal distribution of light pulses with on average 46 photoelectrons (adapted from [267], with kind permission of Elsevier).

good time resolution also for single photons of $\sigma_t^{1\gamma} \approx 100$ ps. Due to the high gain, noise introduced by the subsequent electronics is negligible. SiPMs can be operated at room temperature as well as in magnetic fields up to 7 T, requiring operating voltages of only about 50 V.

The use of SiPMs as excellently performing photodetectors is only limited by the *dark pulse rates*, which are several orders of magnitude higher (10–20 MHz/cm^2) than for PMTs (<1 kHz/cm^2), albeit SiPMs are continuously improving in this respect. Dark pulses in counting SiPMs have the following physical origins:

- the sensitive zone of the substrate or in depleted regions at the surface, their generation being aided (and amplified) by the high electric fields in the amplification region. This also causes a strong temperature dependence.
- 'After-pulsing': These are delayed pulses created by charge carriers which have been trapped during the avalanche amplification and are released after a characteristic trapping time of typically some hundred nanoseconds, thereby initiating new avalanches.
- Optical cross talk: These are photons created during the amplification process reaching across boundaries into a neighbour SiPM cell where they initiate another avalanche.

The dark rate per area $\dot{N}_{\text{dark}}$ multiplied by the response time plus the quench and recovery time Δt_{tot} of a SiPM cell determines the occupancy (by noise hits) P_{occ} of a SiPM:

$$P_{\text{occ}} = \frac{\dot{N}_{\text{dark}}}{n_{\text{pix}}} \Delta t_{\text{tot}} \,, \tag{10.9}$$

where n_{pix} is the SiPM cell density. Assuming a dark rate of 20 MHz/cm^2, cell sizes of $50 \times 50\,\mu\text{m}^2$, and a response and recovery time of $\Delta t_{tot} = 100$ ns yields $P_{\text{occ}} = 5 \times 10^{-5}$. Note, however, that cross talk between neighbouring cells violates the assumption made in (10.9) that the responses of the SiPM cells are uncorrelated.

To suppress the dark count rate one exploits the fact that a true light signal fires SiPM cells in coincidence, thus producing a large energy signal in comparison which is also output (fig. 10.27(a)). A threshold applied to the energy signal therefore suppresses

the dark count rate. The contribution of photon cross talk to the dark count rate is reduced by trenches, as already mentioned above.

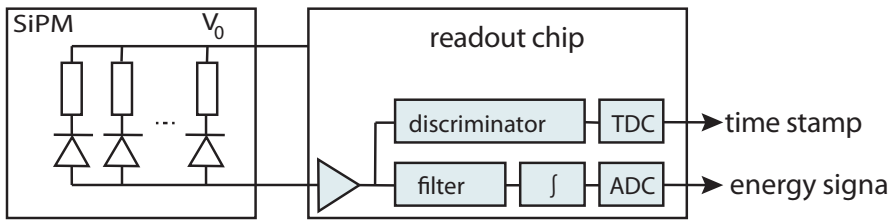
More and more often silicon photomultipliers replace conventional PMTs or other photodetectors in particle or astroparticle physics experiments, in particular for the readout of homogeneous as well as sampling calorimeters (see chapter 15). An example is the development of the CALICE calorimeter [53] for the ILD detector [169] planned at a future Linear Collider.

10.5.2 Digital SiPMs

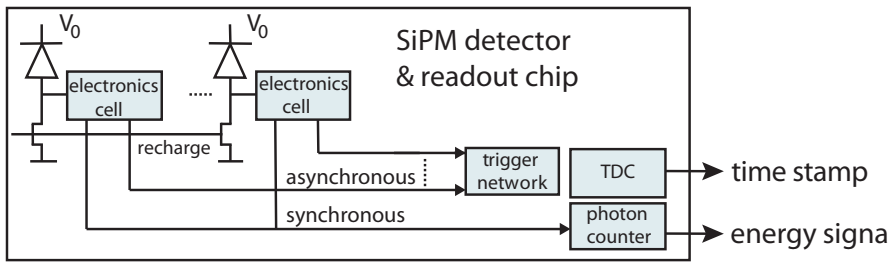
Even though analog SiPMs detect individual single photons, their total count is derived from an analog signal sum. The seemingly obvious approach of directly summing digital 0/1 information of every SPAD is electronically more involved and has been realised only in later phases of the SiPM development.

In digital SiPMs (dSiPM) [422, 423] a digital output signal is formed from the sum of SiPM cells that fired rather than from the analog signal sum of analog SiPMs as explained above, thus treating the SPAD cells as binary switches. Every SPAD cell is an individual unit that can also be switched off if for instance its dark rate is too high. The cell can detect and store one and only one photon. After the photon is detected the quenching process for the hit cell is immediately started. This is realised by MOS transistor circuits replacing the passive quench resistors in fig. 10.24(c), the latter being more difficult to realise on microchips because (large) resistors require a lot of space in integrated circuits. After the voltage has dropped below a certain threshold and the current flow through the diode has come to a halt, the SPAD cell is switched on again, that is, put in Geiger mode with voltage above the breakdown voltage.

The advantage of dSiPMs lies in the integration of all components in the same chip, a fast quenching process, the possibility to disable (noisy) cells in a simple way and the integration of additional intelligent circuit blocks on-chip. Possible disadvantages



(a) Analog SiPM.



(b) Digital SiPM.

Fig. 10.27 Electronics block diagram (schematic) of (a) analog and (b) digital SiPMs (adapted from [421], © SISSA Medialab Srl., with kind permission of IOP Publishing).

are higher power consumption and other consequences inherited from the integration of sensor and chip, for example possible cross talk.

An electronics block diagram and logic of a dSiPM is shown in fig. 10.27(b) compared to that of an analog SiPM (fig. 10.27(a)) [423]. Every cell delivers a fast, synchronous trigger signal which is received by an on-chip trigger network as well as a slower binary output signal which is read out synchronously with other cells. The trigger network starts a TDC (see section 17.7.3) either with the first photon or otherwise (selectable), when dark rates are high, after a minimum number of photon signals. Every other hit of a SiPM cell occurring after the photon triggered first, is regarded as a dark hit and is automatically cleared if it does not lead to another trigger within 15 ns. The slower energy signal corresponding to the number of photons is read out if it is coincident with the (delayed) trigger signal and higher than a given a threshold.

Digital SiPMs were first developed for light intensities with a sufficiently large numbers of photons, but they are also sensitive to single photons. Cherenkov photons, for example, can be detected with very good time resolution (140 ps) and sufficiently low dark rate [421]. The quantum efficiency is somewhat below that of analog SiPMs or photodiodes because the CMOS implantations and metal layers absorb photons undetected, thus reducing the yield [421].

10.6 Photodetectors in comparison

The choice of a photodetector to read out a light generating particle detector depends on many factors, not least the detector's size, its surroundings (available space, magnetic field, etc.) and the price. In table 10.2 we compare important characteristics of photodetectors introduced in the previous sections without claiming completeness. For clarity typical and characteristic values are given; deviations therefrom—often substantial—can occur in dedicated designs for specific applications.

In this chapter we have discussed photodetectors by mainly distinguishing those using photocathodes for photon (to electron) conversion and those purely based on direct detection in semiconductors, and also hybrid detectors featuring elements of both. As shown in section 10.2.1 photocathodes largely determine the quantum efficiency obtainable with such devices, typically yielding about 25–30% . They are sensitive in the lower wavelength range (300–550 nm), whereas semiconductor (i.e. Si)-based direct photodetectors cover a larger wavelength range from about 400 nm to 1000 nm and reach higher quantum efficiencies ($> 50\%$ up to 75%) at wavelengths above 400 nm (see fig. 10.17).

Photomultipliers and variants are bulkier and less compact than photodiodes and APDs, but they are still an excellent choice for many applications when low light intensities (down to single photon detection) need to be detected with high sensitivity and low dark rates. Also the ease of handling commercial PMTs in an experiment or a lab set-up is often an important factor to consider.

PMTs are available in a large variety and in many different sizes (see fig. 10.12 for a very large example), but they are also very expensive. Microchannel plates (MCPs) are unequaled regarding time resolution.

Photodiodes and APDs operating in linear mode are often preferred when sufficient light intensity is available and also when price considerations play a role. Choosing between the two depends on the SNR obtainable for a certain application, considering the balance between the APD's internal amplification and its excess noise in SNR. In particle physics photodiodes and APDs are often used for the readout of crystal

calorimeters (section 15.5.2) in large-scale experiments when price and space requirements play an important role or when magnetic fields impede the electron transport in the dynodes of PMTs. Examples are the calorimeters of the experiments Crystal Barrel (CsI(Tl) crystals, wavelength shifters with photodiode readout) [67] and CMS ($PbWO_4$ crystals, APD readout) [298].

The compact and cheap SiPMs, using Geiger mode single pixel APDs (SPADs) in an array structure, are about to challenge PMTs for many applications. SiPMs equal or even surpass PMTs regarding fast response and single photon sensitivity. However, they are still lagging in acceptance area (size), linearity, and dark count rate.

Hybrid photodetectors are used when a certain disadvantage of a PMT or a (A)PD solution is to be remedied. An example is the detection of light with low intensity in a magnetic field. For this application the hybrid photodiode has been developed (section 10.4.2) featuring a photocathode and then a single stage amplification by accelerating the photoelectron to a pixellated semiconductor diode. This arrangement is much less sensitive to magnetic fields than PMTs with lower gains in the first amplification stage.

The SNR is a measure for the purity that can be reached when detecting multi-photon signals. For all sensor types the light signal generally depends on the entrance area A of the sensor, the quantum efficiency QE, the amplification gain M, as well as on the number N_{γ_1} of generated primary photon signals:

$$S \quad \propto \quad N_{\gamma_1} \times A \times M \times QE\,. \tag{10.10}$$

In table 10.2 we assume $N_{\gamma_1} = 1$ for the SNR values quoted.

This is to be distinguished from single-photon detection, for instance in a Cherenkov detector. Here the SNR (in this case not depending on A) together with sufficiently low dark count rates are important. Photodetectors with intrinsic gain show a better SNR for single photons if the noise contributions are not at all or only weakly amplified.

Contributions to the noise $\mathcal{N}$ contain portions of the photosensor leakage current I_0 (if applicable multiplied by a factor F originating from excess noise introduced by the avalanche process, see page 425) but also portions $\mathcal{N}_{amp}$ if a subsequent amplifier exists (e.g. for photodiode-based readout, dominated by the diode capacitance C_D, see section 17.10.3.3), both added in quadrature ($\oplus$):

$$\mathcal{N}_{tot} \propto \sqrt{2e\,I_0\,F} \;\oplus\; \mathcal{N}_{amp}(C_D)\,. \tag{10.11}$$

Here the $\mathcal{N}$ terms designate noise contributions in the unit $\mathrm{pA}/\sqrt{\mathrm{Hz}}$. The SNR entries in table 10.2 (together with the dark count rate) reflect the suitability of the respective photodetector for single-photon detection applications or for applications with large amounts of light.

For small illuminated area and sufficient light intensity photodiodes have the best cost/performance ratio. By means of wavelength shifters larger surfaces than the surface of the diode itself can also be covered, as shown in fig. 13.17 in section 13.4.2.

Table 10.2 Characteristics of different photodetectors in comparison. Given are typical values without claim of completeness and accounting for all detector variants.

The abbreviations are: PMT = photomultiplier, VPT = vacuum phototriode, MCP = microchannel plate, PD = photodiode, APD = avalanche photodiode operated in linear mode, HAPD = hybrid APD (linear operation mode), HPD = hybrid photodiode, SiPM = silicon photomultiplier (APD arrays operated in Geiger mode).

QE = quantum efficiency at λ_{max}, λ_{max} = approximate wavelength of maximum efficiency, $\Delta\lambda$ = sensitive wavelength region at 50% QE, gain = intrinsic amplification, PDE = typ. photon detection efficiency according to (10.8) at λ_{max}, DR = dark rate at room temperature if the device is sensitive to single photons, SNR (1γ) = signal-to-noise ratio related to a single primary photon (typical, dependent on the bandwidth), σ_t = time resolution, A = typical (maximal) sensitive area, V_{op} = operating voltage, v/c = indicates if a detector is typically rather voluminous or compact/small, sens. 1γ = single photon detection possible (y/n), sens. B = sensitive to magn. fields, sens. T = temperature sensitive gain, price = high/medium/low for small area units.

	PMT	VPT	MCP	PD	APD	HAPD	HPD	SiPM
QE (%)	25 – 40	20 – 22	20 – 25	75	75	25	25	70 – 80
λ_{max} (nm)	400	400	400	750 – 900	750 – 900	400	400	750
$\Delta\lambda$ (nm)	300 – 550	300 – 550	300 – 550	400 – 1000	400 – 1000	300 – 550	300 – 550	400 – 1000
gain	10^{5-9}	10	10^{4-6}	1	50	10^{4-5}	3000	10^{5-6}
PDE (%)	25	20	20	75	60 – 70	20	20	60
DR ($\frac{\text{MHz}}{\text{cm}^2}$)	$< 10^{-3}$		$< 10^{-3}$			$< 10^{-2}$		10 – 20
SNR (1γ)	10^{2-3}	3×10^{-4}	10^{1-2}	5×10^{-4}	3×10^{-4}	≈15	≈20	>100
σ_t (ps)	200	1000	10	2000	20	200	50	100
A (cm^2)	< 2000	< 50	< 200	< 5	< 1	50	2.5	1
V_{op}	1.0–1.5 kV	1 kV	2.5 kV	5–40 V	100–500 V	> 10 kV	< 20 kV	50 V
vol/comp	v	v	v/c	c	c	v/c	v	c
sens. 1γ	yes	no	yes	no	no	yes	yes	yes
sens. B	high	low	medium	low	low	medium	low	low
sens. T$^{(*)}$	med/low	medium	high	low	high	high	low	high/med.
price$^{(*)}$	high	high	high	low	medium	medium	medium	medium

$^{(*)}$ Ratings given are subject to rapid changes, especially for newly developed devices.

11
Cherenkov detectors

The Cherenkov effect is the phenomenon that particles traversing matter with relativistic velocities emit electromagnetic radiation if their velocity exceeds the velocity of light in that medium. (particle identification, see also chapter 14). In astroparticle physics large volume detectors exploit Cherenkov radiation for cost-efficient detection of showers initiated by cosmic particles, such as nuclei, gamma rays or neutrinos (see also chapter 16). In this chapter we first describe the physics of Cherenkov radiation and then focus in particular on its use for particle identification. In this chapter we first describe the physics of Cherenkov radiation and then focus in particular on its use for particle identification.

11.1 The Cherenkov effect

If a charged particle with velocity v traverses a medium with refractive index n, with v being larger than the phase velocity of light $c_n = c_0/n$ in that medium, electromagnetic radiation is emitted. This radiation is called *Cherenkov radiation* after Pavel A. Cherenkov who received the Physics Nobel Prize 1958 jointly with Ilya M. Frank and Igor Y. Tamm for 'the discovery and the interpretation of the Cherenkov effect' [302]. Cherenkov published the discovery in 1934 [303, 304] and Frank and Tamm provided a quantum theoretical treatment in 1937 [426].

Cherenkov radiation is caused by the polarisation of a dielectric medium by a charged particle passing through the medium when the polarisation becomes asymmetric with respect to the flight direction, as shown in fig. 11.1 [571]. The (close to spherical) atoms of the medium become short-time polarised by the interaction with the electromagnetic field of the passing particle and are elliptically stretched. They quickly relax back to their original size while the particle continues flying past further atoms and polarising them. Every infinitesimal region of the medium thereby generates an electromagnetic impulse of dipole radiation.

In an isotropic medium the polarisation is symmetric in azimuth and, at particle velocities v small compared to the propagation velocity c_n of electromagnetic fields in the medium, the polarisation field is also symmetric along the flight direction of

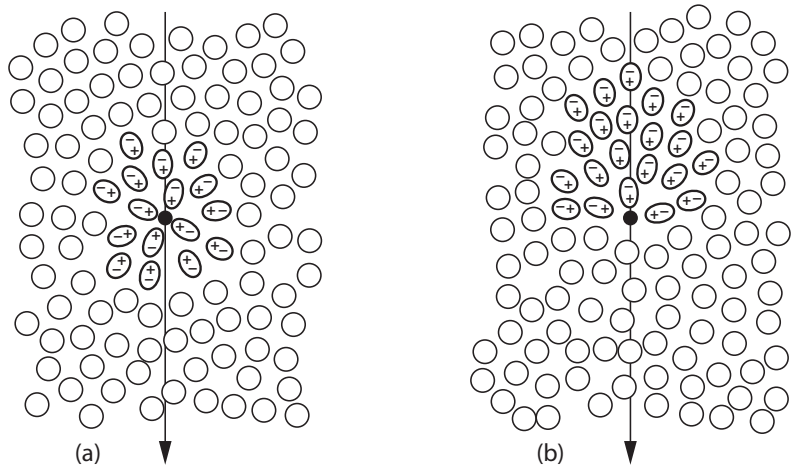


Fig. 11.1 The Cherenkov effect. (a) The particle has a velocity that is small compared to the velocity of light inside the medium ($v \ll c_n = c_0/n$). The medium becomes symmetrically polarised. At a distance the contributions from the local dipole fields cancel. No radiation is observed. (b) The particle's velocity is of the order of the velocity of light c_n in that medium. The medium is polarised asymmetrically along the particle's flight path. A (momentary) polarisation field results being measurable also at a distance. For $v > c_n$ the elementary waves emitted from many emission points along the trajectory superimpose coherently and Cherenkov radiation occurs (adapted from [571], with kind permission of Elsevier).

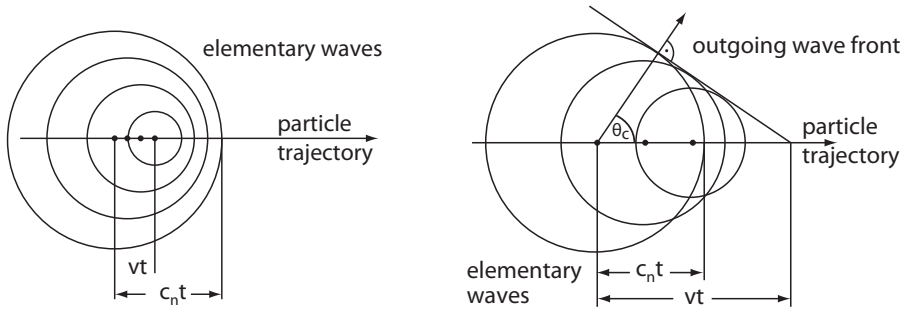
the particle (fig. 11.1(a)). Owing to this complete symmetry of the polarisation field around the track, there is no resulting field at a distant point and hence no radiation.

As soon as v approaches c_n the picture changes (fig. 11.1(b)). The particle's propagation and the polarisation of regions are of the same order in time. The dipoles along the direction of the particle's path can no longer develop and relax fast enough, such that the momentary polarisation field at a point P on the track is no longer symmetric with respect to the actual position (and direction) of the particle. There is a resultant dipole field which will be apparent even at large distances from the particle path. Each dipole element, set up one after another along the flight path, releases its energy in a brief electromagnetic pulse, which corresponds to a characteristic spectrum in the frequency domain (Fourier space).

The Cherenkov relation. The elementary waves created at the individual points along the particle track superimpose according to Huygen's principle. In the general case, integrated over all parts of the track, there is no coherent phase superposition, such that even for asymmetric polarisation along the track there is no net intensity observed at a distance. If, however, the particle velocity exceeds the phase velocity of light c_n in that medium, the elementary waves from different points along the flight path can superimpose constructively, creating a wave front, the Cherenkov radiation.

The phenomenon is analogous to the generation of Mach's sound cone observed for sound waves, for example when airplanes exceed the sound velocity (at a threshold velocity of about $v \approx 1235$ km/h at sea level and 20 °C).

Figure 11.2(b) shows a snapshot of elementary waves along the particle trajectory. They propagate with the medium speed of light $c_n = c_0/n$ and interfere constructively



(a) $v < c_n$. (b) $v > c_n$.

Fig. 11.2 If a charged particle traverses a medium with a velocity greater than the speed of light c_n in that medium, an electromagnetic shock wave can be formed by constructive interference of many elementary waves emitted along the particle track: (a) $v < c_n$, no radiation, (b) $v > c_n$, Cherenkov radiation is emitted under a fixed angle θ_c while the wave front forms under an angle $90^\circ - \theta_c$.

to a wave front under a fixed angle θ_c, the Cherenkov angle, easily computable from fig. 11.2(b) as

$$\cos\theta_c = \frac{\frac{c_0}{n}\,t}{v\,t} = \frac{1}{\beta\,n}\,. \tag{11.1}$$

Here the recoil caused by the photon emission is neglected. Exact quantum mechanical calculations [697], however, show that this approximation is valid for most cases. The index of refraction is frequency dependent $n = n(\omega)$ (dispersion). Hence the simple relation (11.1) is valid only for electromagnetic radiation of a given frequency ω. Radiation occurs if

(1) the particle velocity is larger than the phase velocity of electromagnetic waves of this frequency in the respective medium, that is, $v > c_0/n(\omega)$;
(2) the radiator medium is optically transparent and its length L is much larger than the wavelength of the radiation, $\lambda \ll L$, such that coherent superposition of the elementary waves can occur.

Since $n = n(\omega)$, the emission angle θ_c depends—apart from its dependence on β—on the frequency of the emitted radiation. At sufficient distance from the particle track the radiation waves are transverse waves. They are linearly polarised in a plane containing the direction of wave propagation (given by the Poynting vector[1] $\vec{S} = c^2\epsilon_0(\vec{E} \times \vec{B})$) and the particle direction. A theoretical treatment of Cherenkov radiation employing classical electrodynamics is given for example in [564].

The Cherenkov threshold. From (11.1) one can derive further relations useful for the design and construction of detectors. The Cherenkov angle increases with increasing velocity or momentum of the particle. The maximum angle of Cherenkov radiation ('asymptotic angle') is obtained when the charged particle approaches the velocity of light in vacuum, that is, it is 'asymptotically fast' with $\beta \to 1$, $\gamma \to \infty$:

[1] Directed energy density flow of the radiation.

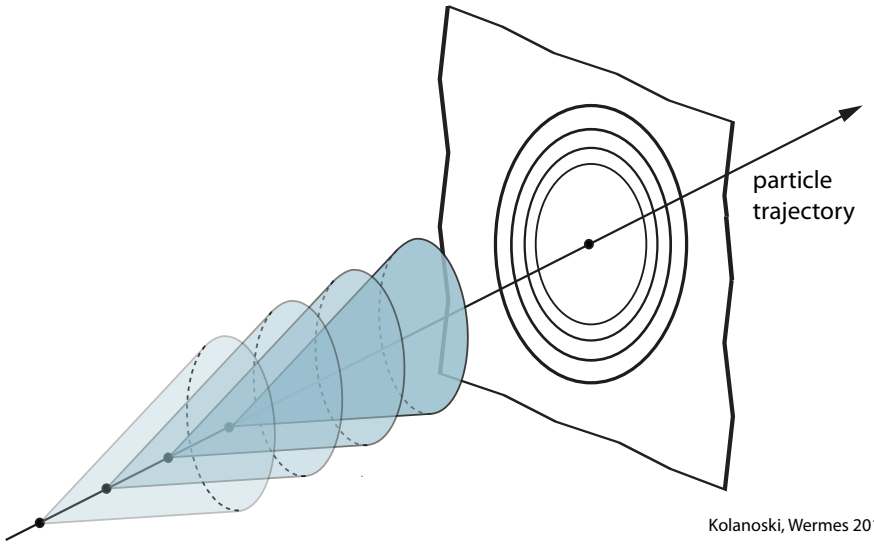


Fig. 11.3 As seen from a radiator length element dx on the particle trajectory, the directions of Cherenkov radiation with frequency ω lie on cone surfaces with fixed opening angle θ_c. The space points of Cherenkov photons on a plane perpendicular to the particle trajectory are circles (Cherenkov rings) with different radii for different emission points. For continuous emission in a thick radiator and without focusing elements the resulting image is a broad circular structure.

$$\cos\theta_{max} = \frac{1}{n}\,, \tag{11.2}$$

hence depending only on the optical properties of the traversed medium.

Furthermore, with (11.1) and $\cos\theta_c \leq 1$ one concludes that $\beta \geq 1/n$ must hold for Cherenkov radiation to occur and that the threshold velocity (under the assumption of a constant index of refraction n) is

$$\beta_{th} = \frac{1}{n}\,. \tag{11.3}$$

The emission angle at threshold velocity is zero, $\theta_{th} = 0°$. The threshold velocity β_{th} hence marks the onset of Cherenkov radiation whose emission angle and also intensity (see section 11.2) further increase with increasing β. For a fixed radiation frequency the direction of the emission lies on the surface of a sharp cone with opening angle θ_c (fig. 11.3) starting from a given point on the particle trajectory. The cone's opening widens with increasing velocity of the particle according to (11.1) up to a maximum angle θ_{max} given by (11.2). No radiation is emitted inside or outside the cone. Dispersion, that is, the frequency dependence of n, introduces a smearing of the precise value of θ_c. The wave front shown in fig. 11.2(b) also forms a cone (with opening angle $90° - \theta_c$), whose envelope is given by the space points of emitted photons at a fixed time t.

Using (11.3) the threshold energy of the particle belonging to θ_{max} can be expressed by the Lorentz factor γ at threshold:

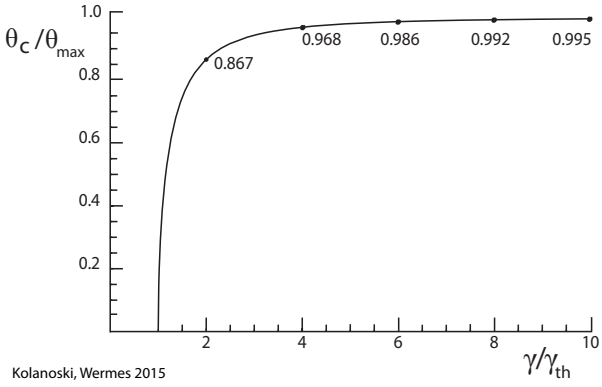


Fig. 11.4 Dependence of the normalised emission angle θ_c/θ_{max} on γ/γ_{th}. The values on the ordinate axis for $\gamma/\gamma_{th} =$ 2,4,6,8,10 are also shown on the curve to illustrate the limited 'dynamic range' restricting the obtainable resolution for γ when measuring θ_c.

$$\frac{E_{th}}{mc^2} = \gamma_{th} = \frac{1}{\sqrt{1-\beta_{th}^2}} = \frac{n}{\sqrt{n^2-1}}\,. \tag{11.4}$$

With (11.2) we then have

$$\sin\theta_{max} = \sqrt{1-\cos^2\theta_{max}} = \sqrt{1-\frac{1}{n^2}} = \frac{1}{\gamma_{th}}\,. \tag{11.5}$$

With (11.4) and (11.5) a relation between the actual emission angle θ_c and the maximum (asymptotic) angle θ_{max} results:

$$\begin{aligned}\sin^2\theta_c &= 1-\cos^2\theta_c = 1-\frac{1}{\beta^2 n^2} = 1-\frac{\beta_{th}^2}{\beta^2} = \frac{1}{\gamma_{th}^2}\frac{\gamma^2-\gamma_{th}^2}{\gamma^2-1}\\ &= \sin^2\theta_{max}\,\frac{\gamma^2-\gamma_{th}^2}{\gamma^2-1}\,.\end{aligned} \tag{11.6}$$

For relativistic particles ($\gamma^2 \gg 1$) and small Cherenkov angles (typical for applications with gaseous radiators) the ratio of the actual emission angle (i.e. for a particle with energy $\gamma = E/mc^2$) and the maximum possible angle θ_{max} (depending only on the refractive index n of the radiator) is

$$\frac{\theta_c}{\theta_{max}} = \frac{R_c}{R_{max}} \approx \sqrt{1-\frac{\gamma_{th}^2}{\gamma^2}}\,. \tag{11.7}$$

Here R_c and R_{max} are the actual and the maximum radius of a Cherenkov ring (see fig. 11.3).

The obtainable resolution on γ by measuring Cherenkov ring radii with a given precision is limited (limited 'dynamic range') due to the saturation of θ_c/θ_{max} with increasing γ, as is evident from fig. 11.4. Already at $\gamma = 2\gamma_{th}$ the ring radius has reached 87 % of the asymptotic radius R_{max}. See section 11.3 for details on the experimental resolution of Cherenkov rings.

11.2 Emission spectrum

The energy emitted by Cherenkov radiation per path length and frequency interval was first calculated by Frank and Tamm [426] yielding

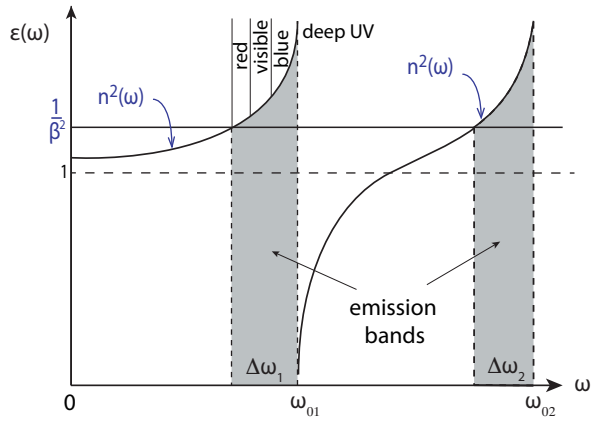


Fig. 11.5 Dependence of $\epsilon(\omega)$ as a function of frequency (schematic) in an ideal medium without damping terms with two assumed atomic resonance frequencies ω_{01} and ω_{02}. Cherenkov radiation occurs in frequency bands $\Delta\omega_i$ for which $n^2 > 1/\beta^2$ is valid, ending at ω_{0i} (adapted from [571]).

$$\frac{d^2E}{d\omega dx} = \frac{z^2e^2}{4\pi\epsilon_0 c^2}\,\omega\,\underbrace{\left(1-\frac{1}{\beta^2 n^2(\omega)}\right)}_{\sin^2\theta_c(\omega)}\,. \tag{11.8}$$

In addition to the original article, the derivation can be found in books on electrodynamics, for example in [564]. Radiation occurs if $\beta^2 n^2(\omega) > 1$, otherwise the term in brackets becomes zero or negative.

For optical, non-magnetic materials ($\mu \approx 1$) the refractive index is $n(\omega) \approx \sqrt{\epsilon(\omega)}$ with ϵ being the relative permittivity of the material. In the neighbourhood of absorption bands, ϵ is complex and can be written as

$$\epsilon(\omega) = \mathrm{Re}\,\epsilon(\omega) + i\,\mathrm{Im}\,\epsilon(\omega)\,. \tag{11.9}$$

The real part of ϵ is a measure of dispersion while the imaginary part arises from absorption. A familiar expression for $\epsilon(\omega)$ is a sum over atomic or molecular resonances

$$\epsilon(\omega) = 1 + a\,\sum_i \frac{f_i}{\omega_{0i}^2 - \omega^2}\,, \tag{11.10}$$

where $a = n_a\,e^2/(\epsilon_0 m_e)$ with n_a = atom density, m_e = electron mass, and with oscillations resonating at $\omega = \omega_{0i}$ with relative oscillator strengths f_i [571, 434] which are normalised as $\sum_i f_i = Z$ (see also eq. (3.114) on page 74). Damping terms have been neglected. Real media always show dispersion such that radiation only occurs in frequency bands $\Delta\omega$ for which $n^2(\omega) > 1/\beta^2$ is valid, as shown in fig. 11.5. These bands tend to lie at frequencies below regions of anomalous dispersion (i.e. where $dn/d\omega > 0$).

The dispersion curve for media that are transparent in the optical regime is shown in fig. 11.6 for a large wavelength range. In the visible region of the spectrum the emission is cut off towards ultraviolet (UV) wavelengths by absorption bands, which occur at transitions from normal to anomalous dispersion. The fact that Cherenkov light, emitted by electrons from neutron decays in the cooling water of nuclear reactors, appears blue is, however, at most partly due to the blue enhancement of the intensity spectrum. Scattering ($\propto \omega^4$) and light absorption in the infrared, folded with the spectral sensitivity of the human eye, must be considered for a quantitative

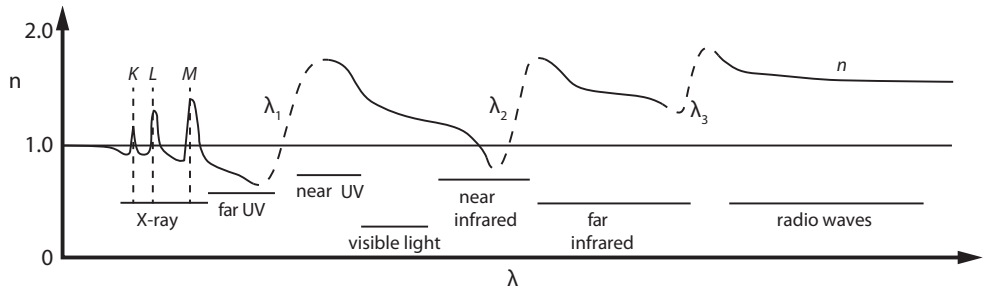


Fig. 11.6 Typical dispersion curve for an optically transparent medium over a large wavelength range of the electromagnetic spectrum. Damping is included in contrast to fig. 11.5. The regions of anomalous dispersion are shown dashed (adapted from [571] with kind permission of Elsevier).

explanation.

In the spectral range where $n(\omega) < 1$, as is the case in the far UV, radiation is forbidden. Further emission bands can be found in the far infrared and the radio spectral regions as well as at discrete X-ray energies [163, 618].

The radiated energy per frequency interval is proportional to ω. Dividing (11.8) by the energy $E = \hbar\omega$ of the emitted photons and with $\alpha = e^2/(4\pi\epsilon_0\hbar c)$ we obtain the photon number spectrum:

$$\frac{d^2N}{dEdx} = \frac{\alpha}{\hbar c} z^2 \sin^2\theta_c = \frac{\alpha}{\hbar c} z^2 \left(1 - \frac{1}{\beta^2 n^2(\omega)}\right) \approx 370 \, \sin^2\theta_c/\mathrm{eV/cm} \quad (\text{for } z = 1)\,. \tag{11.11}$$

The wave length dependence[2] results from (11.11) with $\omega = 2\pi c/\lambda$:

$$\frac{d^2N}{d\lambda dx} = \frac{2\pi z^2\alpha}{\lambda^2}\left(1 - \frac{1}{\beta^2 n^2(\lambda)}\right) = \frac{2\pi z^2\alpha}{\lambda^2} \sin^2\theta_c(\lambda)\,. \tag{11.12}$$

Hence, as a function of frequency or photon energy $\hbar\omega$, the number spectrum is flat, while it has a $1/\lambda^2$ dependence as a function of wavelength. Integrating over x along a path of length L in the radiator, we obtain the number of emitted photons per wavelength interval $\mathrm{d}\lambda$:

$$\frac{dN}{d\lambda} = \frac{2\pi z^2\alpha}{\lambda^2} L \sin^2\theta_c\,. \tag{11.13}$$

For asymptotic particle energies ($\gamma \to \infty$, $\theta_c \to \theta_{max}$) and with (11.5) one obtains:

$$\frac{dN_\infty}{d\lambda} = \frac{2\pi z^2\alpha}{\lambda^2} L \frac{1}{\gamma_{th}^2}\,. \tag{11.14}$$

Figure 11.7 shows the spectrum as a function of wavelength. The region $\lambda \approx 100$–$200\,\mathrm{nm}$ is where the transition from normal to anomalous dispersion occurs, accompanied by absorption. The cut-off of the spectrum hence lies at small wavelengths (blue to UV).

[2]Going from (11.11) to (11.12) a minus sign is suppressed coming from the inverse dependence of the spectrum on λ compared to ω.

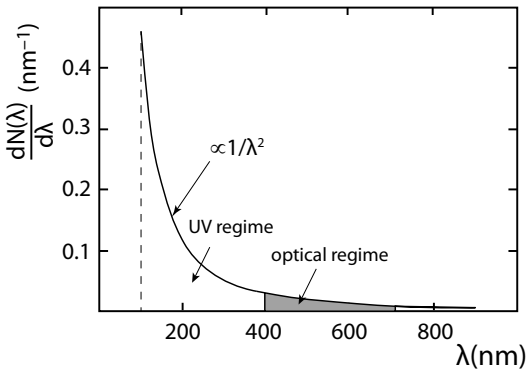


Fig. 11.7 Schematic representation of the wavelength dependence of the Cherenkov photon spectrum. The grey shaded region corresponds to the emission of optical photons ($\lambda \approx 400$–700 nm). At wave lengths around $\lambda = 100$–200 nm the regime of anomalous dispersion begins and absorption sets in which shuts the spectrum off (see text and fig. 11.6).

11.3 Detection of Cherenkov radiation

Cherenkov radiation is used in various ways for particle detection, ranging from the sheer detection of emitted Cherenkov radiation to the explicit reconstruction of Cherenkov rings. Detectors exploit the characteristics of Cherenkov radiation, in particular that radiation sets in above a velocity threshold, that the emission has velocity dependent intensity and direction and also that it is prompt, allowing for precise timing. This is in contrast, for example, to scintillation light which appears delayed and is emitted in all directions (see chapter 13).

In high energy physics experiments a prominent application of Cherenkov light is to distinguish different particle species from each other, for example to separate pions from electrons (see section 14.2.3). Different types of Cherenkov detectors can be distinguished according to whether they only aim to integrally detect the radiation or if also the explicit reconstruction of Cherenkov rings is pursued. Furthermore, Cherenkov radiation is sometimes used as a signal from charged particles in electromagnetic showers for energy measurement, for example in calorimeters using lead-glass crystals (described in section 15.5.2).

Because high energetic particles also produce Cherenkov radiation in transparent natural media such as water, ice and air, it is used in astroparticle physics experiments to detect particle showers initiated by charged cosmic rays, gamma rays, and neutrinos (see chapter 16). Cherenkov telescopes receiving Cherenkov light produced by cosmic gamma rays interacting with the atmosphere (and thereby generating electrons and positrons, section 16.5) as well as high energy neutrino detectors (section 16.6.5) exploit the prompt emission of Cherenkov radiation for time-critical measurements. Arrival time and distribution of the radiation provide information about its creation point and enter in the reconstruction of cosmic showers.

Photon yield. Integrating (11.13) over optical wavelenghts one obtains the number of emitted optical photons:

$$N_{opt} = \int_{optical} d\lambda \, \frac{dN}{d\lambda} = N_{\delta\lambda} \, z^2 \, L \, \sin^2 \theta_c \,, \tag{11.15}$$

where $N_{\delta\lambda}$ contains the integration over the $1/\lambda^2$ part of the spectrum. For optical wavelenghts between 400 nm and 700 nm it assumes the value

$$N_{\delta\lambda} = \int_{optical} \frac{2\pi\alpha}{\lambda^2}\, d\lambda = \ 491 \text{ photons/cm}\,. \tag{11.16}$$

Depending on the radiator medium the number of detectable Cherenkov photons can become very small, in particular for gas radiators featuring a small index of refraction n. Therefore, reconstruction of Cherenkov rings (fig. 11.3) is a real challenge for detector techniques (see section 11.6).

The selection of detector components (mirror, window, photon detector) must be optimised for maximum yield of detected photons. When using photomultipliers, for example, UV transparent entrance windows and UV sensitive cathodes are important. To determine the number of photons (which usually must first be converted into photoelectrons) the losses on their way from their production point to the detector as well as quantum and detection efficiency must be taken into account.

To obtain the number of detectable photoelectrons N_{pe} (11.15) must be corrected for losses:

$$\begin{aligned} N_{pe} \ &= \ 2\pi\alpha\, z^2\, L\, \sin^2\theta_c \int_{\lambda_1}^{\lambda_2} \frac{d\lambda}{\lambda^2}\ T(\lambda)\ Q(\lambda)\ R(\lambda) \\ &=: \ N_0\, z^2\, L\, \sin^2\theta_c\,, \end{aligned} \tag{11.17}$$

where

$$N_0 = 2\pi\alpha \int_{\lambda_1}^{\lambda_2} \frac{1}{\lambda^2}\, T(\lambda) Q(\lambda) R(\lambda)\, d\lambda\,. \tag{11.18}$$

Here $T(\lambda)$ is the transmission efficiency of the detector entrance window, $Q(\lambda)$ is the quantum efficiency of the photodetector and $R(\lambda)$ is the reflectivity of the mirrors used to focus the photons on the detector (see section 11.3). N_0 is regarded as a measure for the quality of Cherenkov detection (*figure of merit*) and is smaller then $N_{\delta\lambda}$ in (11.16) due to efficiency losses caused by T, Q and R.

With the approximation $\sin\theta_c \approx \theta_c$ and with (11.7) one obtains for the ratio of detected photons to those obtained at asymptotic velocities (N_∞):

$$\frac{N_\gamma}{N_\infty} \ \approx \ \frac{\theta_c^2}{\theta_{max}^2} \approx 1 - \frac{\gamma_{th}^2}{\gamma^2}\,. \tag{11.19}$$

Figure 11.8 illustrates the photon yield dependence on γ for two fluorocarbons, perfluorohexane (C_6F_{14}), which is liquid at room temperature (boiling point at 57 °C), and perfluoropentane (C_5F_{12}), which has a lower boiling point (28 °C) permitting the use (at a somewhat increased temperature) as a vapour with substantially smaller refractive index (see table 11.1).

Note that the energy loss caused by Cherenkov radiation of the radiating particle is included in the energy loss described by the Bethe–Bloch formula (see the corresponding discussions by Fermi [398] and in [84]), as (3.25) treats the energy lost and transferred to atoms by electromagnetic interactions in general terms. However, this contribution is small compared to the energy lost due to ionisation. In gases with $Z > 7$ the energy loss due to Cherenkov radiation always remains below 1% of the loss due to ionisation. For light gases like hydrogen or helium the relative proportion reaches up to 5% (see [489] and references therein). Compared with other mechanisms that also generate light, such as the scintillation process, the intensity of Cherenkov light,

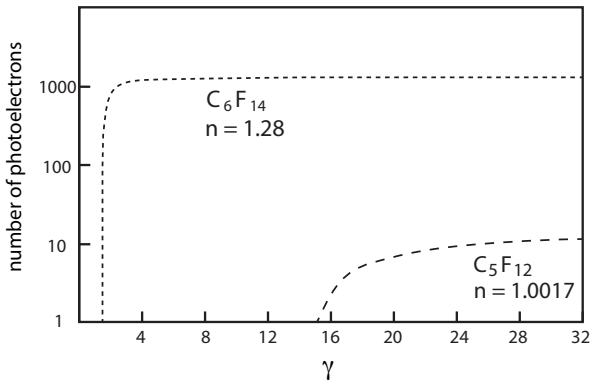


Fig. 11.8 Photoelectron yield for two fluorocarbons ($L = 20$ cm) having different refraction indices and a total detection efficiency $Q \times T \times R$ of 35% as a function of the Lorentz factor γ.

when quantified by the number of generated photons, typically is smaller by several orders of magnitude.

In the following sections we introduce different detector types for the detection of Cherenkov radiation. Hereby we distinguish those which are only sensitive to the total intensity of the Cherenkov radiation, hence being able to separate radiating from non-radiating particles (so-called 'threshold detectors'), and those explicitly detecting the emitted Cherenkov radiation on a cone appearing as a ring on a detector surface, as shown in fig. 11.23 (so-called 'ring imaging Cherenkov' or RICH detectors).

11.4 Threshold Cherenkov detector

To distinguish particles of different mass, the Cherenkov threshold detector employs the mass dependence of the particle velocity for a given energy and the fact that the threshold velocity $\beta_{th} = 1/n$ only depends on the index of refraction of the radiator. With these detectors an explicit measurement of Cherenkov rings is not intended, but only the integrated detection (or non-detection) of Cherenkov radiation. For a long time this technique was the only possibility of employing Cherenkov radiation for particle identification because spatially resolved detection of individual Cherenkov photons was considered too difficult. The threshold method requires much less technical effort. Figure 11.9 explains the principle. The generated Cherenkov light is diffusely reflected, eventually reaching a photosensitive detector (here a photomultiplier, PMT).

For a particle of mass m the threshold momentum, below which no Cherenkov radiation is emitted, is:

$$p_{th} = mc^2 \, (\beta\gamma)_{th} = \frac{mc^2}{\sqrt{n^2 - 1}} \, . \tag{11.20}$$

If the momentum of a particle is known, then by the choice of a suitable radiator medium with refractive index n, Cherenkov counters can be designed which are only sensitive to particles light enough so that their velocity exceeds the Cherenkov threshold. In such a case the momentum needs to be measured simultaneously, usually by deflection in a magnetic field. Often used radiator materials are listed in table 11.1, together with their refraction indices and thresholds γ_{th}.

From table 11.1 one recognises that solid or liquid radiators are dominantly suited to separate p, K, π with momenta from about 1 GeV/c to typically 10 GeV/c, whereas gas radiators are dominantly employed for particle separation at large Lorentz factors

Table 11.1 Table of some typical Cherenkov radiator materials with refractive index n [791] at a wavelength of 589.29 nm (average of the D-lines of sodium) unless indicated differently, the Cherenkov threshold γ_{th}, as well as the threshold momenta for kaons, pions and electrons. The fluorocarbon radiators used by DELPHI [45] and LHCb [51] are specified for $\lambda = 180$ nm [651]. Freon 114 is the trade name for $C_2Cl_2F_4$. STP = standard temperature and pressure, defined in section 2.5.2 on page 20.

Medium	Refractive index	γ_{th}	p^p_{th} (GeV/c)	p^K_{th} (GeV/c)	p^π_{th} (GeV/c)	p^e_{th} (GeV/c)
solid / liquid						
ice	1.310	1.55	1.1	0.58	0.16	6.0×10^{-4}
NaF	1.325	1.52	1.1	0.57	0.16	5.9×10^{-4}
MgF_2	1.374	1.46	1.0	0.52	0.15	5.4×10^{-4}
LiF_2	1.392	1.44	1.0	0.51	0.14	5.3×10^{-4}
CaF_2	1.433	1.40	0.9	0.48	0.14	5.0×10^{-4}
silica (quartz)	1.458	1.37	0.9	0.47	0.13	4.8×10^{-4}
fused silica (400 nm)	1.474	1.36	0.8	0.46	0.13	4.7×10^{-4}
borosilicate	1.474	1.36	0.8	0.46	0.13	4.7×10^{-4}
plexiglass	1.492	1.35	0.8	0.45	0.13	4.6×10^{-4}
plast. scint.	1.580	1.29	0.8	0.40	0.11	4.2×10^{-4}
lead glass	1.670	1.25	0.7	0.37	0.10	3.8×10^{-4}
NaI	1.775	1.21	0.6	0.34	0.10	3.5×10^{-4}
CsI	1.787	1.21	0.6	0.33	0.09	3.5×10^{-4}
C_6F_{14} (180 nm)	1.283	1.60	1.2	0.61	0.17	6.4×10^{-4}
water	1.333	1.51	1.1	0.56	0.16	5.8×10^{-4}
alcohol	1.361	1.47	1.0	0.53	0.15	5.5×10^{-4}
paraffine	1.444	0.69	1.39	0.9	0.47	4.9×10^{-4}
aerogel	1.25 –1.007	1.69 –8.50	1.3 –7.9	0.7 –4.2	0.19 –1.18	0.7–4.3 $\times 10^{-3}$
Gases (1 bar, 0°C (STP), C_5F_{12} at 40°C)						
He	1.000035	119.7	112.3	59.1	16.7	0.061
Ne	1.000066	87.0	81.6	43.0	12.1	0.044
H_2	1.000132	61.6	57.7	30.4	8.6	0.031
Ar	1.000282	42.1	39.5	20.8	5.9	0.022
air	1.000292	41.4	38.8	20.4	5.8	0.021
CH_4	1.000444	33.6	31.5	16.6	4.7	0.017
CO_2	1.000449	33.4	31.3	16.5	4.7	0.017
CF_4 (180 nm)	1.00053	30.7	28.8	15.2	4.3	0.016
freon 114	1.00140	18.9	17.7	9.3	2.6	0.010
C_4F_{10} (180 nm)	1.00150	18.3	17.1	9.0	2.5	0.009
C_5F_{12} (180 nm)	1.00172	16.2	15.1	8.0	2.3	0.008

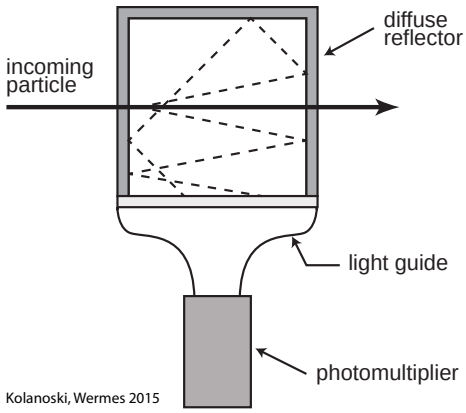


Fig. 11.9 Principle of a threshold Cherenkov counter with light detection after diffuse reflection.

γ, used for example if electrons need to be separated from hadrons. Aerogel is a light, nano-porous silica gel made of SiO_2 and water. It has a density dependent refractive index bridging the gap in n between solids/liquids and gases in the range around $n = 1.03$. The density of aerogel can be varied[3] at the production stage between about $0.03\,\mathrm{g/cm^3}$ and $0.55\ \mathrm{g/cm^3}$ and allows tuning the refractive index between $n = 1.007$ and $n = 1.11$ [602] according to $n = 1 + \alpha\rho$ (Gladstone-Dale relation) with $\alpha \approx 0.2 - 0.25\,\mathrm{cm^3/g}$ (see also the discussion about chromatic dispersion in section 11.6.2 on page 464.

In order to increase the efficiency for collecting the generally not very large number of photons, mirrors are often used to direct the Cherenkov light onto suitably positioned photodetectors. Figure 11.10 shows an example of three threshold Cherenkov counters arranged one behind another for particle identification in the TASSO experiment [263].[4] In the first counter (aerogel) diffusely reflected light is detected by PMTs; in the subsequent counters Cherenkov light is concentrated on PMTs by mirror reflection. To be identified a particle has to cross all three threshold counters with different refraction indices (aerogel, freon, CO_2). From a combination of the counter responses the particle species can be determined. Table 11.2 shows the respective momentum thresholds for pions, kaons, and protons. With this system pions could be identified (i.e. discriminated against K and p) over the entire momentum range occurring in the experiment, and kaons for momenta below $5.5\,\mathrm{GeV}/c$ or above $9.5\,\mathrm{GeV}/c$. For momenta between $5.5\,\mathrm{GeV}/c$ and $9.5\,\mathrm{GeV}/c$ kaons can be confused with protons [263] (see also section 14.2.3 on particle identification).

The following numerical example illustrates some basic considerations when designing a threshold Cherenkov detector. Using (11.17) and (11.16) the number of generated (optical) photoelectrons from a singly ($z = 1$) charged particle in a threshold detector is approximately:

$$N_{pe} \approx 491/\mathrm{cm}\ \ Q \times T \times R\ \ L \sin^2\theta_c \approx 150/\mathrm{cm}\ L \sin^2\theta_\mathrm{c}\,, \tag{11.21}$$

[3] Densities can reach values even below those of air (vacuum aerogel, e.g. $\rho = 0.009\,\mathrm{g/cm^3}$ [930]). The total range in n is given in table 11.1.

[4] The TASSO detector [245] was operated in the 1980s at the e^+e^- storage ring PETRA at DESY, Hamburg. It is mentioned here because of the combination of three threshold Cherenkov counters for particle identification.

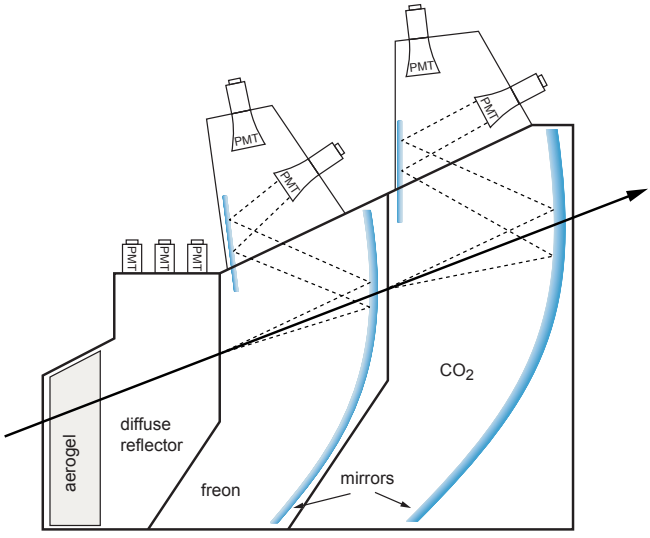


Fig. 11.10 Cherenkov detector arrangement used in the TASSO experiment (adapted from [263, 245], with kind permission of Elsevier). Shown are three successive threshold Cherenkov detectors with different Cherenkov thresholds being traversed by charged particles.

Table 11.2 Threshold momenta of the TASSO Cherenkov detector system [263].

Radiator medium	Refractive index	Momentum threshold (GeV/c) pion	kaon	proton
Aerogel	1.025	0.6	2.2	4.2
Freon 114	1.0014	2.7	9.4	17.8
CO_2	1.00043	4.8	16.9	32.0

where L is the length of the radiator and the efficiencies Q, T and R are understood as averages over the wavelength range; typical values of 40%, 80% and close to 100% have been assumed in (11.21), respectively.

In the following we want to estimate the minimum requirements on the index of refraction n and on the length L of a radiator to separate two particles having the same momentum $p_1 = p_2$, but different masses $m_1 < m_2$. We choose the radiator medium such that particle 2 falls just below the Cherenkov threshold and hence does not radiate, that is, $\gamma_2 = \gamma_{th}$, $\sin^2\theta_2 = 0$, $\beta_2 = 1/n$. Particle 1 has velocities above the Cherenkov threshold β_{th} and hence according to (11.15) radiates a number of photons $N_\gamma(1)$ in the visible spectral range:

$$\begin{aligned} N_\gamma(1) &= N_0\, L \sin^2\theta_c = N_0\, L \left(1 - \frac{1}{\beta_1^2 n^2}\right) = N_0\, L \left(1 - \frac{E_1^2}{p^2c^2n^2}\right) \\ &= N_0\, L \left(1 - \frac{m_1^2c^4 + p^2c^2}{p^2c^2n^2}\right) = N_0 L \left(\frac{n^2-1}{n^2} - \frac{m_1^2c^4}{p^2c^2n^2}\right) \\ &= N_0 L \left(\frac{m_2^2c^2}{p^2n^2} - \frac{m_1^2c^2}{p^2n^2}\right) = N_0 L\, \frac{\Delta m^2 c^2}{p^2n^2}\,, \end{aligned} \tag{11.22}$$

where $\beta_1^2 = p^2c^2/E_1^2$, $\Delta m^2 = m_2^2 - m_1^2$, and $n^2 - 1 = m_2^2c^2/p^2$ according to (11.20) have been used.

Let's assume that at least 10 photons should be detected.[5] With (11.21) and (11.22) we then find

$$L/\mathrm{cm} > \frac{p^2 n^2}{15\,\Delta m^2 c^2}\,.$$

For example, in order to separate electrons ($m_1 = 0.511\,\mathrm{MeV}$) from pions ($m_2 = 140\,\mathrm{MeV}$) at a momentum of 10 GeV, a radiator length of at least

$$L = \frac{(10\,\mathrm{GeV})^2\, n^2}{15 \cdot 0.019\,\mathrm{GeV}^2} = 3.5\,\mathrm{m} \cdot n^2$$

is needed. Pions with a momentum of 10 GeV have a Lorentz factor of

$$\gamma_2 = \frac{\sqrt{p^2c^2 + m_2^2 c^4}}{m_2 c^2} \approx 72$$

such that with (11.4) and $\gamma_2 < \gamma_{th}$ an index of refraction $n < 1.0002$ is found to be suited for the given task, which can be fulfilled by a gas radiator like helium or hydrogen (see table 11.1).

Gases have the additional advantage over solid or liquid radiators that the refractive index n can be changed by changing the pressure in the radiator volume ($n - 1 \propto p$). Thus particle separation can be tuned for a range of momenta by selecting a specific radiator pressure. This method is used, for example, to select certain particle species in so-called 'secondary beams' containing various stable particles like e, π, K and p. Secondary beams are produced off a target by a primary high energy beam, for hadron production usually a proton beam (see also the description on page 12). In charged particle beams, secondary particle momenta are then selected by magnets behind the target.

Resolution of threshold-type Cherenkov detectors. While a threshold counter outputs a yes/no response, one can still ask what resolution on β one can obtain by a measurement of the number of detected Cherenkov photons, that is, photoelectrons. Their number is related to θ_c by (11.17):

$$\sigma_N = \frac{\partial N}{\partial \theta_c}\sigma_{\theta_c} = \frac{2N}{\tan\theta_c}\sigma_{\theta_c}\,, \tag{11.23}$$

where σ_{θ_c} is the spread in θ_c which varies with the particle velocity β. In turn, the β resolution then follows using (11.1) as

$$\frac{\sigma_\beta}{\beta} = \tan\theta_c\,\sigma_{\theta_c} = \tan^2\theta_c\,\frac{\sigma_N}{2N} = \frac{\tan^2\theta_c}{2\sqrt{N}}\,, \tag{11.24}$$

where $\sigma_N = \sqrt{N}$ has been assumed. This relation will be compared in section 11.6.2, page 461 to the resolution obtained by explicitly detecting the Cherenkov ring.

11.5 Differential Cherenkov detector

The so-called differential Cherenkov detector exploits the emission angle of the radiation to increase the discrimination power of the detection without explicitly re-

[5] Such a small number is typical for gas radiators.

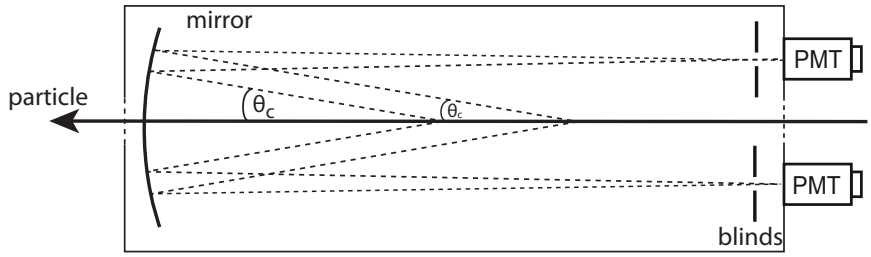


Fig. 11.11 Principle of the differential Cherenkov counter. A spherical mirror focuses the Cherenkov light emitted under θ_c onto a photon detector covered by (ring-shaped) blinds. Due to the blinds only particles having a corresponding ring size generate a signal in the photon detector.

constructing Cherenkov rings. It is hence a precursor stage for *ring imaging* detectors described in the next section, but otherwise has characteristic features similar to threshold Cherenkov counters regarding construction and readout.

Cherenkov light emitted from a particle under the angle θ_c is focused by a mirror onto the photon detector as a ring image with radius $R = f_S \tan\theta_c$ (f_S = focal length), as shown in fig. 11.11. Using ring-shaped blinds in the focal plane only a small area of the emission angle is selected corresponding to a particular velocity of the radiating particle. By changing the radius of the blinds different emission angles θ_c and hence particle velocities β can be covered.

11.6 Ring imaging Cherenkov detector (RICH)

A complete reconstruction of Cherenkov rings is accomplished in so-called *ring imaging Cherenkov* (RICH) counters or RICH detectors. The development of this very ambitious detector concept has its origin in the works of Ypsilantis and Seguinot [883]. Photons emitted under the Cherenkov angle $\theta_c = \arccos(1/\beta n)$ along the particle's flight path are mapped onto the focal plane of a focusing optical system where a photosensitive detector is placed. Figure 11.12 illustrates this RICH principle.
Two focusing methods are used:

- *Mirror focusing* (fig. 11.12(a)): mirror focuses parallel light rays, emitted (under the same azimuthal angle) from different points along the particle trajectory, onto a point on a circle in the focal plane where the photons are detected.
- *Proximity focusing* (fig. 11.12(b)): here the radiator is thin—in general a solid or a liquid—such that the Cherenkov radiation generated in the radiator along the particle trajectory can be regarded to approximately come from the same point in space (pinhole camera principle). The projection of the radiation onto the detector surface represents a ring, also without focusing elements. Its width depends primarily on the thickness of the radiator and the distance to the detector (see fig. 11.18 on page 462).

The ring resolution of the detector, that is, in particular the influence of the granularity of the readout electrodes, should be smaller than the ring width provided by the image.

For both focusing methods the radius R of a Cherenkov ring is approximately

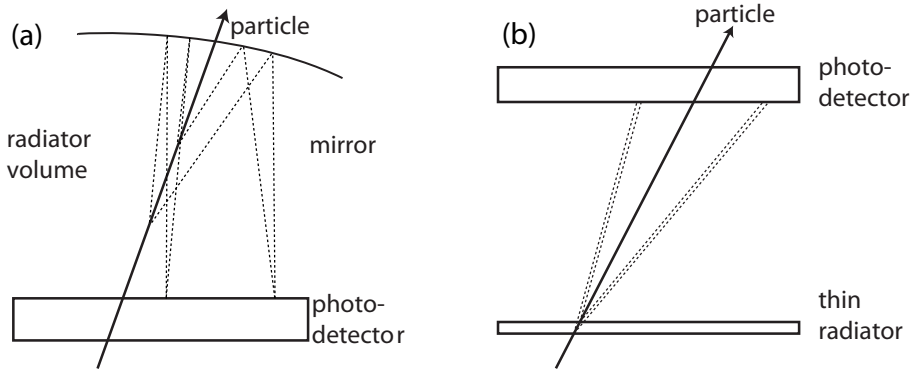


Fig. 11.12 Detection of Cherenkov rings with RICH detectors. (a) Focusing of Cherenkov radiation by a mirror onto a photon detector, (b) thin radiator with *proximity focusing* for which the image merely is a cross-sectional cut through the Cherenkov cone.

$$R \approx d\,\theta_c \approx d\sqrt{1 - \frac{1}{\beta^2\, n^2}}\,, \tag{11.25}$$

where d is the mirror focal length[6] f_m or the distance of the thin[7] radiator to the photodetector, respectively.

From this the particle velocity can be determined using (11.1). For a given momentum

$$p = \beta\gamma mc \tag{11.26}$$

one obtains the mass of the particle:

$$m^2 = \frac{p^2}{c^2}\left(\frac{1}{\beta^2} - 1\right) = \frac{p^2}{c^2}\left(n^2\cos^2\theta_{\mathrm{c}} - 1\right)\,. \tag{11.27}$$

In many experiments p is measured in a magnetic spectrometer and the measurement of the Cherenkov ring is used for particle identification following (11.27). In (11.27) the squared mass m^2 is chosen as in measurements it can in principle also become negative.

Both ring imaging and proximity focusing techniques can be combined using the same photodetector to obtain for example π/K separation by a liquid radiator as well as e/π separation by means of a gas radiator. Figure 11.13 illustrates this method.

11.6.1 Photon detectors for Cherenkov rings

Detecting Cherenkov rings in a RICH counter is a very challenging task. Often—in particular for gas radiators—only very few Cherenkov photons with wavelengths in the visible to UV region are created. The positions of these photons must be determined—generally by converting them first into photoelectrons—with a precision sufficient for the reconstruction of a ring with radius R which can be assigned to a specific particle. The ring reconstruction is substantially easier if the direction and momentum of the particle track is measured (e.g. by a magnetic spectrometer) where the ring centre can

[6]For a spherical mirror $f_m = R_m/2$ and $\theta_c = 2R/R_m$ with R_m being the sphere radius.

[7]Thin compared to the distance d.

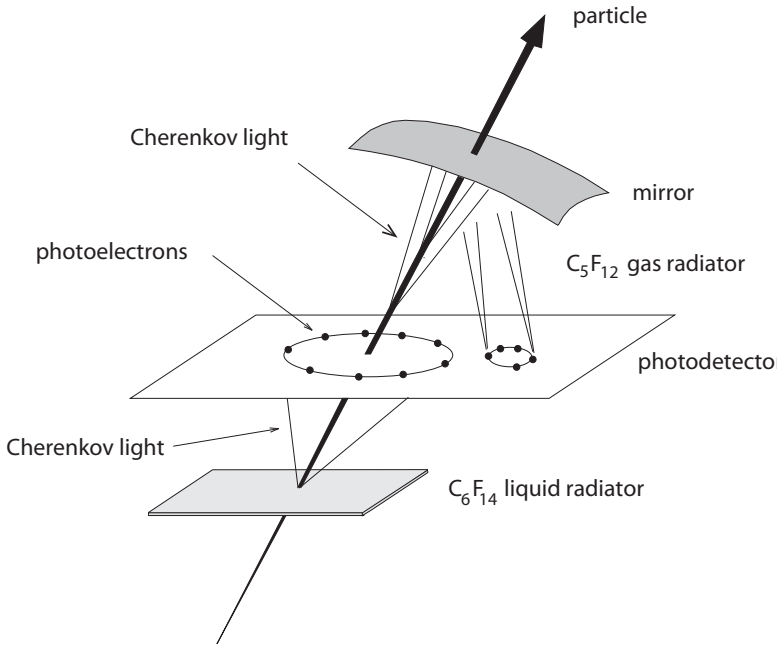
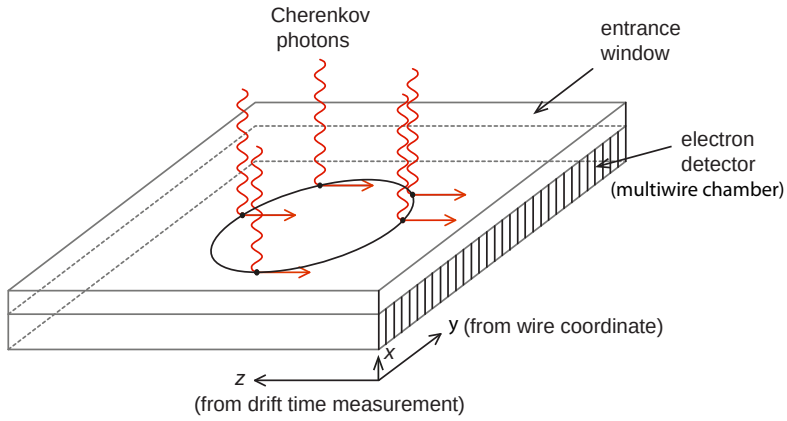


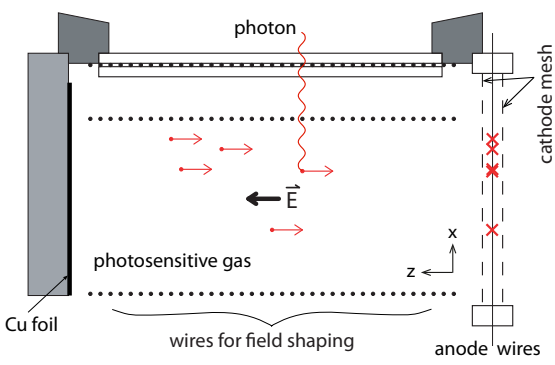
Fig. 11.13 Double RICH system: schematic representation of a RICH detector with two radiators (liquid, gas). The Cherenkov rings from both radiators are detected by a photon detector between them (adapted from [347]). This technique has been used in the DELPHI experiment at LEP [45] (section 11.6.3).

already be determined from track parameters of the particle in the Cherenkov radiator. Only one additional point on the circumference is needed to construct a circle. More points allow improvements in pattern recognition and in reconstructed ring resolution.

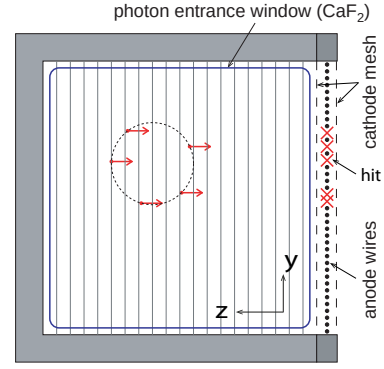
The photon detectors hence must detect single photons with spatial sensitivity. Experimentally this can be achieved by converting the Cherenkov photons into photoelectrons and measuring their creation point, for example by using a gas-filled wire chamber with a photosensitive vapour additive (fig. 11.14(a)) or by using cathode planes covered by a photosensitive layer (e.g. CsI, fig. 11.16). Also, photomultipliers, having entrance windows large compared to typical chamber resolutions, can be used if the rings are expanded to sufficiently large sizes (for an example see fig. 11.27). Which method to use depends on the application in an experiment; in particular it depends on particle rate and radiation tolerances of the photon detectors and the required ring resolutions. In wire chambers with photosensitive vapour additives a limiting factor was often radiation damage. The rapid development of semiconductor detectors and VLSI electronics (see section 17.6) has led to measurement techniques that allow an easier adaptation of the readout technique to the premises of the experiment. Examples are the introduction of multichannel photomultipliers or the development of semiconductor-based photodetectors like HPDs and HAPDs (sections 10.4.1 and 10.4.2) or silicon photomultipliers (section 10.5). Here we discuss some of the detection techniques, illustrating the multitude of detection methods of Cherenkov photons and the recognition of Cherenkov rings.



(a) Sketch of the principle.



(b) Side view.



(c) Top view.

Fig. 11.14 Detection of Cherenkov rings using a gas-filled multiwire proportional chamber (TPC principle) (from [158] with kind permission of Elsevier): (a) sketch of the principle showing the absorption of the Cherenkov photons and the drift of the electrons towards the readout plane; (b) side view ($z - x$), and (c) top view ($z - y$). This detection technique is regarded as the breakthrough for Cherenkov ring imaging by detecting single photons. The hits in (b) are registered on different wires of the chamber, located behind each other in y; the crosses indicate the positions of the electron arrival positions. The photon hits on the ring in (c) are reconstructed from drift time (z) and wire hit (y). The gas fillings are hydrocarbon mixtures (e.g. 75% methane, 25% ethane) with admixtures of a photon absorbing vapour (e.g. TMAE or TEA, see text) that converts the Cherenkov photons into photoelectrons.

The experimental challenge of reconstructing a Cherenkov ring from only few spatially resolved photon signals is sketched in fig. 11.14. The photon detector [158] is a gas volume in which converted photons are detected by means of a multiwire chamber readout at one side. The principle corresponds to that of a time projection chamber (TPC, see section 7.10.10 and fig. 7.33). Photoelectrons created by the photoelectric effect drift through the cathode grid to the anode plane of the chamber where they are gas-amplified near the anode wires and detected. The time of arrival of the electrons at the wire relative to the time of creation of the radiating particle (at collider experiments this is the bunch collision time) is measured. Arrival time and wire number

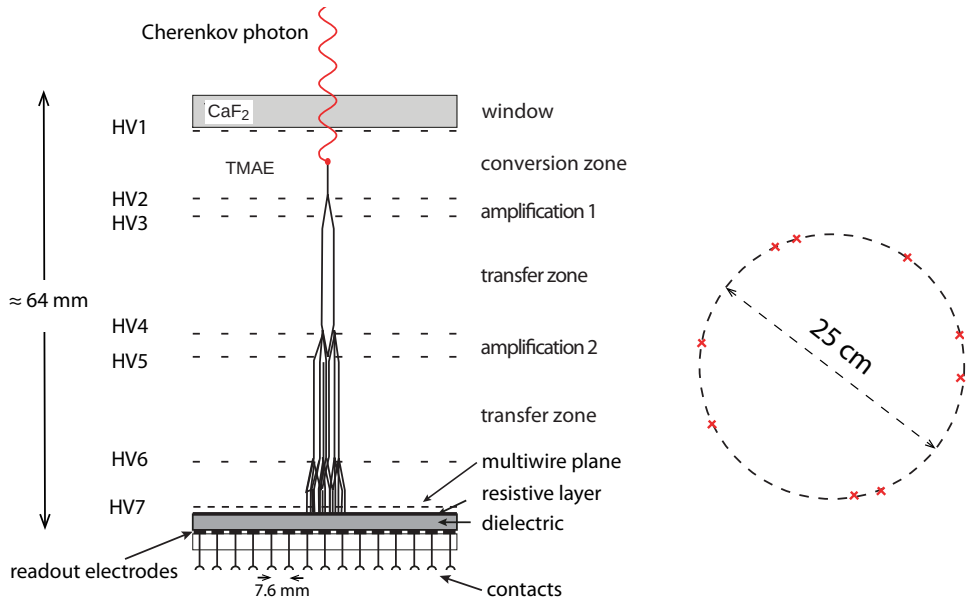


Fig. 11.15 Principle of a Cherenkov photon detector in a parallel plate configuration as used in the CERES/NA45 experiment (adapted from [162], with kind permission of Elsevier). A gas radiator (CH_4) and the mirror system lie outside the figure (see fig. 11.22). The detectors are filled by a gas mixture of He/C_2H_6 plus TMAE vapour. In general the Cherenkov photons are absorbed within the conversion volume. Gas amplification takes place in three gaps between wire grid planes (HV2-HV3 and HV4-HV5) and at a final multiwire plane. On the right: a typical Cherenkov ring reconstructed from a few photons only (see also fig. 11.23).

yield the z- and x-coordinates of the Cherenkov hit, respectively. From several hits a circle can be reconstructed (usually off-line), whose size corresponds to the particle velocity (according to (11.25)). As electron detector any position sensitive detector (in one or two dimensions) can be chosen, for example wire chambers or micro-structured gaseous detectors (section 7.9).

Figure 11.15 shows a gas-based photon detector in a configuration with parallel wire meshes, used in the heavy ion experiment CERES/NA45 (see fig. 11.22 and section 11.6.3) at the CERN SPS [162]. A Cherenkov photon passes the UV transparent window (e.g. quartz or CaF_2) entering a volume filled by a gas mixture containing He + 6% C_2H_6 + TMAE vapour (discussed on page 460). With a certain probability the photon is absorbed in the conversion zone generating an electron. The electron drifts towards the anode traversing several high field amplification zones between wire meshes separated by transfer zones with lower field strengths (see figure). A wire plane at the end of the electrons' paths is the third amplification stage. The electron avalanche created in this final stage is detected with two-dimensional resolution by recording the signals induced on a pad-segmented anode plane, separated from the wire plane by a dielectric. The pad electrodes are connected to integrated electronics circuits for readout. The multi-stage amplification scheme with in-between transfer zones reduces the probability for avalanches caused by photons created in the amplification process that would else generate further avalanches (photon feedback).

An example for a photon conversion by means of a CsI-coated multiwire chamber

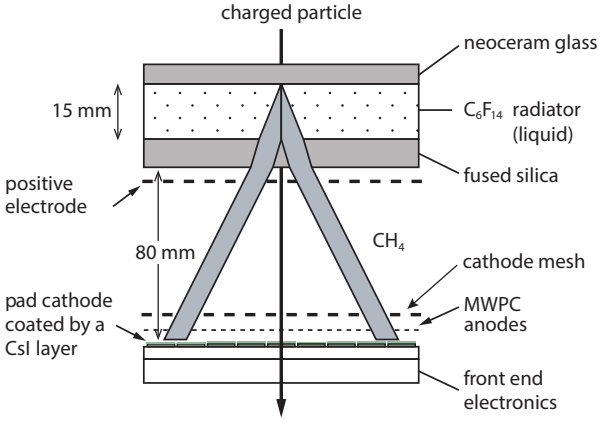


Fig. 11.16 RICH photon detector concept of the ALICE experiment (called HMPID, see also section 14.2.3). Conversion of Cherenkov photons into electrons is achieved by using a CsI-coated cathode of a multiwire chamber (MWPC) [323] (see text for more details).

cathode is shown in fig. 11.16, as used for example in the RICH detector of the heavy ion experiment ALICE at the LHC [323] (see also section 14.2.3). Cherenkov photons generated in the liquid radiator are mapped onto the segmented cathode covered by a 300 nm thick CsI film (*proximity focusing*). CsI has a comparatively large quantum efficiency Q in the UV range of about 25% at 175 nm [244] and has proven to be well suited for cathode coating. The photoelectrons created in the CsI layer drift back to the anode wires and generate a detectable gas-amplified signal (see section 7.4.1) in the CH_4 gas of the wire chamber. The efficiency for photoelectron detection is in the range 80–90%. On average about 14 photons are detected for $\beta \approx 1$ particles. The induced signal on the two-dimensionally structured lower cathode (8×8.4 mm^2 sized pads) allows determination of the coordinates of the signal. The purpose of the positive electrode near the radiator exit window is to suck off electrons coming from unwanted ionisation from background sources occurring in the volume of proximity focusing. They would otherwise be gas-amplified as well in the chamber volume.

Optimisation of photon detection. The interplay of parameters influencing the yield for Cherenkov photon detection is graphically represented in fig. 11.17. Quantum efficiencies of photon converting gas admixtures and transmissions of entrance windows and radiators are plotted as a function of the photon energy. Included are the energy thresholds for photon-converting vapours like TMAE and TEA (described below). Figure 11.17(a) shows the quantum efficiency (Q) for TMAE, TEA and CsI (left hand scale) together with the photon transmission properties (T, right-hand scale) of quartz and CaF_2 window materials. Figures 11.17(b) and (c) illustrate the conditions in which Cherenkov photons pass through radiator and detector window in order to be absorbed by TMAE or TEA: (b) is for gas and liquid radiators and (c) for solid radiators. Transparency ranges are given as horizontal lines drawn at the respective Cherenkov thresholds γ_{th} for the different radiators. Those materials having sufficient photon transparency to be used with TEA are marked by an arrow at the end, others by a bullet. Also shown are the energy regions for detectable photons using TEA in methane as converting vapour (assuming a transparent entrance window), compared to a TMAE admixture with quartz as entrance window.

The most important detector characteristics to be optimised for efficient Cherenkov photon detection are the following:

Photon conversion: For many applications the number of Cherenkov photons and the

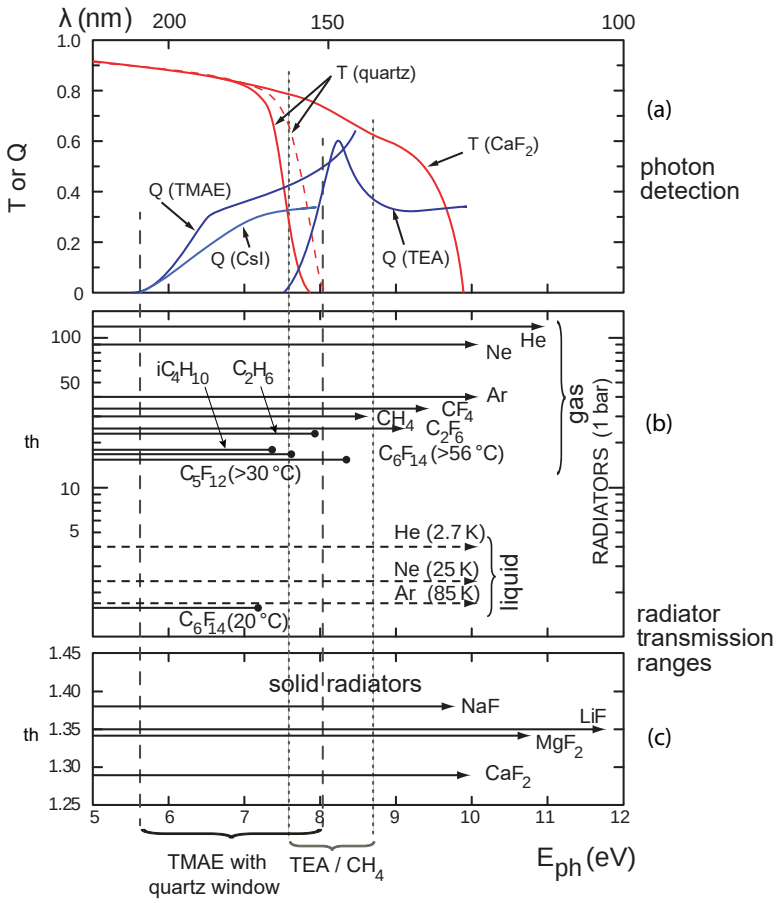


Fig. 11.17 (a) Quantum efficiencies (Q) for TMAE, TEA and CsI [879] as well as transmission (T) for photons using quartz or CaF_2 windows as a function of photon energy. Parts (b) and (c) show the transparency region of various radiator materials: (b) for gas- and liquid radiators, (c) for solids. Shown is the Lorentz factor at γ_{th} for various radiators, drawn as a line the length of which indicates the transparency range in this radiator (γ_{th} is calculated using n at $\lambda = 150\,\text{nm}$). Those materials having sufficient photon transparency to be used with TEA, are marked by an arrow at the end, others by a bullet. The region framed by the dotted vertical lines is that of detectable photons when using TEA in CH_4 as converting vapour (with an assumed transparent ($T = 1$) window). The region framed by the dashed lines corresponds to using TMAE with quartz as entrance window (adapted from [674]). Note that CaF_2 and other crystals are solid radiators, but can also be used as entrance windows with high UV transparency.

resulting number of photoelectrons (see eq. (11.17)) is very small. The product of the wavelength dependent quantities T, Q and R often is not larger than 30%. The quantum efficiency $Q(\lambda)$ of the photon absorbing vapours sets in only at relatively high photon energies (for example at 5.4 eV for TMAE) and stays substantially below 50% in the spectral range transparent for quartz windows (see fig. 11.17). Cathodes coated for example with CsI reach quantum efficiencies of 20–35% [879]

for wavelengths below 185 nm. Photon detectors therefore often see only very few photons for ring reconstruction, typically about 10–20. To extend the usage of the Cherenkov spectrum more deeply into the UV range, particular care over the material choice must be taken (window transparency, mirror reflectivity). If the photon conversion takes place in a gas-filled detector, special molecule vapours are used, such as tetrakis-(dimethylamino)-ethylene (TMAE = $C_2(N(CH_3)_2)_4$) or triethylamine (TEA = $N(CH_2CH_3)_3$), having a particularly large cross section even at low energies for the photoeffect reaction $\gamma + M \to M^{+*} + e^-$ (M^{+*} denoting M excited and ionised). The corresponding threshold energies are $E_{th} = 7.5$ eV (165 nm) for TEA and $E_{th} = 5.4$ eV (230 nm) for TMAE (see also fig. 11.17). TMAE is preferred over TEA as a conversion vapour mainly because it can be used together with a quartz entrance window instead of a single crystal window like CaF_2 or NaF.

Radiator refractive index: According to (11.15), the number of emitted photons per radiator length is proportional to $\sin^2\theta_c$:

$$\frac{dN}{dx} \propto \sin^2\theta_c \approx 1 - \frac{1}{n^2}\,. \tag{11.28}$$

Radiators with large index of refraction n yield more photons than those with small n, their Cherenkov threshold is lower and the ring radius is larger at the same distance, but saturates earlier (fig. 11.8). However, for the separation of different particle types by measurement of Cherenkov ring radii over some momentum range, the choice of n should be chosen such that the dynamic range of rings occurring over the corresponding γ-range is large (see fig. 11.4 and section 14.2.3). For high momenta in the multi-GeV range or when electrons are to be identified while suppressing hadrons or muons, high thresholds γ_{th} are needed corresponding to low refractive indices as provided by gases. Often geometrical constraints imposed by the overall layout of an experiment limit the volume available for a Cherenkov detector and hence also impacts the choice of n. In collider experiments space restrictions often impact the implementation of voluminous, low n gas radiators more than in fixed-target experiments.

Dispersion: The index of refraction n of a medium depends on the wavelength of the radiation (dispersion), leading to chromatic aberrations. To keep these as small as possible one can in principle restrict the wavelength range (for example by the choice of the photocathode and its wavelength sensitivity), however this also reduces the number of photons available for detection. The influence of chromatic aberration on the resolution of Cherenkov detectors is discussed in the next section.

11.6.2 Resolution of RICH detectors

The RICH resolution on β is inferred from (11.1) to be

$$\sigma_\beta = \frac{\partial\beta}{\partial\theta_c}\sigma_{\theta_c} \Rightarrow \frac{\sigma_\beta}{\beta} = \tan\theta_c\,\sigma_{\theta_c} = \tan\theta_c \frac{\sigma_{\theta_i}}{\sqrt{N}}\,, \tag{11.29}$$

where σ_{θ_i} denotes the angular error per photoelectron, whereas $\sigma_{\theta_c} = \sigma_{\theta_i}/\sqrt{N}$ is the Cherenkov angle resolution for N photoelectrons.

Compared to the resolution of a threshold counter given in (11.24), the ratio

$$\left(\frac{\sigma_\beta}{\beta}\right)_{\text{RICH}} \Big/ \left(\frac{\sigma_\beta}{\beta}\right)_{\text{Thr}} = \frac{2\sigma_{\theta_i}}{\tan\theta_c} \tag{11.30}$$

can become as small as 1/250 or smaller depending on wavelength range and chromatic radiator properties [1016, 1015]. This is due to the fact that in threshold counters the resolution infers from N, whereas a RICH detector directly measures θ_c.

Particle identification (particle ID) by Cherenkov radiation is the determination of a particle's mass at a given momentum by measuring its velocity through a measurement of the angle under which Cherenkov radiation has been emitted (see also chapter 14 on particle ID). The resolution of this measurement can hence be written as:

$$\begin{aligned}\frac{\sigma_m}{m} &= \frac{1}{2}\frac{\sigma_{m^2}}{m^2} = \gamma^2\frac{\sigma_\beta}{\beta} = \frac{\gamma^2}{\beta}\left[\left(\frac{\partial\beta}{\partial\theta_c}\,\sigma_{\theta_c}\right)^2 + \left(\frac{\partial\beta}{\partial n}\,\sigma_n\right)^2\right]^{\frac{1}{2}} \\ &= \gamma^2\left(\tan\theta_c\,\sigma_{\theta_c} \;\oplus\; \frac{1}{n}\,\sigma_{\text{n}}\right) = \left(\frac{p}{m}\right)^2\frac{1}{\beta^2}\left(k_R \;\oplus\; \frac{1}{n}\,\sigma_{\text{n}}\right),\end{aligned} \tag{11.31}$$

where $\oplus$ denotes quadratic addition of terms. The first term k_R, called the 'RICH constant', is the contribution to the measurement error due to the angular resolution

$$k_R = \left(\frac{\sigma_\beta}{\beta}\right)_{\text{angle}} = \tan\theta_c\,\frac{\sigma_{\theta_i}}{\sqrt{N}} = \frac{n\beta\,\sigma_{\theta_i}}{z\sqrt{N_0\,L}} \tag{11.32}$$

using (11.17) to express N, together with (11.1). The second term in (11.31) takes variations of the refractive index with wavelength into account leading to a smearing of the Cherenkov ring that does not depend on the velocity of the radiating particle, but rather on the energy of the detected Cherenkov photon (dispersive or chromatic aberration).

Particle separation. The capability of a RICH detector to separate two particles with the same momentum p and with masses m_1 and $m_2 > m_1$ from each other with at least n_σ standard deviations can be defined by the difference of the measured quantities (here $m^2_{1,2}$) divided by the resolution. Using the inverse expressions of (11.31) this is:

$$\frac{m_2^2 - m_1^2}{\sigma_{m^2}} = \frac{\beta_2^2 m_2^2 - \beta_1^2 m_1^2}{2\,p^2\,\left(k_R \oplus \frac{1}{n}\sigma_n\right)} > n_\sigma\,. \tag{11.33}$$

Separation of m_1 and m_2 with n_σ hence is possible for momenta up to

$$p_{max} \approx \sqrt{\frac{m_2^2 - m_1^2}{2\,n_\sigma\,\left(k_R \oplus \frac{1}{n}\,\sigma_n\right)}}\,, \tag{11.34}$$

where $\beta_1 \approx \beta_2 \approx 1$ has been assumed.

Ring resolution. The precision on a θ_c measurement is influenced by track distortions during the generation of the radiation, the quality of the imaging system, as well as the spatial resolution of the detection system. For most RICH systems the latter two (imaging and photon detector resolution) are dominant. Good resolution is particularly important for Cherenkov measurements at high γ/γ_{th} values, at which the relative change in θ_c is small (see fig. 11.4).

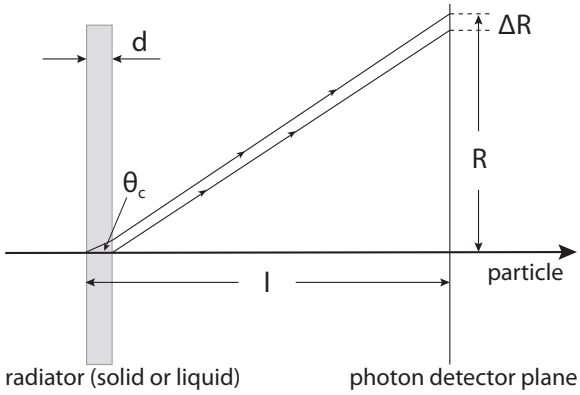


Fig. 11.18 Imaging of Cherenkov radiation by *proximity focusing* using a thin radiator (solid plate or volume with liquid). The resolution is given by the ratio of radiator thickness d and image distance l.

In dense media, solids or liquids, in which the number of generated Cherenkov photons can be sufficiently large even for thin radiators, proximity focusing is most often used. The achievable resolution in a proximity focusing system is mainly governed by the ratio of the radiator thickness d and the distance to the photon detector l, as shown in fig. 11.18:

$$\frac{\sigma_{\theta_c}}{\theta_c} = \frac{\Delta R}{R} = d/l\,. \tag{11.35}$$

In typical applications with gaseous radiators in general spacious volumes and focusing optics (mirrors) as in fig. 11.12 are employed to achieve sufficient photon yield. The radiation emitted from every point along a particle's track in the radiator leads to a ring image on the photon detector surface (see fig. 11.3). The detector must be aligned as precisely as possible in the focal plane of the mirrors. To first approximation the ring centre only depends on the particle direction and not on the emission point.

The quality of the imaging system and that of the photon detector determine the obtainable angular resolution, which is impacted by [1015, 464]:

- optical imperfections: mirror quality, spherical aberrations, flat detectors approximating curved focal planes,
- the quality of the detector readout (granularity, space point measurement),
- multiple scattering of the radiating particle,
- bending of tracks in a magnetic field.

Multiple scattering and bending in a magnetic field. Both multiple scattering as well as magnetic deflection of the particle when passing a radiator of thickness d cause a smearing σ_θ of the observed Cherenkov angle by amounts

$$\sigma_{\theta_{\mathrm{MS}}} = \frac{1}{\sqrt{6}}\,\frac{13.6\,\mathrm{MeV}/c}{\beta p}\,\sqrt{\frac{d}{X_0}}\,, \tag{11.36}$$

$$\sigma_{p_{\mathrm{defl}}} = \frac{2}{\pi}\,\frac{1}{\sqrt{12}}\,\frac{0.3\,\mathrm{GeV}/c}{p}\,\frac{Bd}{\mathrm{Tm}}\,, \tag{11.37}$$

resulting in a total deviation $\sigma_\theta = (\sigma^2_{\theta_{\mathrm{MS}}} + \sigma^2_{\theta_{\mathrm{defl}}})^{1/2}$. Here X_0 is the radiation length, B the magnetic field (in tesla), and p the particle momentum.

Expression (11.36) relates to the definition of the multiple scattering angle θ_{ms} in (3.102). The additional factor $1/\sqrt{6}$ contains two contributions: (a) the average θ_{MS} along the particle path differs from θ_{ms} for the full thickness d. The error-relevant spread of θ_{MS} averaged over the thickness d yields a factor $\sqrt{1/2}$; (b) another factor of $\sqrt{1/3}$ appears because the average plane angle $\langle\psi_{plane}\rangle$ inside the radiator, defined in (3.99) and fig. 3.32, rather than θ_{ms} at the radiator exit, is the relevant angle for this estimate (see also [108]).

Expression (11.37) is obtained using the magnetic deflection of a particle of momentum p transverse to B over a distance d in (9.36):

$$\alpha = \frac{d}{R} = \frac{0.3\,B\,d}{p}\,, \tag{11.38}$$

where R is the track's curvature radius in m and p and B are assumed in GeV/c and tesla units, respectively, following (9.8). The created Cherenkov ring is continuously shifted (smeared) during this bend with maximum shift angle $\alpha_{\max}$ in the horizontal plane, whereas in the vertical plane (no bending) there is no shift. In between these extremes the azimuthal dependence of this shift is $\alpha_{\max}\cos\phi$. The ϕ-averaged shift then becomes $\frac{2}{\pi}\alpha = \frac{2}{\pi}\frac{d}{R}$. Since the shifting from 0 to α_{max} occurs uniformly during the bend the rms error contribution to θ_c is a factor $1/\sqrt{12}$ smaller than the maximum shift.

To estimate the relative importance of the bending effect we take as an example the parameters of the DELPHI RICH detector (see section 11.6.3) with a magnetic field of $B = 1.2$ T and $d = 0.45$ m leading to a deflection angle α of about 0.46° for a momentum of 20 GeV/c. The calculated smearing due to the deflection is 1.5 mrad per photon hit, to be compared to the measured resolution of $\sigma_{\theta_i} = 4.6$ mrad [72].

Chromatic dispersion. The second contribution to the mass resolution in (11.31) is the chromatic error originating from dispersion, that is, the wavelength dependence of the radiator's refractive index:

$$\sigma_n = \frac{dn}{d\lambda}\,\Delta\lambda\,. \tag{11.39}$$

The chromatic error on θ_c

$$\sigma_{\theta_c}\big|_{\rm chrom} = \left(\frac{\partial\theta_c}{\partial n}\right)\left(\frac{\partial n}{\partial\lambda}\right)\Delta\lambda = \frac{1}{n\tan\theta_{\rm c}}\,\sigma_n \tag{11.40}$$

directly depends on the dispersion $dn/d\lambda$ and on the allowed wavelength window $\Delta\lambda$. The dispersion relation can be evaluated from the Lorentz–Lorenz equation (see e.g. [564]):

$$\frac{n^2-1}{n^2+2} = cf(\lambda)\,, \tag{11.41}$$

with $c = 4\pi a_0^3\rho N_A/3M = (0.374\,\mathrm{cm}^3)\,\rho/M$, where N_A is the Avogadro number, $a_0 = 0.529\times10^{-8}$ cm the Bohr radius and ρ/M the radiator's molar density ($\rho/M = p/RT$ for an ideal gas with p = pressure, T = temperature and R = gas constant). The function $f(\lambda)$ accounts for the wavelength (i.e. energy) dependence of the refractivity and is usually obtained from fits to refractivity data using a parametrisation ansatz for

$f(\lambda)$ ('Sellmeier fits'[8]). The parametrisation and resulting fit values for some media are discussed for example in [1015] where further references can also be found.

The derivative of (11.41) then yields the dispersion relation

$$\frac{dn}{d\lambda} = c\frac{(n^2+2)^2}{6n}\frac{df}{d\lambda}\,, \tag{11.42}$$

which can be used to determine the chromatic dispersion of a given radiator. As an example, for fused quartz, $n = 1.474$, $\lambda = 400\,\text{nm}$, $\Delta\lambda = 100\,\text{nm}$, and fit parameters as given in [1015] one obtains a chromatic dispersion of $\sigma_n = 0.011$ yielding an angular smearing of $\sigma_{\theta_c} \approx 6.8\,\text{mrad}$.

In order to keep the chromatic contribution to the error on the Cherenkov angle small a radiator with small dispersion must be chosen (for gases He is an example). Narrowing the $\Delta\lambda$ window is an option at the price of reducing the number of Cherenkov photons. There are no obvious other possibilities to reduce this uncertainty contribution, at least not for large aperture detectors with large angular acceptance. In detectors with limited aperture, for example differential Cherenkov detectors as in fig. 11.11 on page 453, one can reduce the contribution from chromatic aberration by correction lenses.

For gas radiators with $n \approx 1$ (see table 11.1) and for close to asymptotic rings we can make the following approximations:

$$\theta_{max} \approx \sin\theta_{\max} = \frac{1}{\gamma_{th}} = \sqrt{1-\frac{1}{n^2}} \approx \sqrt{2(n-1)}\,, \tag{11.43}$$

$$1 - \frac{\theta_c}{\theta_{max}} \approx 1 - \sqrt{1-\frac{\gamma_{th}^2}{\gamma^2}} \approx \frac{1}{2}\frac{\gamma_{th}^2}{\gamma^2} \tag{11.44}$$

to write the chromatic error contribution to the angular resolution as

$$\sigma_{\theta_c}\Big|_{\text{chrom}} = \frac{\partial\theta_c}{\partial n}\sigma_n \approx \frac{\partial\theta_{max}}{\partial n}\sigma_n \approx \frac{\partial\sqrt{2(n-1)}}{\partial n}\sigma_n = \frac{1}{\sqrt{2(n-1)}}\sigma_n\,, \tag{11.45}$$

$$\frac{\sigma_{\theta_c}}{\theta_c}\Big|_{\text{chrom}} \approx \frac{1}{\sqrt{2(n-1)}}\,\frac{\sigma_n}{\sqrt{2(n-1)}} = \frac{1}{2}\frac{\sigma_n}{n-1}\,. \tag{11.46}$$

As an example we consider a gas RICH with CH_4 radiator ($n = 1.000444$, see table 11.1) of length $L = 0.5\,\text{m}$ with $N_0 = 75\,\text{cm}^{-1}$ and photon angular resolution of $\sigma_{\theta_i} = 1\,\text{mrad}$. The RICH constant then is $k_R = 1.6 \times 10^{-5}$. With (11.34) we derive that $>3\sigma$ π/K separation is achieved up to momenta of $p_{max} = 57.5\,\text{GeV}/c$ without chromatic effects. However, the chromatic dispersion contribution for CH_4 is $\sigma_n/(n-1) = 9.6\%$ [372] yielding $\sigma_n = 4.2 \times 10^{-5}$, hence lowering the separation range to $p_{max} \approx 34\,\text{GeV}/c$. Improving the resolution on θ_c would therefore not improve the particle identification capability of the detector.

Further discussion on particle separation using RICH detectors, in particular on π/K and e/π separation in context with other particle ID methods, is given in section 14.2 on page 548.

[8] Often, especially for solids, the constant c is included in f and the simpler expression $f(\lambda) = n^2 - 1$ is fitted (see e.g. [1015]).

Concluding, one can summarise the particle identification requirements on a RICH detector as follows:

- The measurement error on the radius of the Cherenkov ring $\sigma_R/R = \sigma_{\theta_c}/\theta_c$ should be as small as possible, since it directly translates into the error on β. The spatial resolution of the photon detector should therefore be comparable or better than the resolution contributions caused by other imperfections, for example by chromatic dispersion or surface roughness of the mirrors.
- The chromatic dispersion $\sigma_n = \frac{dn}{d\lambda}\Delta\lambda$ should be as small as possible. Smallest dispersions are found in light noble gases like He or Ne as well as in fluorocarbon gases with a dispersion of approximately $\sigma_n/(n-1) \approx 2$–3% in an energy window of 6.5–7.5 eV (165–190 nm) [372].
- In order to obtain as many Cherenkov photons as possible $N_0\, L$ must be large. This requires maximising the product of transmission, mirror reflectivity and quantum efficiency, $T \times R \times Q$, among others by optimal choices of detector materials and (mirror) surfaces. Note that large L (especially a large gas radiator volume) is often bounded by constraints imposed by the overall detector geometry.

11.6.3 RICH detectors in experiments

RICH detectors in the DELPHI experiment. A use of the RICH technique for the purpose of particle identification has been intensively employed in the DELPHI experiment [11, 45, 72] installed at the electron–positron storage ring LEP (see table 2.2 on page 13) operated until 2001. Figure 11.19 shows the arrangement of the RICH detectors outside of the inner tracking detector, but still inside the calorimeters and the coil of the solenoid magnet of DELPHI. The construction is a Double-RICH design (see fig. 11.13) in which Cherenkov rings originating from a liquid radiator as well as rings originating from a gas radiator are mapped onto a common photon detector. The liquid RICH has perfluorohexane (C_6F_{14}) radiators with refractive index $n = 1.28$ (see table 11.1), the gas RICH has a perfluorobutane (C_4F_{10}, $n = 1.00150$, see table 11.1) radiator in the forward region and a perfluoropentane (C_5F_{12}, $n = 1.00192$, see table 11.1) radiator in the barrel region. The choice of different gases is due to space constraints while trying to keep the same number of Cherenkov photons. The liquid radiators are only 1 cm thick such that proximity focusing on the photon detector can be used. For the gas radiators a mirror system focuses the Cherenkov cones onto the same photon detector. The latter consists of several gas-filled chambers with quartz entrance windows, designed similar to fig. 11.14 as time projection chambers (see section 7.10.10 on page 241). The gas filling is a mixture of hydrocarbons (75% methane, 25% ethane) with 0.1% TMAE vapour addition to convert the Cherenkov photons into photoelectrons.

Table 11.3 summarises the photon yield. The number of photons generated by a particle depends on the Cherenkov angle θ_c and the radiator length L (eq. (11.13)). In both liquid RICHs with $\theta_c \approx 38.8°$ (for $\beta \to 1$) and a radiator length of 1 cm a particle creates about 250 Cherenkov photons in the detectable wavelength range (175–230 nm) [43]. The forward gas radiators are 60 cm deep on average. A $\beta = 1$ particle creates about 150 photons in the detected wavelength range of 165–230 nm at an angle of 3.2° [43]. After multiplication with efficiency factors Q, T and R from (11.17), integrated over λ, table 11.3 presents the number of photoelectrons N_{pe} detected in the photon detector.

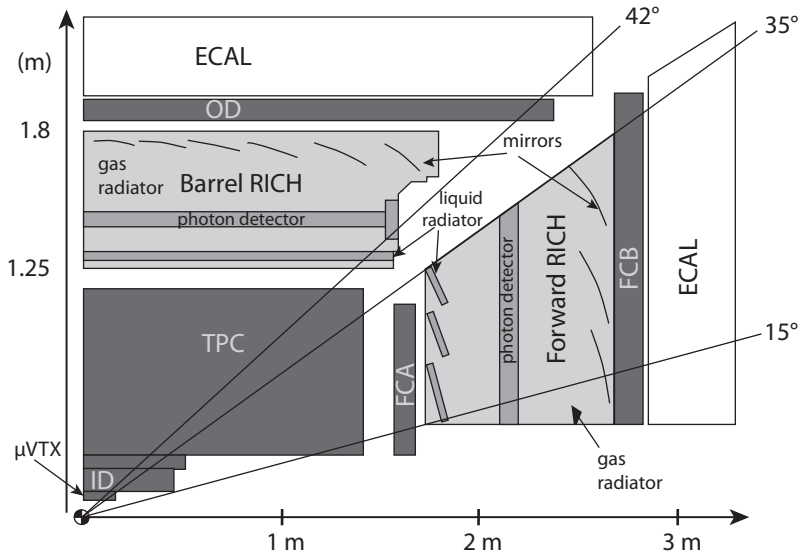
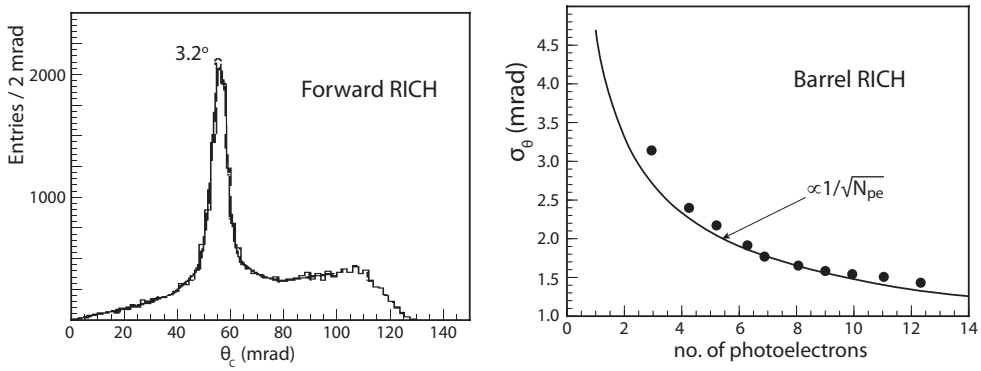


Fig. 11.19 Arrangement of RICH detectors in the DELPHI experiment operated at the electron–positron storage ring LEP. Shown is one quarter of the DELPHI detector in a cut parallel to the beam line. The interaction point is at the origin of the coordinate frame. In the 'barrel' as well as in the 'forward' angular region of the experiment Double-RICH systems are implemented in which the two respective Cherenkov rings of a gas and of a liquid radiator are focused onto the same photon detector (see also fig. 11.13), where the Cherenkov rings are reconstructed. The drawing illustrates the space requirements needed for the voluminous gas radiators. Abbreviations are: ID/OD = inner/outer detector, TPC =time projection chamber, FCA/B = forward chamber A/B, ECAL = electromagnetic calorimeter. Adapted from [72] with kind permission of Elsevier.

Table 11.3 Mean photon and photoelectron yields for ($\beta \approx 1$) particles detected with the RICH detector system of the DELPHI experiment [43, 651]. Particles traverse the different radiators with an average length ⟨length⟩. The radiators are labelled by their state of matter (l = liquid, g = gas) and by their position in the detector (b = barrel, f = forward). N_∞^γ is the number of asymptotically ($\beta = 1$) generated photons. After losses in radiator, mirror system, and entrance window they reduce to $N_{\text{all}}^{\text{pe}}$ photoelectrons in the focal plane of the photon detector without acceptance limitations (according to simulations done for the Forward RICH system). After acceptance, dead time and threshold losses, $N_{\text{det}}^{\text{pe}}$ detected photoelectrons remain. The measured yield is $N_{\text{meas}}^{\text{pe}}$ [43, 72]. $\langle\lambda\rangle$ is the average wavelength of the detected Cherenkov photons.

radiator	⟨length⟩ (cm)	refraction index	θ_c (β=1)	N_∞^γ	$N_{\text{all}}^{\text{pe}}$	$N_{\text{det}}^{\text{pe}}$	$N_{\text{meas}}^{\text{pe}}$	$\langle\lambda\rangle$ (nm)
C_6F_{14} (l, f)	1	1.2827	38.8°	≈ 250	27.0	9.7	7.8	193.1
C_6F_{14} (l, b)	1	1.2827	38.8°	≈ 250	27.0		15	
C_4F_{10} (g, f)	60	1.00150	3.2°	≈ 150	17.1	9.9	10.9	189.7
C_5F_{12} (g, b)	45	1.00192	3.4°	≈ 110	11.1		9	



(a) θ_c spectrum DELPHI-RICH. (b) Resolution as a function of N_{pe}.

Fig. 11.20 Cherenkov angle resolution for muons from $Z^0 \to \mu^+\mu^-$ decays, measured by the gas RICH detectors of DELPHI (adapted from [44], with kind permission of Elsevier). (a) Measured θ_c spectrum in the gas radiator (C_4F_{10}) of the Forward RICH. On average about 10 photoelectrons are detected per radiating muon. The mean value of the peak above background is 55.6 mrad (3.2°), the standard deviation is 2.85 mrad per photoelectron (Gaussian fit). (b) Resolution of θ_c in the gas radiator (C_5F_{12}) of the Barrel RICH as a function of the number of detected photoelectrons. The solid line indicates a $1/\sqrt{N_{pe}}$ behaviour.

Figure 11.20 shows the resolution obtained for the measurement of the Cherenkov angle for muons from $Z^0 \to \mu^+\mu^-$ decays having velocities of $\beta \approx 1$. Panel (a) of fig. 11.20 shows the measured spectrum of Cherenkov angles for single photons in the Forward RICH. Figure 11.20(b) shows the angular resolution for the Barrel RICH for an entire Cherenkov ring as a function of the number of photoelectrons on the ring. The angular resolution per photon is about 7% for the gas radiators and about 2% for the liquid radiators [11].

The resulting performance to distinguish kaons and protons from electrons, pions, and muons is shown in fig. 11.21 (see also chapter 14). The dominant part of the momentum range relevant for hadrons from e^+e^- collisions in the first phase of LEP (LEP 1) is covered reaching up to about 10 GeV/c.

For the LHC experiments ALICE and LHCb the RICH technique has been further developed for particle identification over a wide range of momenta. This is described in chapter 14.

'Hadron-blind' electron identification in the CERES experiment. The experiment CERES/NA45 [162] at the CERN SPS heavy ion beam aimed to detect e^+e^- pairs in heavy ion collisions (for example S+Au or Pb+Pb) as a signature of a hot plasma of quarks and gluons. From the collisions a multitude of charged and neutral hadrons emerge, in central ion–ion collisions up to several hundreds in the acceptance range of the detector [59]. The detection of thermally produced e^+e^- pairs in the presence of much more abundant numbers of hadrons is therefore a formidable task. In addition, π^0's being 10^5 times more abundant then thermal e^+e^- pairs can—besides decaying into two photons that may convert into e^+e^- pairs in the detectors material—also directly decay into $e^+e^-\gamma$ (Dalitz decay), hence constituting 'irreducible' background.

The experimental approach in CERES was 'hadron-blind detection', that is, almost complete suppression of any hadron signals by means of a double Cherenkov detector system, built using minimal material to keep photon conversions at a tolerable level. Pair-produced electrons and positrons are detected by two methane-filled RICH detectors set to a Cherenkov threshold of $\gamma_{th} \approx 32$. This threshold is reached for electrons already at momenta of 16 MeV/c. By contrast, Cherenkov rings created by pions ($p_{th} \approx 4.5$ GeV/c) are strongly suppressed. Only for momenta above 20 GeV/c do pions develop (asymptotic) rings indistinguishable from electron rings.

Figure 11.22 shows the experimental set-up [59] featuring the two interleaving RICH systems. Particles emitting Cherenkov radiation create rings in both RICH detectors. The superconducting coils generate a local magnetic field whose bending effect is essentially confined on a small volume between RICH-1 and RICH-2 where electrons and positrons are deflected in opposite directions in the transverse plane. The corresponding azimuthal displacement of their respective Cherenkov rings in UV detector 2 compared to UV detector 1 allows determination of the radiating particle's charge sign and momentum. By means of (warm) correction coils the magnetic field is shaped such that upstream of the main coils the field is reduced to almost zero and that downstream the field lines in RICH-2 run largely parallel to the trajectories of particles originating from the target so that they are not deflected any further in this part of the detector volume. The photon detectors feature UV-transparent entrance windows and gas fillings of He + C_2H_6 + TMAE. The readout principle of parallel plate detectors with multistage amplification is illustrated and explained in fig. 11.15 on page 457.

Figure 11.23 shows Cherenkov rings in both RICH detectors of the CERES/NA45 experiment. Shown are the rings created by e^+e^- pairs from a p+Be collision (a,b)

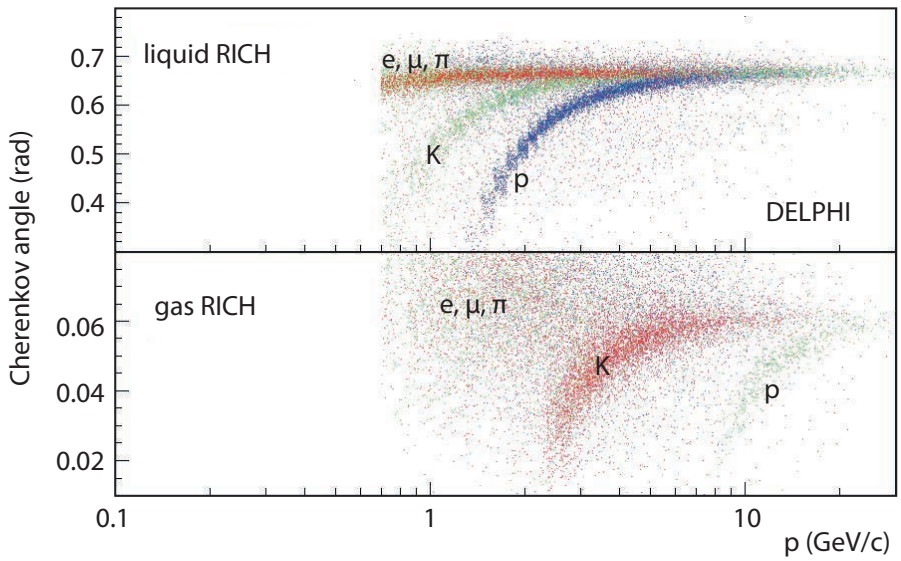


Fig. 11.21 Particle identification with the RICH system of the DELPHI experiment. Plotted are measured Cherenkov angles as a function of (measured) momentum. One observes bands corresponding to different particle species which are well separated at not too high momenta (from [347]).

featuring small backgrounds only, and a S+Au collision (c,d) in which the e^+e^- rings must be reconstructed and separated from a background of pions. The ring centres in UV detector 2 have a larger distance than in UV detector 1 owing to the bending by the local magnetic field between RICH-1 and RICH-2.

Belle-II aerogel RICH with proximity focusing. Photon yield is increased by using a thicker radiator. In the case of solid or liquid radiators together with proximity focusing this, however, bears the disadvantage that the angular resolution for single photons is deteriorated, because the point of emission is less well resolved (see fig. 11.18 and eq. (11.35)). This limitation can be overcome if the radiator is arranged in layers such that the refractive index increases in steps, as shown in fig. 11.24. At the photon detector plane the radiation from different n layers meets at about the same ring radius.

With aerogel as radiator material one can tune the desired index of refraction by changing the aerogel density. By the design shown in fig. 11.24 a focusing effect of the Cherenkov radiation was realised [636] for the Belle-II experiment at the e^+e^- storage ring SuperKEKB in Japan using a two-layer aerogel radiator [554, 778].

Application of the RICH technique for neutrino detection. An example experiment in which the RICH principle is employed for detection of neutrino interactions is the Super-Kamiokande experiment in Japan [432]. The Super-Kamiokande experiment discovered neutrino oscillations in atmospheric neutrino [431] as well as in solar

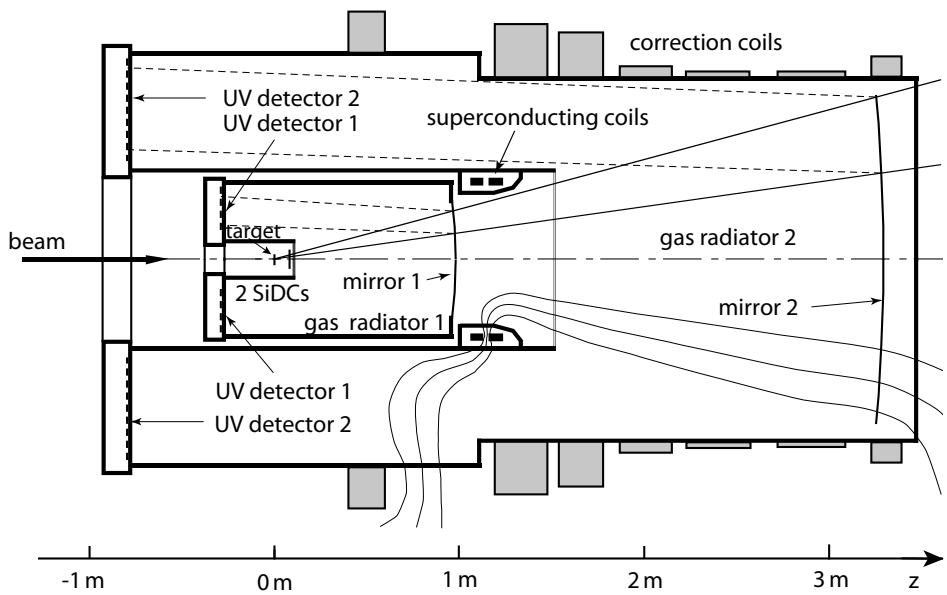


Fig. 11.22 'Hadron-blind' heavy ion experiment CERES/NA45 at the CERN SPS (*fixed target* experiment) [60] (with kind permission of Elsevier). The main elements of the apparatus are two interleaved gaseous Cherenkov detectors. Cherenkov radiation is focused by mirrors onto two UV RICH detectors. The radiating electrons and positrons are azimuthally deflected in a local magnetic field between RICH-1 and RICH-2 (see text). Warm coils shape the field lines in RICH-2 to run parallel to particle trajectories from the target without azimuthally deflecting them. The UV detectors are shown in fig. 11.15.

neutrino detection [541] (see also section 16.6). The experiment as well as its predecessor Kamiokande, were originally designed to search for proton decay $p \to \pi^0 + e^+$ in which the positron emits Cherenkov radiation. To detect the flavour of neutrino reactions Cherenkov emission of electrons and muons, respectively, is exploited.

The Super-Kamiokande detector is a large cylindrical water tank, 42 m high and 39 m in diameter, containing 50 000 tons of ultra clean water. The detector is installed 1000 m underground in a mine near Hida, Japan (see fig. 16.27 on page 694). Light signals are detected by a total of about 13 000 large photomultipliers installed on the tank's walls (fig. 10.12 on page 419). By means of black foils the water volume is optically separated into an inner (facing the PMTs) and an outer part (behind the PMTs) suppressing unwanted low-energy events from residual radioactivity creating unwanted Cherenkov light as well as from cosmic ray muons entering the detector from outside.

The goal of the neutrino part of the experiment is the detection and distinction of electron- and muon neutrinos by means of the reactions

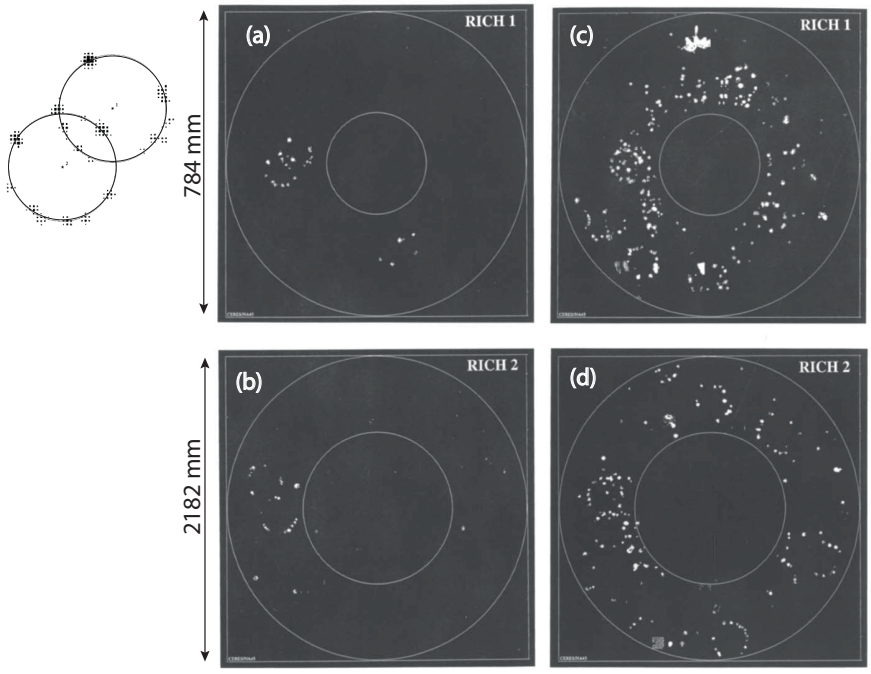


Fig. 11.23 Cherenkov rings in the two RICH detectors of the CERES (NA45) experiment [162] (with kind permission of Elsevier). Shown are the Cherenkov rings of e^+e^- pairs in an almost background-free p+Be collision, likely from the decay $\pi^0 \to e^+e^-\gamma$ (a,b) and from a S+Au collision (c,d) with background originating from pions, each shown in RICH-1 (a,c) and in RICH-2 (b,d). Note that the movement of the upper (lower) ring in (a) is downward (upward) when going from (a) to (b). On the left an enlarged view of the two rings in (a) is shown, displaying those electrodes of the UV detector (fig. 11.15) that have a signal above threshold. Pairs of rings appear in RICH-1 and in RICH-2 at about the same position (approximately at a clockwise position of 9:00 in all pictures). In (c,d) additional electron-like rings (single, no pairs) are visible which are background.

$$\nu + e^- \rightarrow e^- + \nu \tag{11.47}$$
$$\nu + \mathrm{N} \rightarrow l^- + X \tag{11.48}$$

that is, scattering of $\nu = \nu_{e,\mu,(\tau)}$ off shell electrons e^- or nuclei N of the water molecule. The lepton in the final state of the reaction (11.48) signals the flavour of the neutrino that has initiated the reaction.

There are no focusing elements in the Super-Kamiokande detector. The registered light at the wall of the tank originates from the Cherenkov radiation conically emitted over the entire flight path of the generated electron or muon in the tank. The Cherenkov rings are therefore no sharp images as in the RICH examples discussed previously, but are rather broadened in a way that depends on the generating particle's flight path.

The interesting signals in Super-Kamiokande are GeV μ- or e-neutrinos, originating in the atmosphere or MeV neutrinos from the sun. The atmospheric ν_μ interact with the hydrogen or oxygen nuclei of the water in the tank and generate muons that often still stop (range $\approx$ few metres) within the detector volume after a short flight distance (*fully contained muons*). Muons or electrons entering the detector from cosmic showers rather than from a neutrino reaction are background vetoed by observing signals in both detector volumes. Neutrino-induced signals, by contrast, are demanded to originate from the inner volume part (*fiducial volume*) and have no signal in the outer part.

Low energy neutrino reactions ($E \lesssim 1$ GeV) react quasi-elastically and only contain the lepton (μ or e) in the final state (e.g. as in fig. 11.25(b)). At higher energies pions as well as some other other hadrons can also be produced in addition to the lepton (fig. 11.25(a)). Higher energetic muons can also leave the detector (*partially contained muons*), electrons usually not.

Ring-shaped patterns with a sharp outer edge are created by fully-contained muons from neutrino reactions. More highly energetic muons from cosmic rays, usually traversing a larger distance in the detector volume, create light signals in many PMTs within their Cherenkov cone and can thus be identified as background. Scattering of the radiating particles (electron or muon) causes further changes in the ring shapes. Electrons lose energy in the water by bremsstrahlung and can also create a shower if their energy is large enough (fig. 11.26(a)). Muons are only deflected by multiple Coulomb scattering (fig. 11.26(b)). The colour code in fig. 11.26 indicates the light arrival time. Time measurement allows a determination of the incidence angle of the

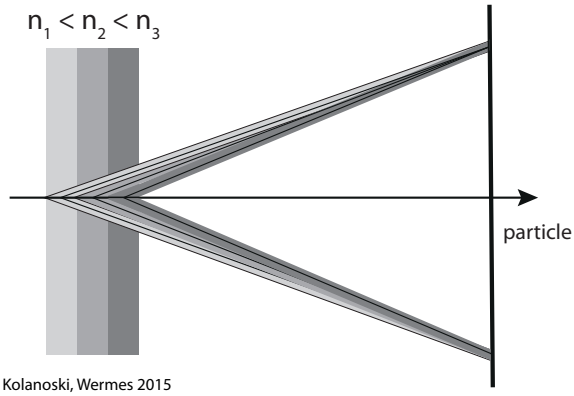


Fig. 11.24 Focusing effect obtained by an inhomogeneous aerogel layer arrangement [738].

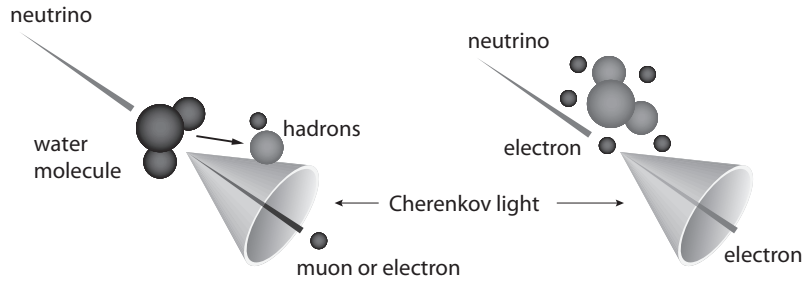


Fig. 11.25 Neutrino induced reactions [921] in the Super-Kamiokande detector. Neutrinos scatter off nuclei (left) or off shell electrons (right) generating electrons or muons in the final state which emit Cherenkov radiation. Since there are no focusing elements the radiation is emitted over the e or μ travel distance within cones around their flight direction (see also fig. 11.3).

radiating lepton relative to the PMT wall.

The sharpness of the outer ring edge can be used efficiently to distinguish ν_e-induced from ν_μ-induced reactions. This was the most important criterion for the neutrino oscillation experiments. The ring identification further allows reconstructing the direction of electrons from solar neutrino reactions (see also fig. 16.28(a)) with a resolution of about 25° at 10 MeV [541].

11.7 Detection of internally reflected Cherenkov light (DIRC)

A special ring imaging detector particularly taking account of the space constraints existing at collider detectors has been developed for π/K separation in the BABAR experiment at the PEP-II e^+e^- storage ring at the Stanford Linear Accelerator Center (SLAC) [1003]. Electrons of 9 GeV collide with 3.1 GeV positrons producing a boosted centre-of-mass system optimised for CP-violation studies. The detector principle is illustrated in fig. 11.27.

The so-called DIRC detector (*Detection of internally reflected Cherenkov light*) [322, 42] brings the Cherenkov light emitted by particles in quartz bars out of the detector volume by angle conserving total reflection. The design exploits the fact that asymmetric beam collisions at the PEP-II B-Factory allow for an asymmetric detector design where the photon detector can be placed in a solid angle region opposite to the boost direction of the colliding system. Here the Cherenkov light is focused onto a matrix of photomultipliers for Cherenkov ring reconstruction, while on the opposite side the quartz bars are metal coated, reflecting the light arriving at this end back to the photomultiplier side. The DIRC has been developed for science at the B-Factory operating at centre-of-mass energies around 10 GeV (see also table 2.2). Detailed study of the final states produced demand excellent π/K separation in the momentum range up to about 5 GeV/c.

The condition for total reflection at a transition from a material with refractive index $n > 1$ into air ($n_0 \approx 1$) is $\sin\alpha > 1/n$. For perpendicular incidence of a charged particle into the rectangular quartz bar (fig. 11.28(a)) we then obtain:

$$\sin\alpha \;=\; \sin\theta_c = \sqrt{1 - \frac{1}{\beta^2 n^2}} > \frac{1}{n}$$

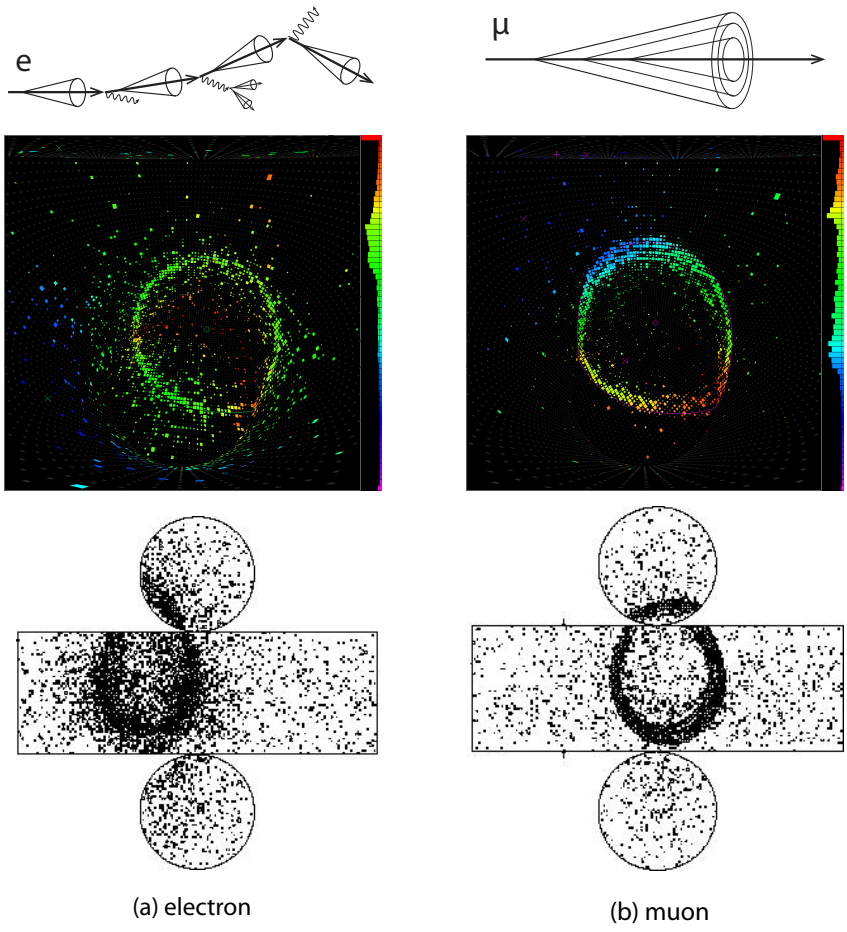
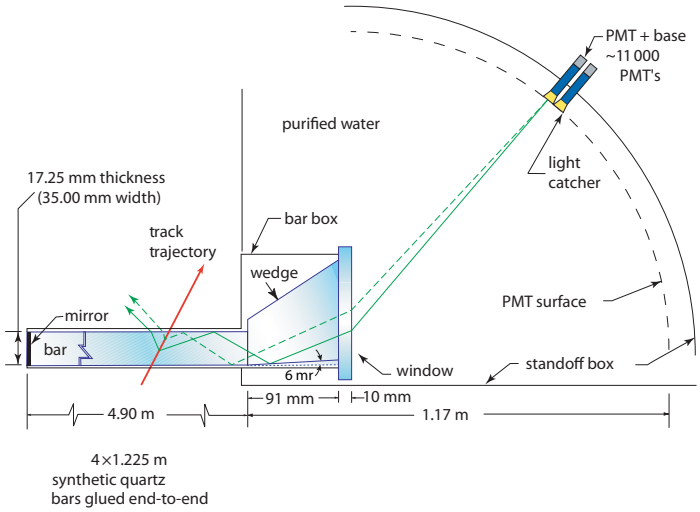


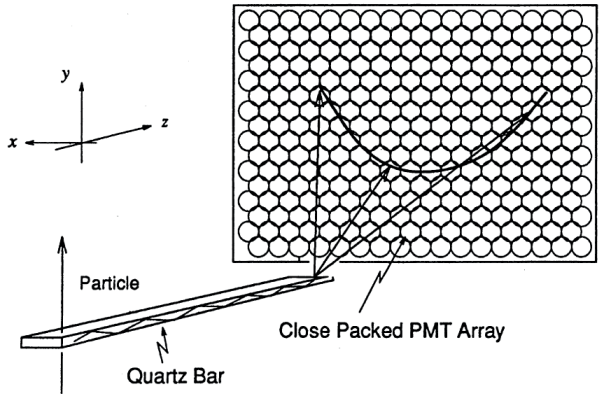
Fig. 11.26 Two different examples each of electron and muon events in the Super-Kamiokande experiment. (a) Electrons lose energy by bremsstrahlung and they can shower (top). Their Cherenkov ring edges are blurred as the two examples of different energy (middle and bottom) show [583]. (b) Muons create rings if they are stopped still in the detector. This corresponds to the case of proximity focusing, discussed at the start of section 11.6. Or they create rings with PMT hits filling the area near the circumference from the outside in, as the muons slow down (high energetic muons, top). The ring edges are comparatively sharp (middle and bottom [583]). The colour code indicates the arrival time of the Cherenkov photons. Source: Kamioka Observatory, Univ. Tokyo, and Super-Kamiokande, with kind permission.

$$\Rightarrow \qquad n > \sqrt{1+\frac{1}{\beta^2}} \qquad \xrightarrow{\beta\to 1} \qquad n > \sqrt{2}\,. \tag{11.49}$$

For amorphous quartz (fused silica) having a refractive index $n = 1.474$ (at $\lambda = 400\,\text{nm}$, see table 11.1) inequality (11.49) is fulfilled for β values larger than 0.925, while the Cherenkov threshold lies at $\beta_{th} = 0.96$. There will therefore always be wavelength parts of the emitted Cherenkov spectrum that are totally reflected at the boundaries



(a) Cross section of the BaBar DIRC detector



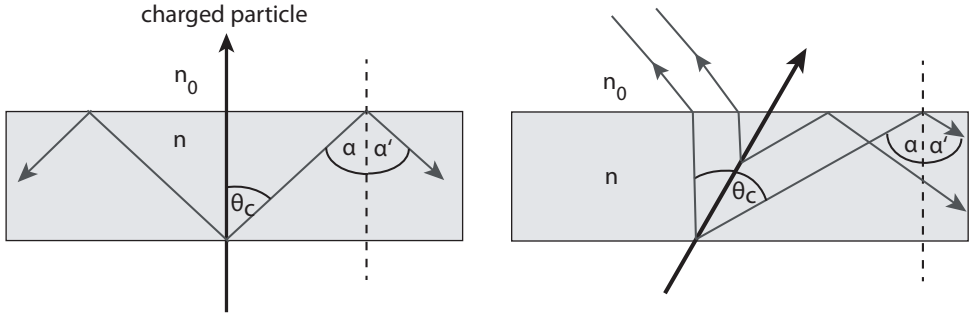
(b) The DIRC principle

Fig. 11.27 Principle of the DIRC detector [41, 42]. In a quartz bar charged particles generate Cherenkov light which is brought to the outside by angle conserving reflection and then focused onto a photon detector. The fraction of light that is not totally reflected is lost (reprint with kind permission of the BaBar Collaboration).

and are transported by multiple reflections to the end of the bar [808]. Different to the case of perpendicular incidence, for inclined tracks about half of the Cherenkov radiation intensity is lost (fig. 11.28(b)) when the condition of total reflection is no longer valid. For the other half of the radiation intensity, however, total reflection is guaranteed such that some portion of the Cherenkov light always reaches the photon detector, if (11.49) is fulfilled.

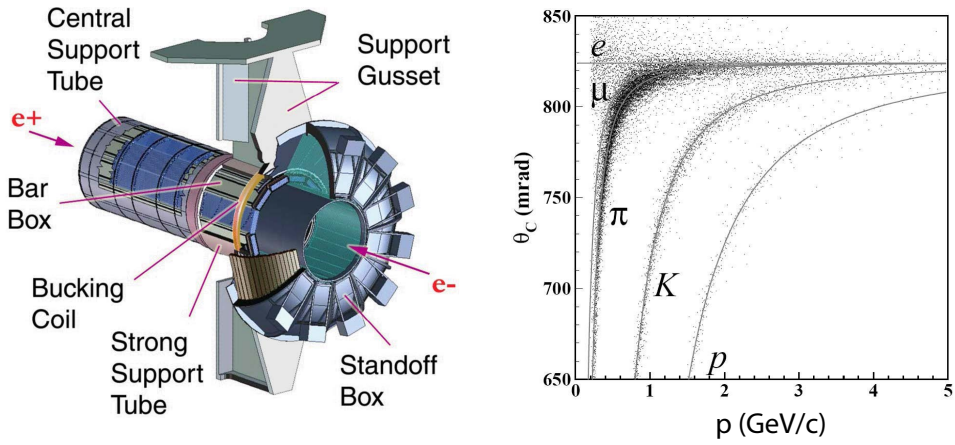
Fused silica is also a good choice because of its large absorption length for light in the Cherenkov wavelength spectrum ($> 30\,\mathrm{m}$ for $\lambda > 300\,\mathrm{nm}$ [808]), low chromatic dispersion and because it enables optical surface finishing.

A total of 144 almost 5 m long rectangular quartz bars ($17.25 \times 35\ \mathrm{mm}^2$) are placed outside the central drift chamber covering the barrel region of the BaBar detector (fig. 11.29(a)). On one side the quartz bars are mirror coated reflecting the light. At the other side the Cherenkov light emerges via a prism wedge into a detection volume filled with ultra pure water where it is expanded onto the PMT matrix arranged as a ring-shaped shell. This simple arrangement reduces the detector coverage needed



(a) perpendicular incidence. (b) inclined incidence.

Fig. 11.28 Relation between the Cherenkov angle and total reflection within the quartz radiator bar: (a) for perpendicular incidence of a charged particle; (b) for inclined incidence.



(a) BaBar DIRC. (b) Particle separation using the DIRC.

Fig. 11.29 (a) Arrangement of the DIRC detector in the BaBar experiment. The quartz bars are arranged barrel-wise in the *bar box* over the full length of the detector. The generated Cherenkov light is focused into the 'standoff box' filled with PMTs. (b) Separation power obtained with the DIRC detector of BaBar. Shown are the measured Cherenkov angles for e, μ, π, K, p as a function of the particle momentum [42]. (Source: BaBar Collaboration and SLAC with kind permission.)

by more than 50% and avoids mirrors. The refractive index of water is close enough to that of quartz such that reflection losses at the boundary can be tolerated. A small bevelling of the prism's bottom surface by 6 mrad (see fig. 11.27) reduces the relative displacement of downward reflected photons at the prism entrance (solid line in fig. 11.27) compared to upward ones (dashed line) due to the finite bar thickness such that both roughly arrive at the same photomultiplier tube.

Particles impinging perpendicularly onto the quartz bar produce about 1000 Cherenkov photons (for $\beta \approx 1$). Less than half is kept in the bar by total reflection. After losses from reflections at the coated end and from coupling out into the 'standoff box', about 10–20 photons are left reaching the photon detector. Particles entering under

steep incidence angles, and hence traversing substantially more radiator material than under perpendicular incidence, result in yielding up to 80 photons at the PMT matrix [878].

The photomultipliers of the PMT matrix measure the position and time of the arriving Cherenkov photons relative to the time of the collision. Both serve to reconstruct the angle of the emitted radiation. Measurement of the light's travel distance provides a spatial resolution of about 10 cm along the quartz bar ($\sigma_t \approx 1.5$ ns). The time measurement also relates directly to the original emission angle of the photons relative to the surface of the bar, because steeper angles cause more reflections and longer travel distances as a consequence. Since position and angle of incidence of the charged particle onto the bar is known, a measurement of the photon arrival time with sufficient resolution is an independent determination of the Cherenkov angle [177].

The arrangement of the DIRC detector in the BaBar experiment can be seen in fig. 11.29(a). Figure 11.29(b) shows the distribution of measured Cherenkov angles for different charged particles as a function of momentum. π/K separation ($2\,\sigma$) is possible for momenta up to 4.2 GeV/c [42]. The DIRC concept is also adopted by the PANDA experiment planned at the FAIR facility (GSI, Darmstadt, Germany) [895].

Further experiments using ring imaging Cherenkov detectors for particle identification are described in chapter 14.

11.8 Cherenkov radiation in astroparticle physics experiments

In astroparticle physics the use of Cherenkov radiation has found many detector applications. Besides water or ice, the Earth's atmosphere is also used as a radiator wherein high energy particles generate air showers. The charged particles of a shower with velocities exceeding the Cherenkov threshold velocity (mostly electrons and positrons) emit Cherenkov radiation, which can be detected as a whole in an energy (calorimetric) measurement or can be detected as an image (imaging detector).

Astroparticle physics experiments using Cherenkov radiation as a detection method can be classified according to how the radiation is used for the scientific goals of an experiment:

- Experiments such as Super-Kamiokande (discussed in section 11.3) that measure Cherenkov rings to detect and identify low to medium energy neutrino reactions (see also section 16.6.1).
- Experiments with high energy neutrinos like IceCube [12] or ANTARES [329] that exploit Cherenkov radiation to measure the direction and the arrival time of Cherenkov radiation to spatially reconstruct the particle track or cascade originating from the neutrino reaction. The experiments use huge radiator volumes in the polar ice, respectively in sea water, in which the Cherenkov radiation is detected (see section 16.6.5)
- So-called *imaging atmospheric Cherenkov telescopes* (IACTs) that focus the Cherenkov light to obtain an image of the air shower, but do not explicitly reconstruct Cherenkov rings. This method is used to detect high energy (above about 100 GeV) gamma rays. One or more Cherenkov telescopes map out the emission cone of the Cherenkov radiation generated by an air shower, as illustrated in fig. 16.24 on page 689. Examples for this measurement technique are the telescopes H.E.S.S. [521], MAGIC [683] and VERITAS [960] (see section 16.5).

– Experiments using Cherenkov light as a measure for the energy loss of particles from the air shower (see section 16.4); examples are AUGER [29] and IceTop [16] (see section 16.4.3 on page 682). Air shower particles with $\gamma > 1.51$, respectively $\gamma > 1.55$ (table 11.1) generate Cherenkov light in water (AUGER) or in ice tanks (IceTop). The light intensity is about proportional to the energy loss dE/dx of the particles in the detector volume. From the total energy deposited in all tanks the energy of the primary particle that initiated the air shower can be estimated.
– Experiments detecting the Cherenkov light from air showers in clear nights in a non-focusing manner for a calorimetric determination of the energy of the shower. The TUNKA experiment [259] at Lake Baikal is an example (section 16.4.3).

For a more detailed description of astroparticle physics experiments we refer to chapter 16.

Already for many decades this effect has been exploited in high energy physics experiments to distinguish charged particles with different masses from each other

12
Transition radiation detectors

A change in an electromagnetic field configuration leads to emission of electromagnetic radiation. For a moving charged particle this can be caused by a change of velocity and/or direction (bremsstrahlung off nuclei or synchrotron radiation by bending in a magnetic field, also in vacuum) or if the polarisation fields of the medium surrounding a traversing particle cannot follow the particle's speed, as we have also seen for Cherenkov radiation. Electromagnetic radiation is also emitted if a charged particle passes from one medium with certain electromagnetic properties (characterised by ε, μ or by the refractive index $n = \sqrt{\varepsilon\mu}$) into another medium with different properties. During the transition process the electromagnetic field determined by the charged particle in its respective surrounding must rearrange even if the particle's movement is uniform. Obeying Maxwell's laws this rearrangement process of electromagnetic fields causes *transition radiation* [458].

As will be shown, the intensity of transition radiation is proportional to the Lorentz factor γ of the traversing particle (up to saturation). Transition radiation can hence be exploited to identify highly relativistic charged particles with velocities $\gamma \gg 100$ for which Cherenkov radiation is no longer attractive since for $\gamma \gg \gamma_{th}$ (γ_{th} = threshold velocity) only small changes $\Delta\theta_c$ in the Cherenkov angle occur (see fig. 11.4 on page 443 and section 14). More detailed descriptions on transition radiation and transition radiation detectors can be found in [357, 119].

12.1 Transition radiation

Transition radiation (TR) is emitted when charged particles traverse the boundary between two media with different indices of refraction. This was predicted in 1946 by Ginzburg and Frank [425] and first observed at optical wavelengths in 1959 by Goldsmith and Jelley [466]. X-ray emission by ultra relativistic charged particles, and hence the possibility to use transition radiation in particle detectors, was pointed out in 1958 by Garibian [443]. Important studies for a deeper understanding of transition radiation and its detection were performed by Alikhanian [82] in the 1960s.

Figures 12.1 and 12.2 illustrate the phenomenon of transition radiation. When a charge in a dielectric medium (ε_1) approaches the boundary to another medium with different dielectric properties (ε_2) the electric field configuration undergoes a transition. The static field configuration is qualitatively displayed in fig. 12.1 and can be readily calculated using the familiar concept of image charges. The concept can also be applied in dielectrics when accounting for the medium polarisation by

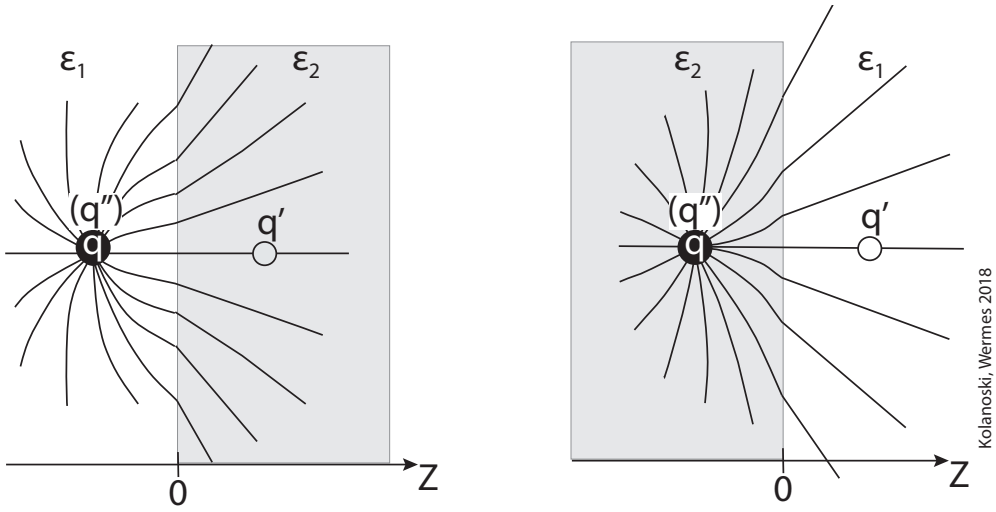


Fig. 12.1 Static field configurations (qualitative) of a charge in a medium near a boundary to a medium with different dielectric properties quantified by ε_1 and ε_2, respectively (see also [564]). Left: charge q residing in ε_1 medium, right: charge q residing in ε_2 medium with $\varepsilon_2 > \varepsilon_1$.

employing image charge values according to the dielectric properties of the two media (see e.g. [564]):

$$q' = -\frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_2 + \varepsilon_1}\, q \qquad \text{and} \qquad q'' = \frac{2\varepsilon_2}{\varepsilon_2 + \varepsilon_1}\, q\,, \tag{12.1}$$

respectively. In the hemisphere containing the physical charge ($z < 0$ in fig. 12.1) the field is that of two point charges, namely q and its image charge q' at the given distance, while in the hemisphere not containing any charge ($z > 0$) there are no singularities and the field there is simply that of the charge q'' defined in (12.1) placed at the point of the original charge q at $z < 0$ [564]. In this static picture, changes of the field configurations at the transition boundary are evident in fig. 12.1.

In a dynamic contemplation the charge (here an electron) moves towards the boundary where the fields suddenly must rearrange with the consequence that dipole radiation fields are to be 'shaken off' in order to guarantee a continuous transition of the electromagnetic field relations at the interface, as required by Maxwell equations. This is sketched in fig. 12.2 where the emission is drawn as a TR photon within an intensity profile cone. As long as this cone is completely contained in the second medium there is no dependence on the angle of incidence of the particle relative to the boundary [119]. In the picture with image charges this corresponds to (partial) annihilation of charge and image charge. The annihilation is complete only in the limit of a vacuum–conductor transition [458].

Transition radiation with sufficient intensity for detection results from a superposition of interfering radiation fields from many interfaces. With appropriate design of the interface layers with respect to thickness, distance and materials, sufficient detectable net intensity remains from multiple superposition of TR radiation fields.

The total energy W emitted by transition radiation increases with the Lorentz factor γ of the traversing particle; for a single interface it is directly proportional to γ

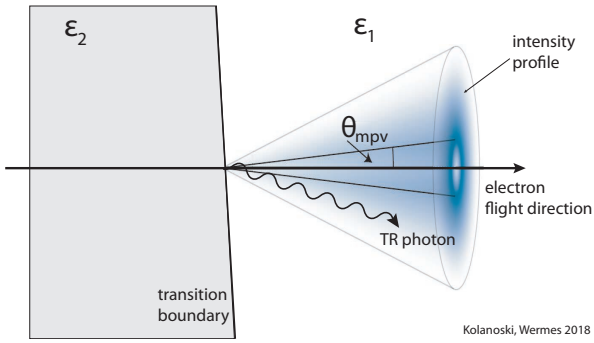


Fig. 12.2 Illustration of transition radiation: a charge in a medium (ε_2) approaches a boundary to a medium with different dielectric properties ε_1. At the transition boundary electromagnetic radiation is emitted.

(see section 12.1.5). As for bremsstrahlung, transition radiation is also strongly peaked forward.

12.1.1 Radiation intensity

The radiation characteristics follows the general dipole form of electromagnetic radiation. As a limiting case one may consider a non-relativistic electron with $z = 1$ moving with constant velocity v in vacuum towards a metal surface [357]. In this case transition radiation is identical to bremsstrahlung because the particle is brought to a complete stop at the metal and the radiation corresponds to the annihilation of the dipole formed by the electron and its image charge. The radiated energy W per solid angle $d\Omega$ and frequency interval $d\omega$ has the same angular distribution as a dipole placed at the metal surface [425, 457]:

$$\frac{d^2W}{d\omega\, d\Omega} = \frac{\hbar\alpha\, v^2}{\pi^2 c^2}\, z^2 \sin^2\theta\,. \tag{12.2}$$

For a relativistic particle and a transition from a dielectric medium ($\varepsilon \neq 1, \mu = 1$) to the vacuum ($\varepsilon = 1$) the Ginzburg–Frank formula for the emission characteristic holds[1] [425, 357] which is:

$$\frac{d^2W}{d\omega\, d\Omega} = \frac{\hbar\alpha}{\pi^2}\, z^2\, \beta^2\, \frac{\sin^2\theta\cos^2\theta}{(1-\beta^2\cos^2\theta)^2} \times \left[\frac{(\varepsilon-1)(1-\beta^2-\beta\sqrt{\varepsilon-\sin^2\theta})}{(1-\beta\sqrt{\varepsilon-\sin^2\theta}\,)(\varepsilon\cos\theta+\sqrt{\varepsilon-\sin^2\theta}\,)}\right]^2. \tag{12.3}$$

Transition radiation is linearly polarised with the electric field vector $\vec{E}$ lying in the plane spanned by the particle velocity $\vec{v}$ and the radiation vector $\vec{k}$ [424]. This polarisation is, however, of no use in any practical detector application, so we will not consider it any further.

12.1.2 Interface materials

For highly relativistic particles ($\gamma > 1000$) transition radiation is emitted with energies up to the hard X-ray regime. These energies are much larger than the binding energies

[1] The Ginzburg–Frank formula for arbitrary dielectrics with $\varepsilon_1 \neq \varepsilon_2$ is given for example in [357], eq. (2).

Table 12.1 Properties of some materials used for transition radiation detectors [932, 545].

Material	Density ρ (g/cm^3)	Plasma energy $\hbar\omega_p$ (eV)	Linear absorption coefficient at 10 keV μ (cm^{-1})	Radiation length X_0 (cm)
air	1.205×10^{-3}	0.71	6.2×10^{-3}	30 400
He	1.66×10^{-4}	0.26	4.1×10^{-5}	56 819
Li	0.534	13.8	0.18	155
Be	1.85	26.1	1.2	35.3
Al	2.7	32.8	70.8	8.9
polyethylene	0.93	20.53	1.94	50.3
Mylar*	1.38	24.6	4.8	28.5
polypropylene	0.90	20.51	1.88	49.7
Rohacell**	0.032–0.075	23.09	0.11–0.25	1275–544

* Trade name for polyethylene terephthalate (PET, $(C_{10}H_8O_4)_n$).
**Trade name for foam of variable density based on polymethyl methacrylate (PMMA, $(C_5O_2H_8)_n$).

in solids and exceed by far the resonance frequencies of a medium. Hence the medium can be considered as a free electron gas whose properties are characterised by the plasma frequency ω_p introduced in section 3.2.1.3 (page 33):

$$\omega_p = \sqrt{\frac{e^2 n_e}{m_e \varepsilon_0}}\,, \tag{12.4}$$

(in SI units) where n_e is the electron density (3.12) and m_e the electron mass. The plasma frequency ω_p is the frequency of collective oscillations of the electrons of the medium. With the mass density given in units of g/cm^3 the corresponding plasma energy can be expressed as:

$$\hbar\omega_p \approx 28.8\,\text{eV}\sqrt{\frac{\rho}{\text{g cm}^{-3}}\frac{Z}{A}}\,. \tag{12.5}$$

For air and some materials often used in transition radiation detectors like Mylar, Rohacell, and others the plasma energies are given in table 12.1.

Far above atomic resonance frequencies of a medium the dielectric constant ε can be approximated as [357]:

$$\varepsilon(\omega) \approx 1 - \frac{\omega_p^2}{\omega^2} =: 1 - \xi_p^2\,, \tag{12.6}$$

where in the X-ray regime $\xi_p \ll 1$ always holds.

In the following sections we first discuss the emission characteristic and the energy spectrum of a single transition interface between two media. Thereafter we discuss one foil (two interfaces) and finally a sequence of many interfaces are considered.

12.1.3 Emission characteristic

In a transition between two media (e.g. air/Mylar) we label the respective plasma frequencies by ω_{p1} (air) and ω_{p2} (foil), constituting the only material parameters remaining for a single interface. The frequency ω (without index) denotes the frequency

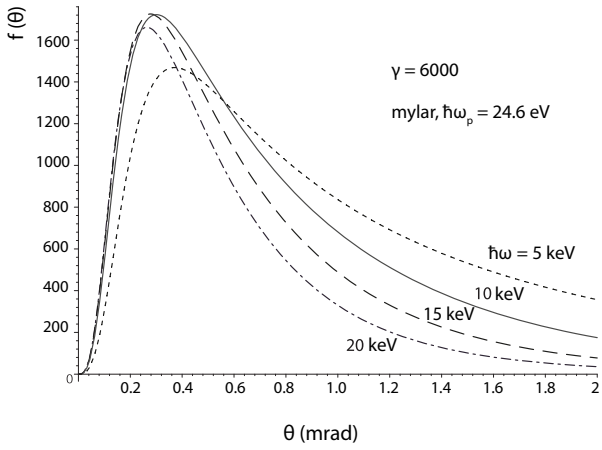


Fig. 12.3 Angular distribution of transition radiation according to (12.7), integrated over ϕ for an air/Mylar interface; $\theta =$ emission angle relative to the particle flight direction (adapted from [357]).

of the transition radiation photons and $\gamma = \frac{E}{m}$ always is the Lorentz factor of the particle passing the transition boundary.

For highly relativistic particles ($\gamma \gg 1$) and for $\omega \gg \omega_{p2} \gg \omega_{p1}$ the Ginzburg–Frank formula (12.3) can be approximated [119, 396] for small emission angles $\theta \ll 1$ as

$$\frac{d^2W}{d\omega d\Omega} \approx z^2 \frac{\hbar\alpha}{\pi^2} \theta^2 \left(\frac{1}{\gamma^{-2} + \xi_{p1}^2 + \theta^2} - \frac{1}{\gamma^{-2} + \xi_{p2}^2 + \theta^2} \right)^2 , \tag{12.7}$$

with the abbreviations $\xi_{p1} = \omega_{p1}/\omega \ll 1$ and $\xi_{p2} = \omega_{p2}/\omega \ll 1$. Using the approximation $d\Omega \approx \theta\, d\theta\, d\phi$ and after integrating over the azimuth ϕ, (12.7) becomes

$$\begin{aligned} \frac{d^2W}{d\omega d\theta} &\approx 2z^2 \frac{\hbar\alpha}{\pi} \theta^3 \left(\frac{1}{\gamma^{-2} + \xi_{p1}^2 + \theta^2} - \frac{1}{\gamma^{-2} + \xi_{p2}^2 + \theta^2} \right)^2 \\ &= 2z^2 \frac{\hbar\alpha}{\pi} f(\omega, \theta) \,. \end{aligned} \tag{12.8}$$

Figure 12.3 displays the distribution $f(\omega, \theta)$ for different photon energies $\hbar\omega$ and for an air/Mylar interface. Transition radiation is emitted azimuthally symmetric in a cone without a sharp boundary around the flight direction of the traversing particle, as illustrated in fig. 12.2. Different to Cherenkov radiation (chapter 11) the emitted radiation does not have a fixed emission angle but follows the distribution given by (12.7). The intensity reaches a maximum at an angle determined by the particle's Lorentz factor γ and the plasma frequencies ω_{p1}, ω_{p2} of the media involved and falls steeply off towards larger angles. If, as often is the case, $\omega_{p2} \gg \omega_{p1}$, then the most probable emission angle peaks approximately at [305]

$$\theta_{\mathrm{mpv}} \approx \sqrt{\gamma^{-2} + \xi_{p1}^2} \approx \frac{1}{\gamma} \,, \tag{12.9}$$

where the last approximation holds if $1/\gamma \gg \xi_{p1}$ which is the case for example for hard X-rays (larger than some 10 keV) and $\gamma \gtrsim 1000$. The relation $\theta_{mpv} \approx 1/\gamma$ can be regarded as representing the typical emission angle of transition radiation. The spectrum roughly extends until θ^2 becomes larger than $1/\gamma^2 + \xi_2^2$ making the difference

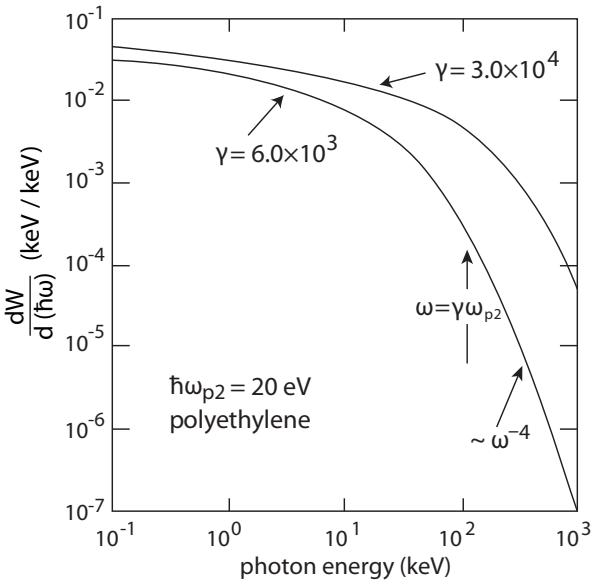


Fig. 12.4 Spectral energy distribution of transition radiation for polyethylene and two different particle energies γ (adapted from [357] with kind permission of Elsevier).

in (12.8) decrease faster to zero.

Equations (12.7) and (12.8) also hold for inclined incidence of the particle onto the interface [444] as long as the angle θ is defined with respect to the direction of the passing particle and not to the vertical of the boundary.

12.1.4 Energy spectrum

Integrating (12.7) over Ω yields the energy spectrum [306, 119]:

$$\frac{dW}{d(\hbar\omega)} \approx z^2 \, \frac{\alpha}{\pi} \left[\left(\frac{\omega_{p2}^2 + \omega_{p1}^2 + 2\omega^2/\gamma^2}{\omega_{p2}^2 - \omega_{p1}^2} \right) \ln \left(\frac{\omega_{p2}^2 + \omega^2/\gamma^2}{\omega_{p1}^2 + \omega^2/\gamma^2} \right) - 2 \right] . \quad (12.10)$$

Figure 12.4 displays the spectrum for a polyethylene/vacuum interface. Depending on the ratio $\omega/\gamma\omega_p$ we recognise three regions:

(1) $\omega \ll \gamma\omega_{p1}$: In this region (typical X-ray region from about 0.1 keV to 10 keV) there is only weak dependence on ω. On a very rough scale the intensity is about constant and for $\omega_{p2} \gg \omega_{p1}$ given by

$$\frac{dW}{d(\hbar\omega)} \approx 2z^2 \, \frac{\alpha}{\pi} \left[\ln \left(\frac{\omega_{p2}}{\omega_{p1}} \right) - 1 \right] . \quad (12.11)$$

The extension of this almost flat region increases with increasing γ of the radiating particle.

(2) $\gamma\omega_{p1} < \omega < \gamma\omega_{p2}$: The intensity increases logarithmically with γ and decreases logarithmically with ω:

$$\frac{dW}{d(\hbar\omega)} \approx 2z^2 \, \frac{\alpha}{\pi} \left[\ln \left(\frac{\gamma\omega_{p2}}{\omega} \right) - 1 \right] . \quad (12.12)$$

(3) $\omega \gg \gamma\omega_{p2}$: The intensity increases with the fourth power of γ and falls with the fourth power of ω:

$$\frac{dW}{d(\hbar\omega)} \approx z^2 \frac{\alpha}{6\pi} \left(\frac{\gamma\omega_{p2}}{\omega}\right)^4 . \tag{12.13}$$

The ω^{-4} dependence of (12.13) becomes evident when writing the square bracket of (12.10) as $(1+\frac{2}{x})\ln(1+x)-2$ for small $x = \gamma^2\omega_{p2}^2/\omega^2$ and assuming $\omega_{p1} \ll \omega_{p2}$ in (12.10); the logarithm is expanded to second order.

For a single transition the quantity $\gamma\omega_{p2}$ is thus an upper frequency bound for the intensity yield.

Equation (12.10) also exhibits the γ dependent characteristics of transition radiation. The intensity is low as long as the term in square brackets is small. It becomes sizeable only for $1/\gamma^2 \lesssim \omega_{p1,p2}^2/\omega^2$ or $\gamma \gtrsim \omega/\omega_{p1,p2}$, hence constituting an effective γ threshold for detection depending on the energy of the radiation: for $\hbar\omega = 10\,\mathrm{keV}$ and $\hbar\omega_{p2} \approx 20\,\mathrm{eV}$, for example, $\gamma \gtrsim 1000$ is needed. Although the radiated intensity, when integrated over all frequencies, linearly increases with γ (see next section), for TR detection the ω range is usually limited, resulting in a logarithmic γ dependence according to (12.12). The intensity eventually saturates for $\gamma \gg \omega/\omega_{p1}$ (eq. (12.11)), which is also evident from the fact that the terms with γ in (12.10) become negligible for large γ.

12.1.5 Photon yield

Integration of (12.10) over $\hbar\omega$ yields the total radiated energy of transition radiation for a single interface:

$$W = z^2 \frac{\hbar\alpha}{3} \gamma \frac{(\omega_{p2}-\omega_{p1})^2}{\omega_{p2}+\omega_{p1}} \approx \frac{z^2\alpha}{3}\gamma\,\hbar\omega_{p2} . \tag{12.14}$$

On the right-hand side of (12.14) $\omega_{p2} \gg \omega_{p1}$ has been used. The total intensity is directly proportional to the Lorentz factor γ of the particle and to the plasma frequency of the medium with the larger ω_p.

The sequence of the media at the interface is not important for the total intensity. The interface foil/air yields the same intensity as air/foil. The total radiated energy increases linearly with the Lorentz factor $\gamma = E/m$ of the particle, resulting from the fact that the photon spectrum decreases less steeply with ω for higher γ's (fig. 12.4).

To compute the average photon yield we assume that photons with an energy smaller than $\hbar\omega_{\min}$ are not detected, for example because of absorption or some detection threshold (typically in the keV range or even higher). For a foil/vacuum transition with foil plasma frequency ω_{p2} we obtain [746, 564]:

$$\langle N_\gamma\rangle_{\omega>\omega_{\min}} = \int_{\hbar\omega_{\min}}^{\infty} \frac{1}{\hbar\omega}\frac{dW}{d(\hbar\omega)}d(\hbar\omega) \approx z^2\frac{\alpha}{\pi}\left[\left(\ln\frac{\gamma\omega_{p2}}{\omega_{\min}}-1\right)^2+\frac{\pi^2}{12}\right] . \tag{12.15}$$

The number of radiated photons for a single transition is of the order $\alpha = \frac{1}{137}$, that is, very small. For a polyethylene foil with $\hbar\omega_{p2} \approx 20\,\mathrm{eV}$ and a threshold energy of 1 keV for the detection we obtain for a 5 GeV electron ($\gamma = 10^4$) an average number of photons of $\langle N_\gamma\rangle \approx 6.7\,\alpha \approx 0.05$. For the construction of a transition radiation detector one

therefore needs many (some hundred) transitions. They can be realised for example by a stack of foils or—as done in modern transition radiation detectors—as fibre structures or foams made of polypropylene or similar materials in which many (unordered) transition boundaries occur. To produce a net signal intensity the contributions from individual transitions must predominantly add constructively (see section 12.2.3 and example detectors in chapter 14).

If the momentum or energy of the radiating particle is measured independently of the TR measurement yielding an estimate of γ, the particle's mass and thus its identity can be inferred. In typical high energy physics experiments, transition radiation of sufficient intensity is only emitted by electrons. A completely linear increase of the emitted intensity with γ as suggested by (12.14) is in practice not given, since the detection capability of detectors depends on the energy of the TR photons (see also the comment on page 485 in section 12.1.3). The number of emitted photons above a given detection threshold increases logarithmically (see eq. (12.15)) and saturates at some value of γ, as discussed further in the following sections.

12.2 Multiple interfaces

To generate detectable TR intensity many subsequent interfaces along a particle trajectory are needed. In this section we first discuss two interfaces (one foil) and then multiple foils. If more than one interface is considered the emitted waves radiated from the different boundaries superimpose, leading to interference effects. A characteristic *formation length* relating the distance between the emission fronts and the phase relation of the emitted waves is used for the description.

12.2.1 Formation length

The emission of radiation is formed over some region corresponding to a formation time t_F over which the photon wave remains coherent with the electron wave. To illustrate this time consider a particle moving with $\vec{v} = \vec{\beta}c$ in a medium with refractive index n emitting waves at an angle θ with respect to its flight path (fig. 12.5). The formation time must be short compared to the time within which the phase difference between two wave trains emitted at different points along the path under the same angle θ develops:

$$\Delta\phi = |\phi_{t_1} - \phi_{t_2}| = |\omega t_F - k\, v t_F \cos\theta| = \left|\omega t_F - \frac{\omega n}{c}\, v t_F \cos\theta\right| , \tag{12.16}$$

where the phase factor at a given wave point $(\vec{r}, t)$ is $\mathrm{e}^{i(\vec{k}\vec{r}-\omega t)}$ and the two wave trains are emitted at times t_1 and t_2 with $t_F = t_2 - t_1$ as shown in fig. 12.5.

A characteristic length $Z = vt_F$ called *formation length* or *formation zone* [459, 936, 357] is defined (together with t_F) for a value[2] of the phase difference $\Delta\phi = 1$, for which the intensity of the interfering waves drops to $1/e$:

$$\left|\omega \frac{Z}{v} - \frac{\omega n}{c} Z \cos\theta\right| := 1 \tag{12.17}$$

$$\Rightarrow Z = \frac{\beta c}{\omega\,|1 - \beta n \cos\theta|} = \frac{\lambda \beta n}{2\pi\,|1 - \beta n \cos\theta|}$$

[2] Some arbitrariness in the definition of the formation length exists in the literature. Definitions with $\Delta\phi = 2\pi$ are used for example in [459], definitions with $\Delta\phi = 1$ (adopted here) can be found in [119, 357].

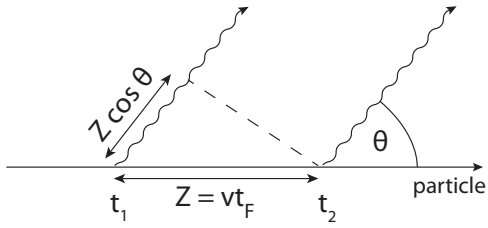


Fig. 12.5 Illustration for the definition of the formation length (see text) [459].

$$= \frac{c}{\omega \left| \left(1 - \frac{1}{\gamma^2}\right)^{-1/2} - \sqrt{1 - \xi_p^2}\,\cos\theta \right|} \approx \frac{2c}{\omega\,\left(\gamma^{-2} + \theta^2 + \xi_p^2\right)} \tag{12.18}$$

with in-medium wavelength $\lambda = 2\pi c/n\omega$ and refractive index $n = \sqrt{\varepsilon} = \sqrt{1 - \xi_p^2}$ with $\xi_p = \omega_p/\omega$ from (12.6). The root terms as well as $\cos\theta$ in the denominator of (12.18) have been Taylor expanded up to second order.

The formation length characterises the distance it takes until the field formed by the particle in its environment is readjusted in the new medium after the boundary, at which point in time the radiation field (TR photon) is separated from the mother particle. Its definition is analogous to that of other characteristic interactions lengths, such as the formation lengths l_f for bremsstrahlung (section 3.3.2, eq. (3.71)) or for the LPM effect (section 15.2.2.3 and appendix G, eq. (G.1)), respectively. The meaning of Z becomes particularly obvious for forward radiation ($\theta = 0$) in vacuum ($n = 1$). Within the time t_F the radiation leaves the particle behind by $\frac{\lambda}{2\pi}$, where the factor $\frac{1}{2\pi}$ is due to the definition (12.17).

Figure 12.6 shows the formation length Z from (12.18) at $\theta = 1/\gamma$ as a function of the photon energy for air ($\omega_p \approx 0.7$) and polyethylene ($\omega_p \approx 20$) as well as for different γ values. For large ω ($\omega \gg \omega_p,\ \xi_p \ll 1$) and for $\theta \approx 1/\gamma$ we have

$$Z \approx \frac{c\gamma^2}{\omega} = \frac{n\lambda\gamma^2}{2\pi} \tag{12.19}$$

and the maximum of Z is found at $\omega \approx \gamma\omega_p/\sqrt{2}$ assuming a value of

$$Z_{max} = \frac{2\sqrt{2}c}{\gamma\omega_p}\left(2\gamma^{-2} + \left(\frac{\sqrt{2}}{\gamma}\right)^2\right)^{-1} = \frac{c\gamma}{\sqrt{2}\omega_p} = 139.5 \times 10^{-9}\,\frac{\gamma}{\hbar\omega_p/\mathrm{eV}}\ \mathrm{m} \tag{12.20}$$

where $\hbar c = 197\,\mathrm{MeV\,fm}$ has been used. For highly relativistic particles ($\gamma \gg 1$) and with $n = \mathcal{O}(1)$ the formation length Z hence exceeds the wavelength $\lambda = 2\pi c/n\omega$ substantially. This has important consequences when multiple interfaces are considered for transition radiation detectors.

Typical values for $Z \approx Z_{max}$ lie in the region of some ten to some hundred micrometres: $Z(\text{polyethylene}) \approx 7\,\mu\mathrm{m}$, $Z(\text{air}) \approx 200\,\mu\mathrm{m}$ for $\gamma = 1000$ (2-GeV electrons). For thicknesses, respectively, distances smaller than Z the transition radiation yield is reduced, an effect referred to as *formation-zone effect* (see e.g. [119]). This effect (yield suppression for $l < Z$) occurs for both interface media, that is, usually the foil

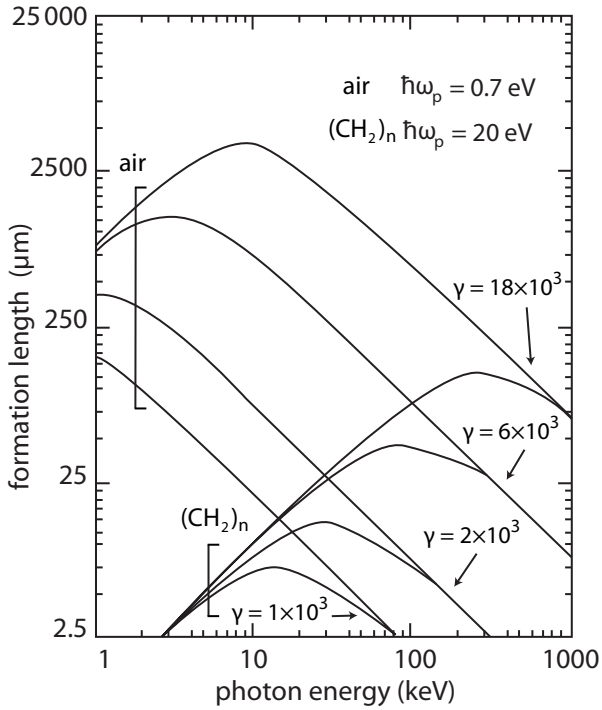


Fig. 12.6 Formation length for transition radiation as a function of the photon energy after (12.18) using $\theta = 1/\gamma$ for air and polyethylene $(C_2H_4)_n$ and for different γ values of the traversing particle. The parentheses show the regions of the formation length that are covered for air and for polyethylene for different γ values and typical photon energies. The formation length in air is about two orders of magnitude larger than in polyethylene (adapted from [357] with kind permission of Elsevier).

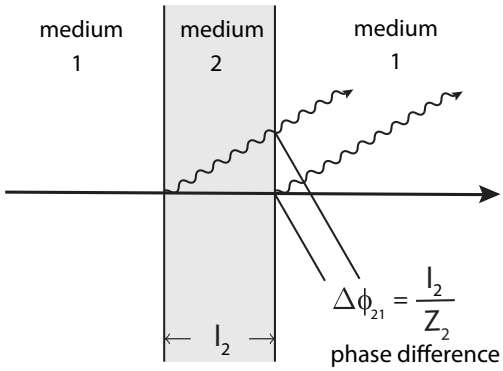


Fig. 12.7 Interference of transition radiation at the two boundaries of a foil (two interfaces).

thickness and the air gap [1017].

12.2.2 Transition radiation from two interfaces (one foil)

When a charged particle passes through a foil with two transition interfaces (e.g. air/Mylar and Mylar/air) the emitted wave trains interfere constructively or destructively depending on the phase difference between the waves emitted at the entrance and the exit of the foil (fig. 12.7). The interference depends on the frequency ω of the emitted radiation.

Since the formation length Z also depends on ω (and θ) the intensity yield of the oscillation can be expressed in terms of the formation length Z and the foil thickness l_2. Neglecting absorption in the foil itself, it is given by the yield of a single interface

multiplied by the interference term [119, 396]:

$$\left.\frac{d^2W}{d\omega d\Omega}\right|_{\text{foil}} = \left.\frac{d^2W}{d\omega d\Omega}\right|_{\substack{\text{single}\\ \text{interface}}} \times 4\sin^2\frac{\Delta\phi_{21}}{2}\,, \tag{12.21}$$

where the phase difference $\Delta\phi_{21}$ is given by the ratio of foil thickness and formation length (12.18):

$$\Delta\phi_{21} = \frac{l_2}{Z_2}\,, \tag{12.22}$$

which follows from the definition of the formation length Z in (12.17) as the length for which the phase difference $\Delta\phi$ between two waves emitted at an angle θ is one. Since—due to the inverted sequence of the dielectrics—the wave train at the downstream boundary has an additional phase shift of π relative to the wave train at the upstream boundary, destructive interference results for $\Delta\phi_{21} = 2n\pi$, while for $\Delta\phi_{21} = (2n+1)\,\pi$ constructive interference occurs. Furthermore, for thicknesses l_2 smaller than the formation length Z_2 the photon yield is suppressed (formation-zone effect).

For the opposite extreme case $l_2 \gg Z_2$ the interference term oscillates rapidly in ω at frequencies that are usually not resolvable by detectors. The yield for a foil then assumes an average value of about two times that of a single interface [306]. If l_2 is of the same order of magnitude as Z_2 the interference maxima are important for the yield, in particular the lowest order maximum ($n = 0$). This has to be taken into account when building a transition radiation detector.

If we neglect (for not too large ω) in (12.18) θ^2 and $1/\gamma^2$ compared to $\xi_p^2 = \omega_p^2/\omega^2$, we obtain

$$\sin^2\left(\frac{\Delta\phi_{21}}{2}\right) \approx \sin^2\left(\frac{l_2\omega_{p2}^2}{4c\,\omega}\right) \tag{12.23}$$

and can identify the yield maxima at frequencies ω_n:

$$\frac{l_2\omega_{p2}^2}{4c\,\omega_n} = (2n+1)\frac{\pi}{2} \quad \Rightarrow \quad \omega_n = \frac{l_2\omega_{p2}^2}{2\pi c}\,\frac{1}{2n+1}\,, \qquad n = 0, 1, 2, \ldots \tag{12.24}$$

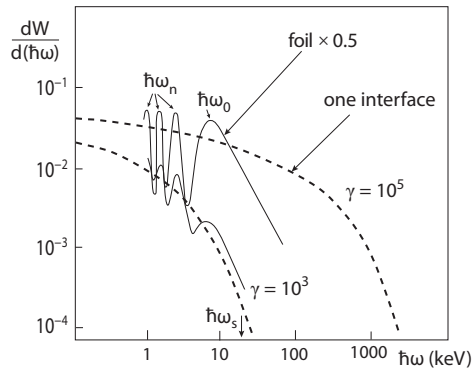
where ω_{p2} again is the plasma frequency of the foil.

The maximum at the highest frequency corresponds to $n = 0$. It is the most important one because the photons occurring at lower frequencies ($n > 0$) are absorbed more strongly in the foil (in particular if there are multiple foils, see next section). The upper frequency bound for a single interface $\gamma\omega_{p2}$ derived from (12.13) on page 485 is therefore replaced by a new effective upper bound $\omega_s = \pi\omega_0 = l_2\omega_{p2}^2/2c$ [396]. This new bound lies roughly half the width[3] of the curve of the 0th maximum higher than its peak value and is independent of γ, but dependent on the foil thickness l_2. From a certain value $\gamma = \gamma_s$ onwards, this new frequency bound starts to be lower than the frequency bound for a single interface $\gamma\omega_{p2}$, that is, when $\gamma_s\omega_{p2} = \pi\omega_0$ or

$$\gamma_s = \frac{l_2\omega_{p2}}{2c}\,. \tag{12.25}$$

Hence γ_s represents a saturation point for the transition radiation yield [396] beyond which the yield only increases logarithmically (dashed curve in fig. 12.8(b)). For

[3]The factor π is an arbitrary factor to account for the decrease of the resonance curve.



(a) TR-photon spectrum. (b) TR intensity as a function of γ.

Fig. 12.8 (a) Calculated TR photon spectrum [396] for a polypropylene foil (20 µm thick) in helium. Shown is the yield of a single interface (dashed line) compared to the yield of a foil normalised to one interface (solid lines) for two different values of γ. (b) Total radiation intensity as a function of γ shown for one interface, for a foil (two interfaces), as well as for a stack of 400 foils (20 µm thick) spaced by 180 µm helium gaps with and without absorption (adapted from [396]). Saturation is evident starting at γ_s (foil) and $\gamma_{s|N}$ (N foils, see text and eq. (12.30)). Perhaps surprising is the fact that saturation for many interfaces differs from a naiv superposition assumption, even for the case without absorption.

polyethylene or similar foils with $\hbar\omega_{p2} \approx 20\,\text{eV}$, this saturation is reached for γ values above $\gamma_s \approx 50\,l_2/\mu\text{m}$. For a typical foil thickness of $l_2 = 20\,\mu\text{m}$ this corresponds to $\gamma_s = 1000$ or electron energies of only about 0.5 GeV.

Explicit measurements of γ are possible with so-called 'precision TRDs' exploiting the γ dependence of transition radiation between threshold and saturation, as employed for example in cosmic ray experiments (see e.g. [964] and fig. 16.6(b) in chapter 16). A measurable range of $500 < \gamma < 10^4$ has been achieved in [964] using comparatively thick foils (76 µm) with optimised spacings. Cosmic ray experiments also benefit from the z^2 dependence of the TR yield when detecting ionised high energy nuclei.

In particle physics experiments transition radiation detectors are mostly operated as threshold detectors (above or below an intensity threshold), usually distinguishing electrons with very high γ values in TR saturation (thus benefiting from high intensities) from other particles.

The formation-zone effect hence introduces bounds on ω of the radiation as well as on the particle's Lorentz factor γ. The upper frequency bound for a practically relevant TR yield is

$$\omega < \omega_s = \min\left(\gamma\omega_{p2},\, \gamma_s\omega_{p2} = \frac{l_2\omega_{p2}^2}{2c}\right) . \tag{12.26}$$

Towards lower frequencies (below ω_0) the intensity first falls over a larger range and rises again only towards the maximum of the next ($n = 1$) order peak (see fig. 12.8(a)).

Figures 12.8 and 12.9 illustrate the various consequences of the formation-zone effect. Figure 12.8(a) displays the yield of one interface in comparison to the yield of a foil (two interfaces) showing interference as calculated for a polypropylene foil with $l_2 = 20\,\mu\text{m}$ [396]. Figure 12.8(b) is the total intensity (integrated over angle and

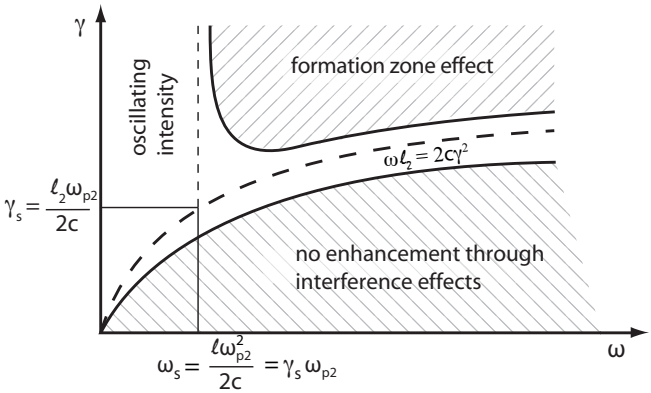


Fig. 12.9 TR intensity in the $\omega-\gamma$ plane: regions of high (light areas) and low (hatched areas) photon yield due to either the formation-zone effect or due to incoherent superposition of wave trains in the $\omega-\gamma$ plane (adapted from [119] with kind permission of American Physical Society).

energy) as a function of γ for one interface, a single foil ($l_2 = 20\,\mu$m), and a stack of 400 foils (polypropylene with $l_2 = 20\,\mu$m in helium at a distance of 180 μm), respectively. Saturation is evident already for one foil (two interfaces) and is even stronger for a stack (see next section). The proportionality of the TR intensity to γ in (12.14) thus only holds for a single interface, not for multiple interfaces. In addition the threshold behaviour due to absorption within the multiple foil stack is seen in fig. 12.8(b). The intensity sets in only for γ values of the order of 10^3 and above. This threshold effect and the saturation in TR intensity and respectively the number of detectable TR photons, are the most important consequence of the interference phenomena and the absorption of TR photons occurring for multiple interfaces. It must be paid attention to in the design of transition radiation detectors (section 12.3).

Figure 12.9 illustrates the yield dependence of the TR intensity in the ω–γ plane [119]. The hatched areas are regions of small photon yield. The dashed drawn curve corresponds to $l_2 = 2Z_2$ with Z_2 given in (12.19). For γ values larger than $\gamma_s = l_2\omega_{p2}/2c$ and $\omega > \omega_s$ (where Z_2 is larger than l_2) the yield is determined by the formation-zone effect; for high ω and low γ values the TR yield does not profit from the interference maxima (lower hatched area) as is evident also from fig. 12.8(a).

12.2.3 Transition radiation from many interfaces

To increase the TR yield to detectable intensities multiple transitions are needed. One possibility is to stack many foils with thickness l_2 and distance l_1 after each other where l_2 usually is as thick as several formation lengths Z_2 and the foils are separated by gas-filled gaps of thickness l_1 being (much) larger than Z_1. This arrangement leads to a further modulation of the photon spectrum compared to that of a single foil. In addition to the formation-zone effect limiting the yield as a function of photon energy $\hbar\omega$ as well as a function of γ, the energy dependent partial absorption of emitted radiation by the stack comes into play. This causes a further saturation effect for the detectable TR intensity when the number of foils is increased.

Apart from the condition for constructive interference for one foil (maxima at half integer multiples of the wavelength due to opposite phases at the two boundaries, cf. page 489 in section 12.2.2)

$$\frac{l_2}{Z_2} = (2n+1)\,\pi\,, \qquad n = 0, 1, 2, \ldots \tag{12.27}$$

we receive an additional condition for an air/foil system with multiple interfaces.

Neglecting absorption of the radiation at first, one obtains for N layers [306, 305] the total intensity

$$\frac{d^2W}{d\omega d\Omega} = \underbrace{\left.\frac{d^2W}{d\omega d\Omega}\right|_{\substack{\text{single}\\\text{interface}}} \times\, 4\sin^2\left(\frac{l_2}{2Z_2}\right)}_{\text{single foil}} \times \frac{\sin^2\left[N\left(\frac{l_1}{2Z_1}+\frac{l_2}{2Z_2}\right)\right]}{\sin^2\left(\frac{l_1}{2Z_1}+\frac{l_2}{2Z_2}\right)}\,, \tag{12.28}$$

where $Z_{1,2}$ are the formation lengths of the two media involved. The last term is the new interference term, in complete analogy to light diffraction on a slit compared to a grid in optics. Constructive interference occurs if

$$\frac{l_1}{Z_1}+\frac{l_2}{Z_2} = (2n+1)\,\pi\,, \qquad n = 0, 1, 2, \ldots\,. \tag{12.29}$$

Saturation sets in at γ_s for a single foil and somewhat higher at $\gamma_{s|N}$ for a foil stack; see fig. 12.8(b). As a function of γ a stack of foils develops a saturation in per-foil yield (with and without absorption) which quickly reaches a limiting value, in contrast to a single foil where the saturation (without absorption) is reached much more slowly. Therefore, maybe surprisingly, the photon yield of a stack cannot be regarded as a simple sum (not even just an incoherent sum) of interface contributions, independent of absorption effects. The position of the 0^{th} maximum for a stack remains about the same as in (12.26) for one foil, as can be verified by approximating $Z_i \approx 2c\omega/\omega_{p_i}^2$ as in (12.26) and assuming for the constructive interference condition in (12.29) that $l_1\omega_{p1}^2 \ll l_2\omega_{p2}^2$, even if $l_1 \gg l_2$. Since the stacking of foils is necessary to reach a measurable photon yield the observed saturation means that for γ values above $\gamma_{s|N}$ an energy measurement via transition radiation is effectively impossible [357].

An estimate of $\gamma_{s|N}$ for a stack of N foils for radiation energies near the 0th order maximum is derived in [305]:

$$\gamma_{s|N} \approx 0.6\,\omega_{p2}\frac{\sqrt{l_1\,l_2}}{c}\,, \tag{12.30}$$

where an explicit N-dependence does not appear.

Absorption of transition radiation in subsequent foils can be accounted for by an additional factor in (12.28) which can be computed assuming equal radiation emission at every foil boundary and subsequent exponential photon absorption with depth [305]:

$$\eta(\omega) = \frac{1-\mathrm{e}^{-N\,\sigma}}{N(1-\mathrm{e}^{-\sigma})} \approx \frac{1-\mathrm{e}^{-N\,\sigma}}{N\sigma}\,, \tag{12.31}$$

where $\sigma = \sigma(\omega) = \mu_1 l_1 + \mu_2 l_2$ with $\mu_{1,2}$ being the linear absorption coefficients of the two media (see table 12.1). Equation (12.28) then becomes

$$\frac{d^2W}{d\omega d\Omega} = \eta(\omega)\ \left.\frac{d^2W}{d\omega d\Omega}\right|_{\substack{\text{without}\\\text{absorption}}}\,. \tag{12.32}$$

Since the radiated intensity only weakly depends on the atomic number through ω_p ($\propto \sqrt{Z}$, eq. (12.5)), while photon absorption (dominated by photoeffect) has a very

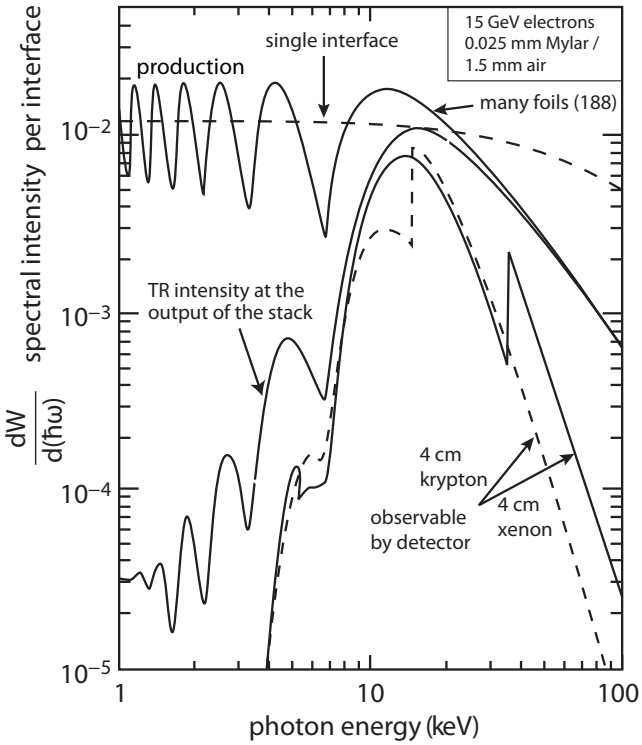


Fig. 12.10 Spectral intensity of transition radiation for a set-up consisting of a radiator and an X-ray detector, as for example shown in fig. 12.12. Shown is the produced intensity for a single interface (top dashed line) and for many foils (normalised to intensity per interface) without and with absorption within the foil stack (solid lines). The lower curves show the yield obtained by absorbing the TR photons in an X-ray detector filled with xenon (solid line) or krypton (dashed line), not accounting detecting efficiency. Adapted from [306], with kind permission of the American Physical Society.

strong Z dependence ($\approx Z^{4-5}$ for a given X-ray energy, see section 3.5.6), the radiator material should have low atomic number ($Z \approx 6$ for polyethylene).

For sufficient yield the foils must have a distance of at least the gap medium's formation length Z_1 (about 0.5 – 2 mm for air). For distances substantially larger than the air formation length the resulting yield is roughly the same as the (incoherent) sum of single foil photon yields. Absorption limits the number of foils useful for photon yield to [412, 357]:

$$N_{\text{eff}} = N\eta(\omega) \approx \frac{1 - \mathrm{e}^{-N\sigma}}{1 - \mathrm{e}^{-\sigma}} \tag{12.33}$$

using (12.31), which for $N \to \infty$ approaches saturation: $N_{\text{eff}} \to (1 - \mathrm{e}^{-\sigma})^{-1}$. Hence there is no benefit in using a number of foils for a transition radiation detector much larger than N_{eff}.

Figure 12.10 shows the TR spectral intensity for 15 GeV electrons traversing a stack of foils (25 μm Mylar/1.5 mm air). Both the production spectrum and the absorbed spectrum in a detector are shown (normalised per interface) for (i) one interface and for (ii) 188 foils [306] (see section 12.3). To detect transition radiation photons, detector materials with high photon absorption, that is, high atomic numbers, are needed. If wire chambers are used as transition radiation detectors, xenon and krypton are suitable gases. Apart from the typical oscillatory behaviour for multiple interfaces, fig. 12.10 illustrates how the detectable spectrum is damped both at the low photon energy end (due to absorption within the foil stack) as well as at the high energy end (due to the falling detection efficiency for X-ray photons), hence developing a maximum

at a photon energy of 10–20 keV. According to (12.24) the maximum ($n = 0$) is located at about

$$\omega_0 \approx \frac{l_2 \omega_{p2}^2}{2\pi c}\,, \tag{12.34}$$

assuming $\omega_{p2} \gg \omega_{p1}$ and $l_1 \gg l_2$, typical for transition radiation detectors.

12.3 Detectors for transition radiation

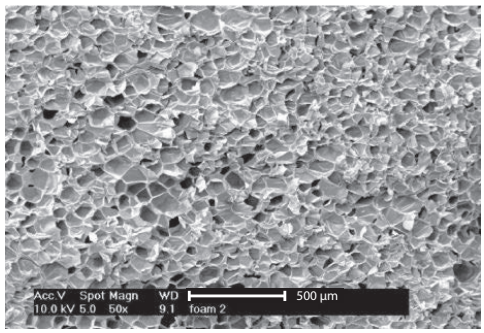
On the basis of the discussions in the previous sections we can now formulate the design criteria to be considered when transition radiation is to be utilised in particle detection.

12.3.1 General design characteristics

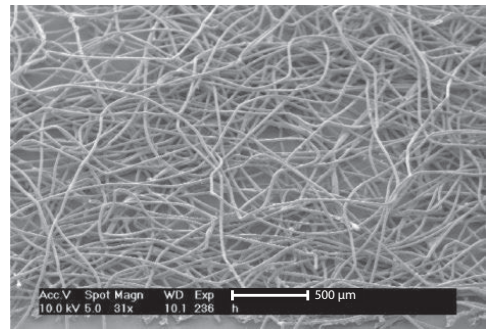
The design characteristics for transition radiation detection can be summarised as follows:

- The transition radiation intensity obtained from a single interface is insufficient to be exploited in a detector. Multiple interfaces are necessary.
- The total TR intensity for an interface rises with γ, but saturates for a foil and for arrangements with many interfaces starting at a saturation value[4] γ_s as a consequence of the occurring interference effects (see fig. 12.8(b)). The saturation occurs at γ_s of the order of 10^3 and can only be slightly changed by other parameters (like distance and thickness of the foils) without causing other negative effects for the yield. An example of such an effect is too strong spectral hardening, that is, suppression of the lower energy part of the emission spectrum (e.g. by absorption in the radiator). A very hard transition radiation X-ray spectrum on the other hand suffers from low detection efficiency in transition radiation detectors. Absorption within the foils further reduces the yield (see fig. 12.8(b)). The effectively usable γ range is limited to values around γ_s and above, for $z = 1$ particles maximally one order of magnitude below. Depending on the application the γ range can be larger when the TR yield is increased, for example for nuclei ($z \gg 1$), and when γ_s is tuned to high values (see the 'precision TRD' example mentioned in section 12.2.2 on page 490).
- In order to obtain a sufficiently large number of TR photons to identify and discriminate a radiating particle as an electron, a sufficient number of produced X-ray photons must be obtained that are not absorbed in the radiator (high energies preferred). They must then be detected (absorbed, low energies preferred) by an X-ray detector (see fig. 12.10). These requirements, together with the interference maxima and minima and the steep intensity decrease ($\propto \omega^{-4}$) at high ω, cut out a usable region in the γ–ω plane of transition radiation. The usable spectral range lies around the energetically highest lying 0th order maximum (typically 5–20 keV) and the γ range is around and above γ_s as illustrated in figs. 12.10, 12.8(b) and 12.9.
- These conditions correspond to a soft effective threshold ('threshold TRD') $\gamma_{th} \approx \gamma_s$ for transition radiation detectors (see fig. 12.8(b)), only beyond which detection of transition radiation becomes practically possible.

[4] Here γ_s denotes and is equal to $\gamma_{s|N}$ of (12.30) for saturation in the multi-interface situation.



(a) Rohacell foam.



(b) Polypropylene fibre.

Fig. 12.11 Electron microscopy pictures of Rohacell foam and polypropylene fibres (see table 12.1 on page 482) which are often used in place of foil stacks to generate transition radiation (source: CERN/ALICE CDS record 628225).

12.3.2 Design parameters

Many interfaces are required for sufficient TR intensity. The regular foil stacks discussed so far are often geometrically inconvenient, in particular for collider detectors. Therefore the radiators are usually made using fibre- or foam-type polyethylene or similar materials (fig. 12.11) which can be integrated as filling material into tracking detectors. An example is the ATLAS TRT discussed in section 12.3.4. Fibres or foam materials have irregular dielectric/gas interfaces which, depending on the density, can on average satisfy the requirements on l_1 and l_2 such that some detectable TR photons are generated per particle track.

Important parameters for the radiator choice are foil, respectively (average) fibre thickness l_2 and (average) distance l_1 in an air or gas volume relative to the respective formation lengths $Z_{1,2}$ as well as the plasma frequency ω_{p2} of the foil/fibre/foam material:

Foil thickness: According to (12.34) the X-ray energy of the yield maximum is proportional to the foil thickness. Thicker foils lead to a harder and more penetrating photon spectrum. Note however, that photon absorption also increases with foil thickness. Typical foil or rather fibre/foam thicknesses are in the range 10–20µm.

Foil distance: The Lorentz factor at which saturation sets in, is proportional to $\sqrt{l_1 l_2}$ according to (12.30). The pore size in radiator foams and the average fibre distance in radiator fibres should be larger than the formation length Z_1 of the air/gas interspace to ensure that the yield does not drop too low. For example, for 20 µm thick Mylar foils at $\gamma = 10^4$ and with $\hbar\omega_0 \approx 10\,\text{keV}, Z_1 \approx 2.2\,\text{mm}, \gamma_{s|N} \approx 20\,000$, one should choose a distance of $l_1 \approx 5\,\text{mm}$. For smaller values of γ the formation length Z_1 shrinks and smaller distances l_1, better suited for fibres/foams, can be chosen.

Material: The difference between plasma frequencies of the interface media involved should be large (intensity $W \propto \Delta\omega_p$, eq. (12.14)), effectively asking for a large ω_{p2}. Large plasma frequencies correspond to large electron densities and hence also to large material densities ρ, causing again enhanced photon absorption. To reduce self absorption in the radiator, materials with low Z are preferred. As typical ra-

diators for transition radiation detectors lithium or Mylar foils, or polypropylene fibres/foams are often chosen materials.

With a suitable choice of these parameters one can roughly reach a photon yield—omitting the detection efficiency—of N_{eff} times the yield of a single foil where N_{eff} is the effective number of foils according to (12.33).

The detectable transition radiation yield depends on many parameters. For the final design of TR detectors simulation computations are indispensable. In [316] and in [119, 357] some practical formulae for optimised detector design are given. The photon yield integrated over angle of a foil stack with N foils including absorption is [388]:

$$\frac{dW^{(N)}}{d(\hbar\omega)} = \frac{4\alpha}{\sigma(\kappa+1)}\left(1-\mathrm{e}^{-N\sigma}\right) \times \sum_n \theta_n \left(\frac{1}{\rho_1+\theta_n} - \frac{1}{\rho_2+\theta_n}\right)^2 (1-\cos(\rho_2+\theta_n)) \,, \tag{12.35}$$

where

$$\rho_i = \frac{\omega l_2}{2c}(\gamma^{-2}+\xi_i^2)\,, \qquad \kappa = \frac{l_1}{l_2}\,, \qquad \theta_n = \frac{2\pi n - (\rho_2+\kappa\rho_1)}{1+\kappa} > 0$$

and σ is the absorption as defined in (12.31). The index n refers to the maxima of the spectrum.

In [119] is shown that a description of the TR yield is possible using the dimensionless variables

$$\Gamma = \gamma/\gamma_s\,, \qquad \nu = \frac{\omega}{\gamma_s\omega_{p2}}\,, \qquad \kappa = \frac{l_1}{l_2}\,,$$

with γ_s from (12.25). A high yield radiator should have values of $\Gamma \approx 1$, $\nu \approx 0.3$ and $\kappa \approx 5$–20.

12.3.3 TR detector concept

Most commonly, TR photons that exit the radiator are detected by highly absorbing (high Z) gas-filled multiwire chambers (e.g. drift chambers, fig. 12.12(a)) or an arrangement of gas-tubes (see below). The radiating particle also traverses the chamber, thereby leaving an ionisation signal. The angular distribution of radiated photons follows fig. 12.3 with a maximum at $\theta_{max} \approx 1/\gamma$ (eq. (12.9)). For typical Lorentz factors of $\gamma > 1000$ and correspondingly small θ_{max} in the range of milliradians a spatial separation of the transition radiation signal and the track signal is usually not aimed for. Transition radiation is detected solely by identifiable additional charge deposition adding to the charge signal from the ionising particle track. Figure 12.12(b) shows a possible time sequence of signals seen on a wire of the detecting chamber from longitudinally drifting charges generated by deposited energy in the chamber gas. Observed pulses originate from ionisation charges including highly ionising δ-electrons (see section 3.2.2), from TR-photon absorption, as well as from noise.

On the other hand, one can profit from this simultaneous detection of transition radiation and ionisation signals by combining tracking capability and electron identification in the same detector. For electrons TR identification is easiest due to their very different γ value compared to hadrons at the same momentum. In cosmic ray experiments transition radiation is also exploited to identify high energetic nuclei and

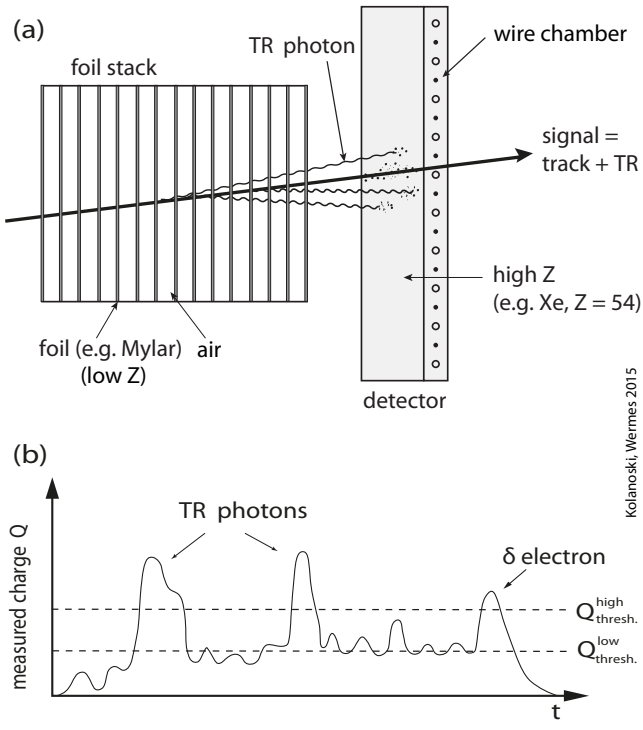


Fig. 12.12 (a) Schematic arrangement of a transition radiation detector. The particle's ionisation charge as well a the TR X-ray radiation are detected. The radiation is distributed cone-like about the particle track with small opening angle (cf. fig. 12.2). The detector chosen here is a multiwire chamber with a large drift volume, filled with a heavy gas. (b) A scenario of a pulse sequence, originating from longitudinally drifting charges, on a wire showing the occurrence of enlarged pulse heights by transition radiation that can be recognised by a high threshold. Also δ-electrons can cause large signals.

measure their energies (i.e. γ) [64] (see also section 16.2.1 and fig. 16.6). The large collider experiments ZEUS [170] at HERA as well as ATLAS [4,125] and ALICE [751] at LHC had or have such combined transition radiation and tracking detectors (see below and section 14.2.4).

Spatial separation of transition radiation signals from ionisation signals of the track is possible for distances at the detector face larger than about 50–200 µm. The separation distance can be increased for example by a magnetic field deflecting the track. This has been shown for demonstration purposes in [287] using a wire chamber or in [435] by using a highly segmented ($\sigma_x < 5\,\mu m$) pixel detector (see section 8.10.2).

Figure 12.13 demonstrates how multiplication of a several TR units, containing a foil stack and a TR detection chamber, filled with xenon or krypton, improves π/e separation. Since pions do not generate transition radiation and leave only an ionisation signal in the chamber, the total signal is larger for electrons than for pions. The electron spectrum becomes narrower with increasing number of TR units.

12.3.4 The transition radiation tracker of ATLAS

In the ATLAS experiment the *transition radiation tracker* (TRT) [4] serves as tracking detector in the outer region of the inner detector (ID) and can in addition separate electrons from hadrons (especially pions) by detecting TR photons covering a track momentum range from 1 GeV/c up to 200 GeV/c.

The TRT detector (fig. 12.14) consists of drift tubes, arranged cylindrically in the central (barrel) region and radially in the forward (endcap) region, of 4 mm diameter surrounded by 19 µm thick polypropylene fibres (in the barrel region), respectively, foils (in the endcaps) as radiators. The foils are 15 µm thick and are kept at an average

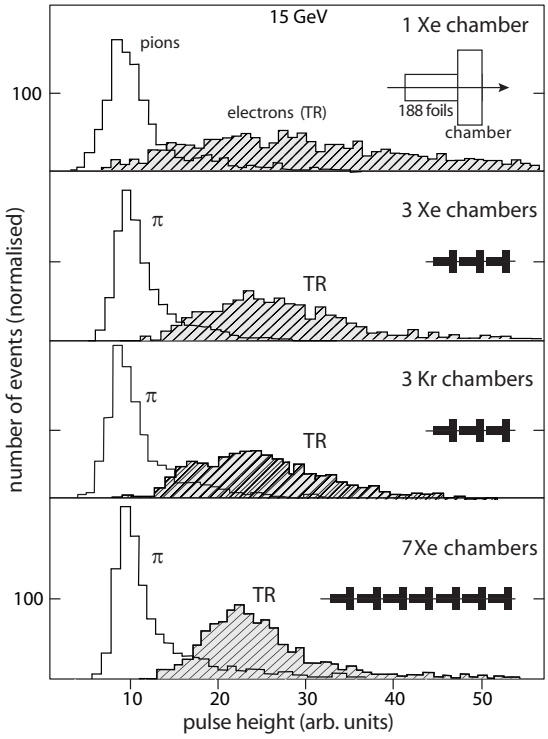


Fig. 12.13 Pulse height spectra obtained from 15 GeV pions and electrons, respectively, for different configurations of foil stack (188 foils, 25μ thick, 1.5 mm air) and gas-filled chambers (Xe or Kr). By using more than one TR unit (foil stack and photon detector, marked by the corresponding symbol) a narrower distribution for electrons is obtained and hence a better e/π separation (adapted from [307] with kind permission of Elsevier).

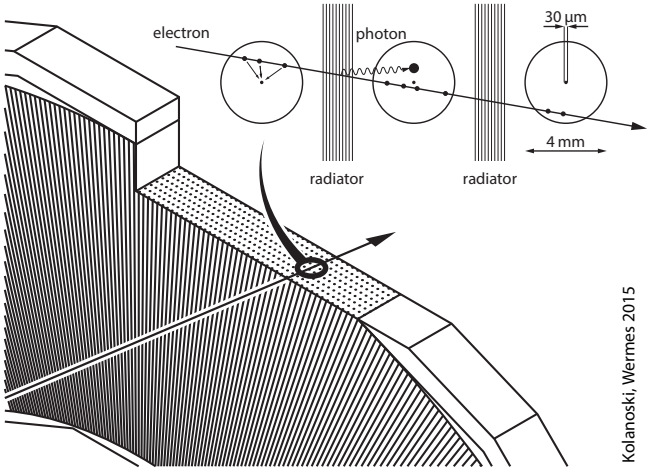


Fig. 12.14 Schematic representation and detection principle using drift tubes and transition radiation radiators in the *Transition Radiation Tracker* (TRT) of ATLAS. The drawing shows geometry and arrangement of drift tubes in the endcaps.

distance of 250 μm by a grid of polyimide fibres [4]. The tubes are filled with a mixture of Xe–CO_2–O_2 (70%, 27%, 3%). In the central region of ATLAS a track always traverses at least 36 drift tubes. Additional TR signals are electronically distinguished from 'normal' ionisation signals by defining two thresholds (low/high) (fig. 12.14). While δ-electrons (see section 3.2.2) also generate large signal pulses, a large number of hits having passed the upper threshold allows the conclusion that with high probability the particle track was an electron.

The achievable separation with TRDs, in particular between electrons and pions, is described in chapter 14, section 14.2.4.

12.3.5 TR detectors in cosmic ray experiments

Transition radiation is also exploited in cosmic ray experiments on satellites or balloons. For example, a 'precision TRD' (see also section 12.2.2 on page 490) is used in the balloon experiment CREAM [64], as described in section 16.2.1. It is used to measure ultra-relativistic ions with energies up to the 'knee' of the cosmic ray spectrum at about 10^{15} eV. Figure 16.6 on page 663 illustrates how the summed signal of transition radiation and ionisation (dE/dx) compared to the pure ionisation signal allows for a determination of Lorentz factor γ (from the TR signal) and charge (from the dE/dx-signal) of the ions and hence their energy and mass.

The AMS experiment (section 16.2.2 on page 666), which is installed at the International Space Station (ISS) [673], targets the search for antimatter in the universe. AMS employs a transition radiation detector mainly to separate positrons from protons and antiprotons from electrons.

13
Scintillation detectors

13.1 Overview

Detecting ionising radiation by observing scintillation light, generated in some materials upon impinging particles, is one of the oldest techniques for radiation detection. In 1910 Rutherford detected α particles with his eyes when they created scintillation flashes impinging on a zinc sulfide screen. Today scintillation photons are detected electronically.

By *scintillation* we understand the creation of luminescence by absorption of ionising radiation. *Luminescence* is light emission with a characteristic spectrum caused by certain inherent molecular or crystalline properties of materials. We distinguish inorganic crystal scintillators (e.g. NaI, $PbWO_4$, section 13.3) and organic materials discussed in section 13.2. Depending on the time delay of the light emission process we further distinguish *(prompt) fluorescence*, *phosphorescence* and *delayed fluorescence* (section 13.2.1).

Under *organic materials* we count organic crystals like anthracene or stilbene, but also liquids (e.g. xylene[1]) or organic scintillators dissolved in liquids as for example *p*-terphenyl ($C_{18}H_{14}$) dissolved in benzene (C_6H_6), sometimes also with the addition of a wavelength shifting material (see section 13.2.2.1 on page 507) like POPOP ($C_{24}H_{16}N_2O_2$). By polymerisation of the base material liquid scintillators can be solidified and can be formed into almost any geometrical shape. An example commonly employed has the scintillator molecule dissolved in a styrene monomer which is polymerised to form a solid plastic. These are called plastic scintillators (section 13.2.2). A recent review is, for example, given in [603].

Inorganic scintillators usually are crystals that generate scintillation light through transitions inside the lattice band gap, where photons with optical wavelengths can be created employing luminescent centres. Such centres which can be intrinsic, that is, molecular systems inside the lattice or molecular systems, or extrinsic, that is, created by doping, enable radiative transitions at optical wavelengths (see section 13.3). Examples are CsI(Tl) for the latter, where thallium is the dopant, and CeF_3 which is

[1] Also termed dimethylbenzene.

self-activated, meaning that oxy-anion complexes from lattice atoms form the luminescence centres. Inorganic scintillators are comprehensively described in [647].

Luminescence by ionising radiation. Charged particles can cause luminescence by ionisation with subsequent recombination in the scintillating medium or by molecule excitation with subsequent de-excitation into the ground state under emission of light. The recombination can occur directly or as a result of a reaction chain.

In the detection of photons in the X-ray (keV) or gamma energy range (MeV) scintillation light is created indirectly by electrons having been generated in the scintillation medium via the photoeffect, Compton effect or via pair creation and which for their part have transferred energy to atoms or molecules by ionisation or excitation.

Neutrons are detected indirectly if photons are created in (n,γ)-reactions or if protons are kicked off in (n,p)-reactions, generating light as described above.

Scintillators for particle detection. Scintillators for particle detection should ideally possess the following properties:

- The energy deposited in the scintillator should be converted into light with high efficiency.
- The light yield $L_S = \langle N_{ph} \rangle / E$, defined as the number of photons of a certain wavelength per deposited energy E, should be proportional to the energy released (linearity).
- The scintillation medium should be transparent for the wavelength of the scintillation light.
- The decay time of the luminescence, that is,r the duration of the excitation and light emission processes, should be as short as possible in order to obtain a fast signal pulse.
- The refractive index of the scintillating material should be close to that of the attached readout unit in order to ensure efficient light transmission. Photomultiplier glass tubes have indices of refraction near $n \simeq 1.5$.
- The light collection efficiency should be as large as possible.
- Depending on the application the scintillating medium should have high Z (e.g. for good photon absorption) or low Z and low atomic mass (e.g. for neutron detection).

Application areas of scintillation detectors on the one hand are energy measurement of photons, electrons or neutrons using totally absorbing detectors, and on the other hand the generation of fast signals for charged particles for timing purposes, which can for example be used in time-of-flight measurements or for fast trigger signals. For these two application areas different kinds of scintillator materials are used (see section 13.5): preferentially organic scintillators for time critical applications, inorganic crystals for energy measurement.

Figure 13.1 shows a typical arrangement of a scintillation detector consisting of the following components:

Scintillator: The scintillator converts ionising radiation into light. In order to prevent light from escaping the scintillator medium or entering from outside, the scintillator is made light-tight by a foil which is reflecting on the inside, for example an aluminium foil or diffusely backscattering white paper.

Light guide: The part (i.e. the areal cross section) of the scintillator through which the light must pass towards the photomultiplier often has a geometry very different from that of the photomultiplier entrance window. A light guide, most often made

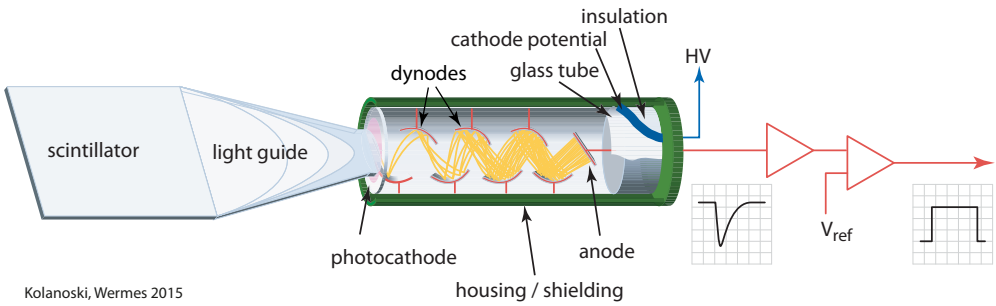


Fig. 13.1 Scintillator (here a plastic scintillator) with signal processing chain (photomultiplier, amplifier (if applicable), discriminator)

of plexiglass, can serve as an adapter in such cases when a direct connection is inefficient or is not possible.

Photomultiplier: The photomultiplier tube (PMT) converts scintillation photons into electrons in its photocathode and provides low-noise amplification via the subsequent *dynode* system. The electrons are amplified by a factor of up to about 10^{8-9} and generate a current pulse at the output (anode) of the PMT (see section 10.2.1). Alternatives to using PMTs also exist (see chapter 10).

Amplifier: The output pulse of the photomultiplier can in most cases be processed without further electronic amplification; only in applications with very low light output it must be further amplified in some cases. The signal is usually discriminated and passed on as a digital pulse to the backend electronics.

The entire arrangement of fig. 13.1 is enclosed in a light-tight wrapping, for example black tape. In order to shield against magnetic fields (even as weak as the Earth's magnetic field) the photomultiplier must be housed in a shielding cage, for example a μ-metal cylinder, a magnetically soft Fe–Ni alloy with high permeability. This is because the electron paths inside the dynode system are very sensitive to magnetic field effects and thus also is the PMT amplification.

Scintillation mechanism. The mechanism generating scintillation light depends on the type of the scintillator material and is very different for organic and inorganic scintillators. The absorbed energy leads to excitations of atomic/molecular states or crystal excitations, respectively, the de-excitation of which lead to light emission. The emitted light can be reabsorbed in the scintillator. The emission energies are however lower than or at most equal to the absorption energies (see fig. 13.4 in section 13.2 and fig. 13.9 in section 13.3). Therefore the emitted scintillation light is shifted to larger wavelengths, a phenomenon called *Stokes shift* (fig. 13.2). For detectors one aims for a high light yield with little self-absorption. The wavelength ranges for absorption and emission should hence overlap as little as possible.

13.2 Organic scintillators

13.2.1 Scintillation mechanisms in organic materials

The scintillation mechanism of organic materials is largely determined by the electronic structure of the carbon atom. To understand the emission mechanisms we therefore must study the energy level schemes of carbon compounds.

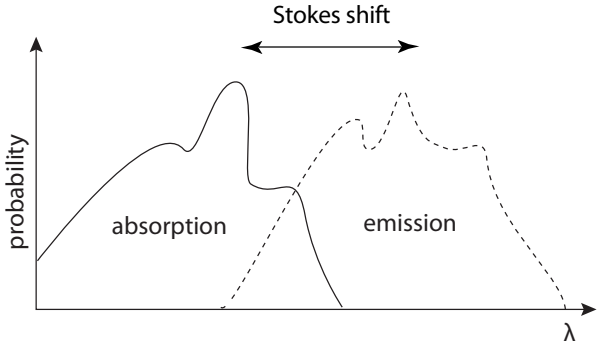
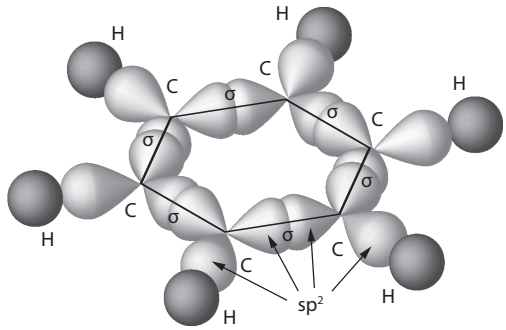
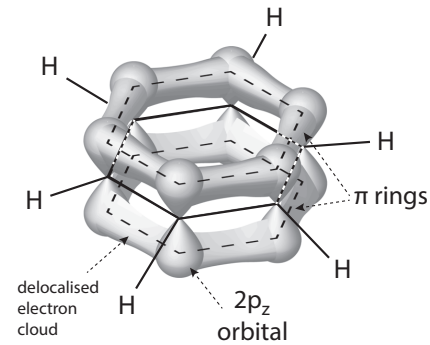


Fig. 13.2 Absorption and emission spectrum of scintillators (schematic).

The carbon atom has six electrons which fill the K-shell ($1s$) completely and partially the L-shells ($2s$ and $2p$). The configuration is: $1s^2$, $2s^2$, $2p^2$ where the superscripts indicate the number of electrons in that orbital. In compounds this configuration often changes into $1s^2, 2s^1, 2p^3$, which means that a $2s$ level is raised to the energy of a $2p$ state due to the influence of the other compound atoms. Three different types of orbital mixing configurations (hybridisations) are possible, depending on whether the created s–p mixed orbitals are energetically equal (sp^3 hybridisation) or if only two or one of the three $2p$ orbitals mix with the $2s$ orbital (sp^2 or sp hybridisation, respectively). Luminescent transitions can be found in materials with sp^2 or with sp hybridisations for which at least one p orbital remains unchanged compared to the unbound state of the atom. The electrons in this p level are called π electrons. The mixed-orbital electrons are called σ electrons. They form strong covalent bindings. The more weakly



(a) σ hybrid orbitals.



(b) π orbitals.

Fig. 13.3 Electron orbitals of carbon in benzene (C_6H_6). The light and dark grey bubbles represent electron orbitals. Each carbon atom is sp^2 hybridised and forms σ bonds with two adjacent carbon atoms (six C–C bonds) and another σ bond with the hydrogen atoms (six C–H bonds). (a) Sketch of only σ orbitals which are arranged in a planar hexagonal structure and are tightly bound, since their orbitals overlap strongly. (b) π orbitals with the σ orbitals kept simplified as a line structure. The six $2p_z$ orbitals of carbon are orthogonal to the σ orbitals and from (three) π bonds. Since the π orbitals also extend to their neighbours (delocalisation) the π bonds form hexagonal rings in which the electrons can move freely.

bound π electrons are in orbitals orthogonal to the σ orbitals. Three-dimensional orbital structures are illustrated for the example of the benzene molecule in fig. 13.3.

Responsible for luminescence are the transitions between the molecular π electron levels. In fig 13.4 a π-level spectrum of an organic molecule is drawn which features two level schemes that are almost independent of each other (singlet and triplet). The main levels are split into sub-levels corresponding to different vibration modes. The energy distances between the main levels S_0 and S_1 lie in the range 3–4 eV corresponding to 400–300 nm.

Absorption of energy in the scintillator leads to an excitation from the ground state S_{00} to higher levels (S_{1i}, S_{2i}, S_{3i}, ...). De-excitation from S_{1i} to S_{10} levels is radiationless and takes place in the picosecond range. The same holds for S_{2i} to S_{20} transitions and so forth. Transitions from there to the ground levels S_{0i} takes place at the nanosecond time-scale and is called *prompt fluorescence*. This transition is most important for fast scintillation signals. Since the difference between energy levels for emission is smaller than the one for absorption (fig. 13.4) the emission wavelength is Stokes shifted relative to the absorption wavelength.

Transitions between singlet and triplet terms are spin forbidden for radiation. For a fraction of molecules the T_{1i} triplet states[2] can, however, be populated by radia-

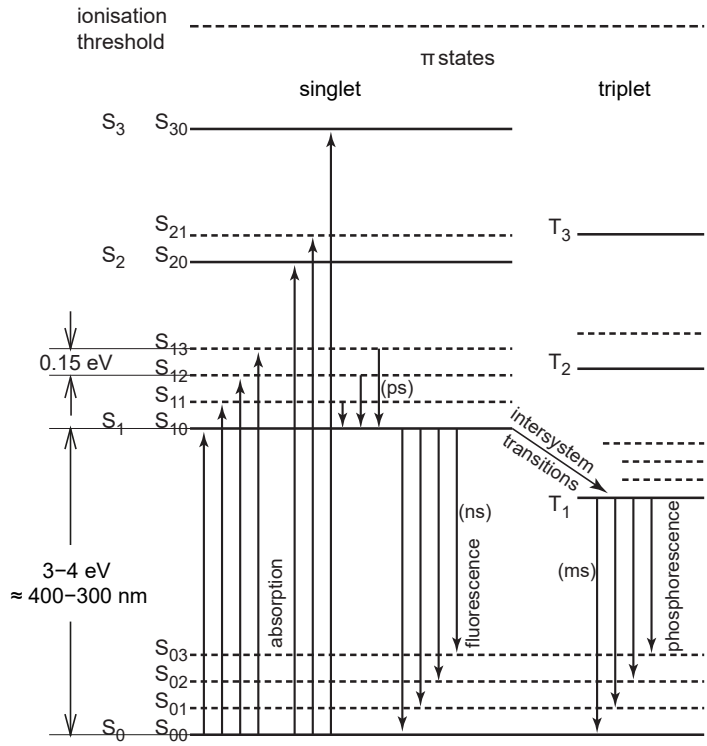


Fig. 13.4 Typical π energy levels in organic molecules (adapted from [206], with kind permission of Elsevier).

[2]The triplet states have a smaller radius in the *s* orbital due to the antisymmetric spin alignment. They are thus more strongly bound and lie energetically lower than the singlet states.

Table 13.1 Properties of some selected organic scintillators (after [1014,616]). λ_{max} denotes the wavelength of maximum emission, τ_{dec} the decay constant of the transition.

	Commercial name	Density (g/cm^3)	λ_{max} (nm)	Photons per MeV	τ_{dec} (ns)
anthracene $C_{14}H_{10}$		1.25	447	16 000	29
stilbene $C_{14}H_{12}$		1.16	410	8 000	5
naphthalene $C_{10}H_8$		1.14	348	<5 000	11
plastic scintillator	NE102	1.032	423	10 000	2
plastic scintillator	BC-420	1.032	391	10 500	1.5
liquid scintillator	EJ-301, BC-501A	0.874	425	12 500	3.2

tionless transitions (inter-system transitions). The T_{1i} states are metastable and the subsequent de-excitations to the ground state S_{0i} take place in the millisecond range (*phosphorescence*).

During the time that the metastable T-levels are occupied, their electrons can—by another energy take-in (for example thermal or by a second ionisation)—undergo transitions back to the S-levels, followed by immediate de-excitation. This on average leads to *delayed fluorescence* (µs to ms).

Phosphorescence and delayed fluorescence should be avoided if possible. In very pure organic materials as well as in inorganic crystals (see section 13.3) they are both suppressed.

13.2.2 Organic materials for scintillation detectors

Organic scintillators in solid, crystalline form are naphthalene ($C_{10}H_8$), anthracene ($C_{14}H_{10}$) and stilbene ($C_{14}H_{12}$). Of these anthracene crystals have the largest light yield with about 15 000 optical ($\lambda \approx 300$–700 nm) photons per MeV deposited energy. They are, however, brittle when handled and difficult to obtain in large shapes. In addition, the scintillation efficiency depends on the direction of incidence of the ionising particle with respect to the crystal axis orientation [616].

Dissolved in suitable solvents, called aromatic compounds,[3] as for example benzene or xylene, many organic scintillators are available as liquids. Examples are *p*-terphenyl ($C_{18}H_{14}$), PBD ($C_{24}H_{22}N_2O$), DPO ($C_{15}H_{11}NO$), as well as POPOP ($C_{24}H_{16}N_2O_2$). Aromatic compounds are usually the scintillators by themselves, however with only low usable quantum efficiency. This means they feature fairly low light yield per deposited energy and yield only few photons at detectable wavelengths and suitably short decay times. Efficient scintillators hence need agents to be added to the base material, like the ones mentioned above. In polymerised form they are solids, called *plastic scintillators*, which can be put into almost any shape. Due to their excellent and easy handling they are often used in particle and nuclear physics experiments. Base materials most often are polystyrene, polyvinyl toluene or polymethyl acrylate with densities in the range 1.03–1.20 g/cm^3 [746]. A selection of often used organic scintillator materials, their properties and their commercial names are listed in table 13.1 (see also [1025]).

A typical organic scintillator is formed using a solvent doped with a high concentration of primary fluor (= fluorescence emitter) and a much lower concentration of

[3] In chemistry, the term *aromatic* specifies cyclic, planar molecule arrangements which are very stable.

a secondary fluor. The solvent acts as the initial energy absorber. The most efficient solvents all have a benzene molecule incorporated within their structure so that we can employ benzene for a discussion of the most important features. Referring to fig. 13.4 the primary excitation processes occurring when a charged particle passes an organic scintillator can be classified as follows [1025]:

(1) excitation of the π singlet states;
(2) ionisation of π electrons;
(3) excitation of other electron groups (such as σ electrons);
(4) ionisation of other electron groups.

The first process is predominantly responsible for the fast (fluorescent) component of the scintillation process. The second process, followed by ion recombination, seems to be responsible for the population of some triplet states and predominantly leads to the slow component in the scintillation process. Any excitation via the third process is dissipated thermally. The fourth excitation mode seems to be the fundamental cause of radiation damage to the material via ionising radiation.

13.2.2.1 Plastic scintillators

For a scintillator to be transparent the Stokes shift between emission and absorption wavelengths must be large. Also, for most photomultipliers the quantum efficiency maximum is located at larger wavelengths (typically with a maximum at 400 nm, see fig. 10.10 on page 418) than the maximum of the emission of the scintillating base material (the solvent). Large Stokes shifts can be achieved by addition of two or more agents to the base material.

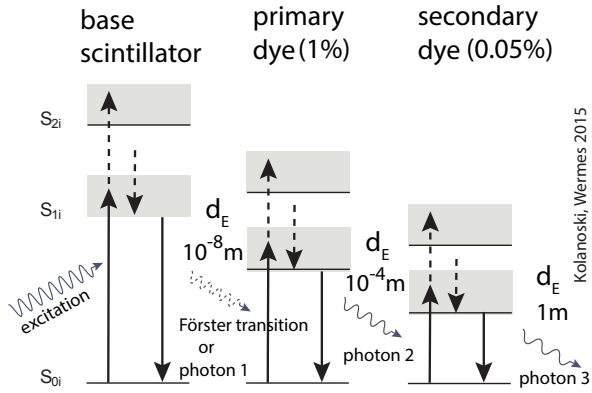
Most plastic scintillators use as the base scintillator polyvinyl toluene (PVT) or polystyrene (PS). An often used combination is PVT (as solvent and first scintillator), *p*-terphenyl (as primary agent and second scintillator) and POPOP (as secondary agent and third scintillator. POPOP has a large Stokes shift and acts as a *wavelength shifter* (WLS) in the dye combination facilitating larger light transmission lengths. Alternative combinations are put together in table 13.2.

The primary agent ensures a larger overall yield and a shorter rise time than the base material, provided it is added in a saturated solution (typical are mass concentrations of about 1% or more) and has good quantum efficiency as well as good spectral

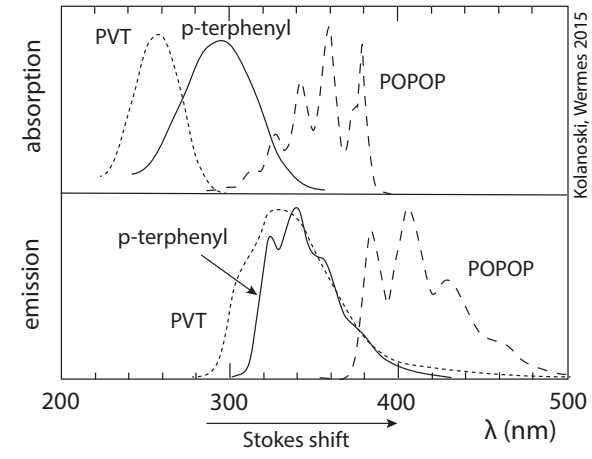
Table 13.2 Often used compositions of organic scintillators in liquid and plastic scintillators.

Scintillator	Base	Primary additive	Secondary additive
liquid	benzene	*p*-terphenyl	POPOP
	toluene	DPO	BBO
	xylene	PBD	BPO
plastic	polyvinyl benzene (PVB)	*p*-terphenyl	POPOP
	polyvinyl toluene (PVT)	DPO	TBP
	polystyrene (PS)	PBD	BBO/DPS

Abbreviations are as follows: PBD = $C_{24}H_{22}N_2O$; DPO = $C_{15}H_{11}NO$; BPO = $C_{21}H_{15}NO$; TBP = $C_{12}H_{27}O_4P$; BBO = $C_{27}H_{19}NO$; POPOP ($C_{24}H_{16}N_2O_2$); DPS = diphenylstilbene.



(a) Wavelength shifting: principle.



(b) Absorption and emission spectra.

Fig. 13.5 Wavelength shifting in a plastic scintillator (schematic) consisting of a base material (here polyvinyl toluene, PVT), a primary dye (here *p*-terphenyl) and a secondary dye (here POPOP): (a) principle of absorption and emission with increasing wavelength. The vertical arrows indicate excitations and de-excitations (drawn dashed when many transitions are subsumed). Also shown is the average spatial distance d_E between centres for the energy transfer to occur. The base to primary dye transition can be a fast Förster transition with small d_E. (b) Absorption and emission spectra of the three scintillator materials. The curves in (b) are not calibrated absolutely. The (linear) y axes have arbitrary divisions and units.

conformity, that is, the emission spectrum of the base material should overlap with the absorption spectrum of the primary agent. Then the mean distance between the molecules of base material and primary agent are in the range of about 10 nm, shorter than the wavelength of the emitted light. In this case radiationless dipole–dipole resonance interactions (Förster transitions [419]) can occur, the range of which, however, decreases with the sixth power of the intermolecular distance. It is the strongest and fastest coupling between base scintillator and agent and provides short rise times in the nanosecond range as well as—for maximal agent addition—a substantial yield increase such that attenuation lengths of 3–4 m can be achieved [1025]. In order to maintain sufficient transparency a secondary addition is necessary for wavelength shifting. A typical example is POPOP. The WLS concentration is small (typically about 0.05% in mass) to minimise self-absorption. In fig. 13.5(a) the principle of wavelength changes in organic scintillators is depicted.

Figure 13.5(b) shows absorption and emission spectra for a typical plastic scintillator composed of three often used scintillating components. One can see that base scintillator and primary agent strongly overlap in emission and absorption, respec-

tively, whereas the secondary agent emits in regions of larger wavelength. The Stokes shift obtained all together (horizontal arrow in fig. 13.5(b)) largely separates absorption and emission wavelength regions such that good transparency is achieved.

In cases where the scintillation process needs to remain local (and fast), for example in thin (< 1 mm) scintillating fibres (see section 13.6), often only the primary agent is added (only Förster transitions contribute to the light yield).

Usage. Plastic scintillators are mostly used to detect charged particles and neutrons (by detection of the proton after a (n,p) reaction). They stand out in terms of their simple handling, cheap fabrication to almost any desirable shape, and by their robustness and reliability. The spatial resolution of plastic scintillator detectors is limited by the size of the scintillator piece to areas of mm^2 to cm^2 at best. Better spatial resolutions can be achieved with scintillating fibre arrangements (see section 13.6).

As disadvantages one must list ageing effects and sensitivity with respect to some solvents and greases, as well as sensitivity to high temperatures and, especially, to ionising radiation. The light yield of scintillator material as well as the attenuation length for light transmission typically decreases after ionising radiation doses of a few kGy.[4] Therefore plastic scintillators or fibre trackers are usually not suited for the inner zones of hadron collider experiments. Far away from the interaction points, however, the radiation levels are less, such that scintillating fibres are attractive for small angle scattering measurements very close to the beam line (see section 13.6.2).

For the testing of plastic scintillators the radioactive β^+(EC) isotope $^{207}_{83}$Bi is a good choice (see table A.1 in appendix A.2). The decay scheme has two monochromatic electron lines at 975 keV and at 481 keV from inner conversion of the excited daughter nucleus $^{207}_{82}$Pb. These monochromatic electrons are typically absorbed by a few millimetre thick plastic scintillator and serve as calibration standards for a scintillation detector. Figure 13.6 shows a ^{207}Bi spectrum measured with a plastic scintillator. An

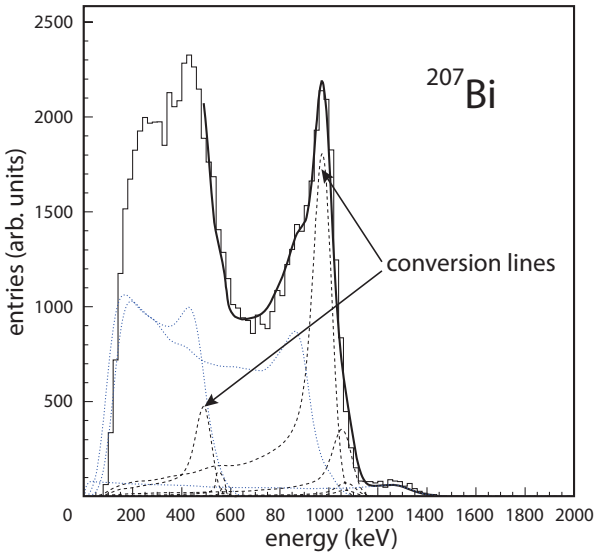


Fig. 13.6 Spectrum of the ^{207}Bi isotope measured with a plastic scintillator (PVT) [721]. Prominent is the monoenergetic electron conversion line at 975 keV which serves

[4] Special developments of radiation tolerant organic scintillators achieve values around 100 kGy, see for example [604].

absorbed energy of 975 keV corresponds to the energy deposit of a minimum-ionising particle traversing 5 mm of plastic scintillator.

13.2.2.2 Wavelength shifters

The term *wavelength shifter* (WLS) is used for the chemical agent providing a large Stokes shift in organic scintillators, but is also used for plastic plates or other geometrical shapes, including fibres, which absorb and re-emit (wavelength shifted) light with high efficiency. They can for example be used to improve the light collection in more complex or space restricted geometries (see sections 13.4.2 and 13.4.3). The isotropic emission distribution of a WLS can also be exploited to efficiently bring light emitted for example from a scintillation crystal onto the surface of a (small) photodiode and—at the same time—adapt the wavelength to the diode's quantum efficiency coverage. Typical light collection efficiencies by means of wavelength shifters are only in the per cent range which, however, is acceptable for many applications, for example in calorimeters which provide ample amounts of scintillation light (see chapter 15 with the principle shown in fig. 15.27 on page 618). An example of an efficient blue (430 nm) to green (476 nm) shifting WLS which is often used in detector applications and for fibres is the so-called Y-11 dye in polyvinyl toluene (PVT) or polystyrene (PS), which has an attenuation length of more than 3.5 m.

13.2.2.3 Liquid organic scintillators

What has been said in the previous section also implies and holds for liquid scintillators. A primary agent is dissolved in an aromatic solvent to increase the quantum efficiency. The requirements on purity standards are high in order to avoid unwanted effects like the decrease of transparency or, most importantly, a degradation in light output. In particular oxygen in the solvent is an aggressive quencher, since energy can be transferred without radiation to oxygen molecules with a high cross section. Also water vapour destroys the scintillation properties. Liquid scintillators are hence operated under dry nitrogen atmosphere [1025].

13.2.3 Light yield

In organic materials UV photons are created in the ionisation process which generate the excitation transitions shown in fig. 13.4. The scintillation light yield by ionisation is much smaller (in the range of 5%) than by direct UV light excitation (50–90%) [206]. The remainder of the deposited energy is dissipated without radiation, mainly into heat. The typical light yield from ionisation is about one scintillation photon per 100 eV of deposited ionisation energy [746]. In a 1 cm thick plastic scintillator about 20 000 photons are generated by a minimum-ionising particle.

For the passage of charged particles, the light yield L (here the number of photons of a certain wavelength λ) of organic scintillators is to first order proportional to the specific ionisation density (see section 3.2) along the particle's path through the scintillator [206]:

$$\frac{dL}{dx} = S\frac{dE}{dx}\,, \tag{13.1}$$

with the scintillation efficiency S being defined by the proportionality factor. This linear behaviour practically holds for all minimum-ionising particles, but only approximately for slow protons or other heavy particles with momenta below 100 MeV. A

high ionisation density along the particle path leads to quenching effects, due to already excited or ionised molecules, leading to an overall reduction in light yield. This quenching again is proportional to the specific ionisation loss with quenching strength k_B. Putting both together the empirical *Birks formula* [206] is obtained:

$$\frac{dL}{dx} = \frac{S\frac{dE}{dx}}{1 + k_B\frac{dE}{dx}}\,, \tag{13.2}$$

which was originally proposed by Birks for organic scintillators, but also holds for inorganic crystals (see [206], chapter 11 and [132,133]). Extensions of (13.2) including higher order effects [326] contain an additional term in the denominator $(C\,(dE/dx)^2)$ with an empirical constant C to be determined by fits to measurements. For high energetic, singly charged particles with energies much larger than the rest mass of the particle as well as for gamma rays, we can assume that energy loss and ionisation density are small compared to the energy of the particle. Equation (13.2) then reduces to (13.1). For strongly ionising radiation, for example for α particles in the MeV range, the dE/dx term in (13.2) dominates such that the light yield becomes

$$\left.\frac{dL}{dx}\right|_\alpha = \frac{S}{k_B}\,. \tag{13.3}$$

The k_B value must be determined experimentally for every scintillator. This can be achieved for example by measuring the ratio of the light yields for electrons and for α particles:

$$k_B = \frac{(dL/dx)_e}{(dL/dx)_\alpha}\,\frac{1}{(dE/dx)_e}\,. \tag{13.4}$$

For the common scintillator type NE102 k_B is in the range of $k_B \approx 10^{-2}$ g cm^{-2} MeV^{-1} [206] where x in (13.4) is the *column density* of eq. (3.28), that is, dE/dx is density normalised, and S has an approximate value of 10 000 photons (at $\lambda \approx 423$ nm, blue) per MeV deposited energy.

The mean number of detected photoelectrons is given by the following expression:

$$\langle N_{pe}\rangle = \int_{\lambda_1}^{\lambda_2} S(\lambda)\,T(\lambda)\,Q(\lambda)\,d\lambda \approx \langle S\rangle\langle T\rangle\langle Q\rangle\,\Delta\lambda\,. \tag{13.5}$$

$S(\lambda)$ is the wavelength spectrum of fluorescence photons (integral is the number of photons produced per wavelength interval), $T(\lambda)$ denotes the light transfer and collection efficiency, and $Q(\lambda)$ is the quantum efficiency of the photodetector. The integration is carried out over the emission spectrum from λ_1 to λ_2. Often the terms contributing to $\langle N_{pe}\rangle$ only weakly depend on the wavelength such that the approximation on the right-hand side of (13.5) is valid. The average 'source strength' $\langle S\rangle$ is the product of average energy deposition of the primary particle and the number of photons per wavelength interval and per unit of deposited energy:

$$\langle S\rangle = \left\langle\frac{dE}{dx}\right\rangle\,\Delta x\,\left\langle\frac{\Delta N_\gamma}{\Delta E\,\Delta\lambda}\right\rangle\,. \tag{13.6}$$

13.2.4 Signal shape

The response in time of scintillating organic molecules is predominantly determined by prompt fluorescence exhibiting an exponential decay law for the intensity:

$$I(t) = I_0 \, \mathrm{e}^{-\frac{t}{\tau_{dec}}} , \tag{13.7}$$

where τ_{dec} is the decay constant (as given in table 13.1) for prompt fluorescence and $I(t)$ and I_0 are the light intensities at time t and $t = 0$, respectively. Note that for very fast scintillators with short decay times the times for excitation and occupation of the energy levels S_{1i} in fig. 13.4 are in the order of several hundred picoseconds, such that the rise time of the pulse must also be accounted for. This can be achieved by introducing another time constant τ_{rise} or by multiplication with a Gaussian $g(t)$:

$$I(t) = I_0 \, (\mathrm{e}^{-\frac{t}{\tau_{dec}}} - \mathrm{e}^{-\frac{t}{\tau_{rise}}}) \qquad \text{or} \qquad I(t) = I_0 \, g(t) \, \mathrm{e}^{-\frac{t}{\tau_{dec}}} . \tag{13.8}$$

For many scintillators one can also observe the intensity contribution with longer decay time, caused by delayed fluorescence and being in the order of several hundred nanoseconds. The intensity curve then consists of a dominant prompt part with decay time τ_f in the nanosecond region and a smaller slow part with time constant τ_s in the 100 ns to 1 µs range:

$$I(t) = I_1 \, g(t) \, \mathrm{e}^{-\frac{t}{\tau_f}} + I_2 \, \mathrm{e}^{-\frac{t}{\tau_s}} . \tag{13.9}$$

The relative contributions I_1, I_2 of both components often depend on the type of the impinging radiation $(\mathrm{p}, \alpha, \mathrm{n}, \gamma)$ due to their different ionisation densities. There is strong indication that this is caused by re-excitation of the triplet states responsible for delayed fluorescence (fig. 13.4, T_{1i} levels). Their de-excitation happens via intermolecular interactions between two excited molecules, thus leading to delayed fluorescence [616]. The probability for this to happen depends on the ionisation density and is the stronger the larger the specific energy loss of the projectile particle. The slow intensity component thus should be stronger for particles with larger $\langle dE/dx \rangle$. One can exploit this effect to distinguish different particles in the scintillator (*pulse shape discrimination*). Electronic time windows can be defined in which the intensity of the pulse tail towards long times is measured. Figure 13.7 shows pulse shapes with different pulse tails for particles with differently strong specific energy losses.

The measured pulse shape also depends on time constants arising from the readout. A typical readout chain—here with an additional amplifier whose effect on the pulse shape is similar to an RC time of an oscilloscope—is shown in fig. 13.8. Assuming a general decay scheme with decay constant τ_{dec} we find for the current:

$$i(t) = \frac{v(t)}{R} + C\frac{dv}{dt} \;\Rightarrow\; \dot{v} + \frac{1}{\tau_{RC}} v \;=\; \frac{i_0}{C} \, \mathrm{e}^{-\frac{t}{\tau_{dec}}} , \tag{13.10}$$

where τ_{RC} is the RC time constant of the amplifier input. The differential equation (13.10) has the following solution:

$$v(t) = \frac{\tau_{RC}}{\tau_{RC} - \tau_{dec}} \frac{Q_0}{C} \left(\mathrm{e}^{-\frac{t}{\tau_{RC}}} - \mathrm{e}^{-\frac{t}{\tau_{dec}}} \right) \tag{13.11}$$

with $Q_0 = \int i dt$ being the total charge of the current integrated in the preamplifier.

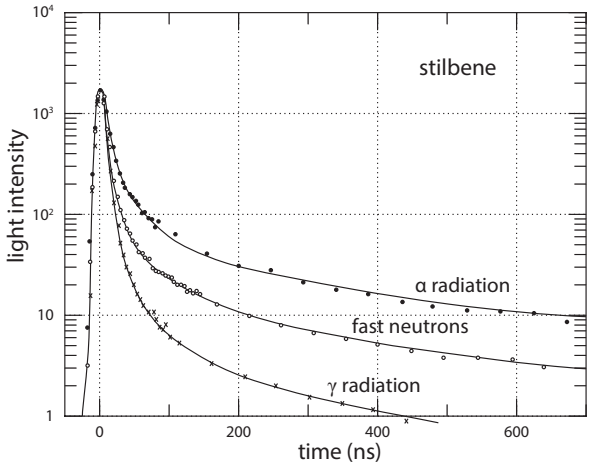


Fig. 13.7 Intensity shape in an organic scintillator (stilbene) upon excitation by different particle types with different ionisation densities [229]. The curves are relatively normalised to each other at the maximum. Note that neutrons deposit energy in the scintillator via (n,p) reactions.

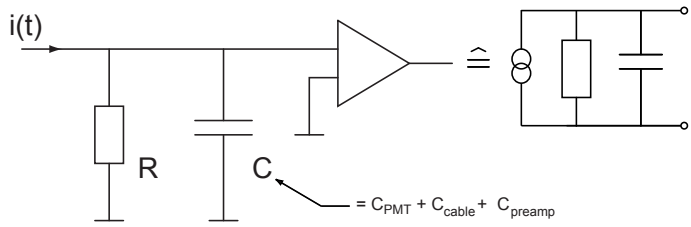


Fig. 13.8 Typical readout chain of a scintillator. C denotes all appearing capacitances of the readout chain (PMT, cable, amplifier). The equivalent circuit (right) consists of a current source with R and C in parallel.

For the two extreme cases the following simplifications can be made. If

$$\tau_{RC} \gg \tau_{dec} \qquad \text{then} \quad v(t) \approx \frac{Q_0}{C}(1 - \mathrm{e}^{-t/\tau_{dec}})$$

with the pulse rise time being determined by τ_f. If

$$\tau_{RC} \ll \tau_{dec} \quad \text{then} \quad v(t) \approx \frac{\tau_{RC}}{\tau_{dec}} \frac{Q_0}{C}(1 - \mathrm{e}^{-t/\tau_{RC}})$$

the signal is small because of $\tau_{RC}/\tau_{dec} \ll 1$, but the rise time can principally be short, given by τ_{RC}.

13.3 Inorganic scintillators

13.3.1 Scintillation mechanisms in inorganic crystals

Different to the molecular origin for scintillation in organic materials the scintillation mechanism in inorganic crystals rests on the lattice structure depending considerably on the band-structure properties (see also chapter 8) and the existing energy levels created by impurities or self-ions or other complexes in the band gap. A complete treatment of the multiple underlying processes is beyond the scope of this book; a

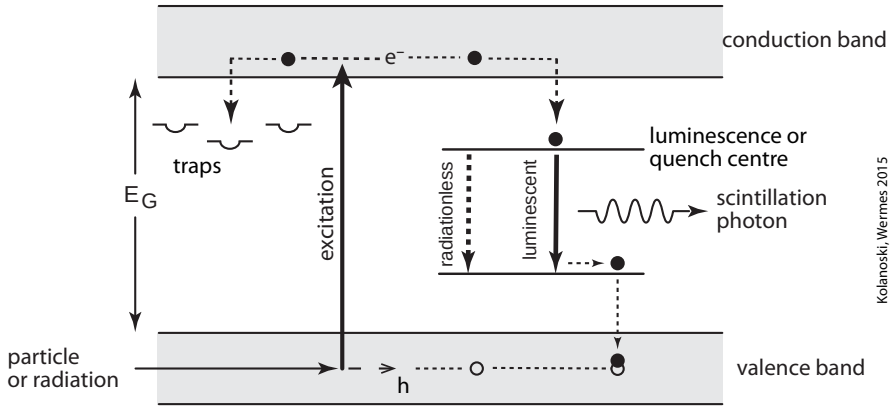


Fig. 13.9 Simplified compilation of mechanisms contributing to the scintillation phenomenon in inorganic crystals (adapted from [206]).

comprehensive treatise is given in [647]. We here describe the essential mechanisms and discuss scintillation crystals which are relevant for particle detectors.

In a perfect non-metallic crystal the electrons are in separated energy bands (fig. 8.7 on page 263) consisting of extremely dense energy levels. The valence band (VB) is the energetically highest band still populated; atom core bands lie underneath. The conduction band (CB) is above the valence band and is—in the case of an insulator—not populated with electrons. All energy levels of the valence band, by contrast, are filled with electrons. The electrons do not move since all locations of the band are occupied. Freely moving holes in the valence band (and electrons in the conduction band) arise when a valence band electron is lifted into the conduction band by supplying energy. In scintillating crystals the band gap adds up to about 4–12 eV and hence is substantially larger than in semiconductors. The energy released upon recombination of an electron and a hole thus corresponds to a photon wavelength which normally lies outside of the visible range and does hence not contribute to the scintillation yield.

In particle physics experiments scintillation crystals are predominantly used to detect high energy electrons or gamma rays in the above MeV energy range. Photons interact via photoeffect, Compton effect or pair creation (see section 3.5), thus creating electrons, respectively, electrons and positrons. At still higher energies photons as well as electrons create an electromagnetic shower (see section 15.2). The high energy (compared to the lattice electrons) electrons and positrons interact with the electrons of the crystal lattice and stimulate electron transitions from the valence band into the conduction band, if sufficient energy is transferred to the crystal. By means of further processes described in the following and sketched in fig. 13.9 visible scintillation light is produced.

Luminescence centres. In order to scintillate a crystal must have so-called luminescence centres whose energy level distances are smaller than the band gap distance and which have photon transitions in the visible range. Luminescence centres can be externally induced, for example by doping ('extrinsic' luminescence, see also page 518), but can also arise from the crystal's own ions or by crystal defects which are often created by the interaction process of the radiation to be detected (self-activated 'intrinsic' luminescence). Crystal doping to increase the light yield is often found in inor-

ganic scintillation detectors. A well-known example is thallium-doped sodium iodide, NaI(Tl).

Absorption of an electron and a hole (similar to direct photon absorption) in a luminous centre or capture of an exciton by a centre (see more below) cause a transition of an electron of the centre into an excited state. The relations of the potentials for the ground state and an excited state are sketched in fig. 13.10 as a function of a 'configuration coordinate' (instead of absolute coordinates). The configuration-coordinate concept is used in theoretical formalisms to treat differences between two different lattice configurations by formally putting all lattice coordinates into one coordinate q (see e.g. [681] or [747]). Here the two configurations are ground and excited state of a system comprising the luminescence centre in its surrounding lattice and in particular the shapes of their respective potential energies as a function of q. In free space, without the existence of a lattice environment, the shapes of the potential energy of ground state and excited state would lie exactly on top of each other in this coordinate. Due to the differing polarisation which the excited centre induces on the lattice neighbourhood, however, the binding relations and hence the locations of the minimum of the centre's ground state and excited state relative to the position of the minimum of normal lattice atoms (zero point) have changed. Normally the distance from the zero point position of the configuration coordinate is larger for the excited state (more strongly influencing the neighbourhood) than for the ground state [647], leading to the relative shift drawn in fig. 13.10. The neighbouring atoms adjust to the new equilibrium configuration consuming thereby part of the absorbed excitation energy, which together with thermal (vibrational) dissipation causes the C→B transition. The reconfiguration process is long ($\mathcal{O}(10^{-13}\,\mathrm{s})$) compared to the electronic transition ($\mathcal{O}(10^{-15}\,\mathrm{s})$) and we can assume that the latter occurs at a fixed lattice configuration (Born-Oppenheimer approximation). The transitions A→C and B→D can therefore be drawn as vertical lines in fig. 13.10.

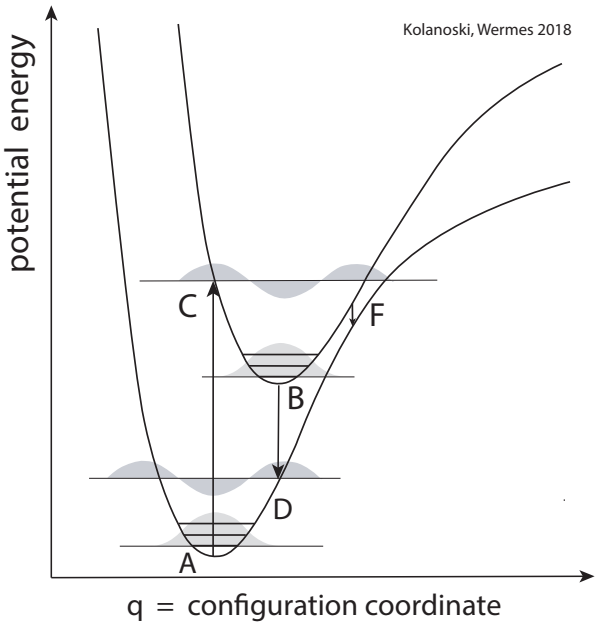


Fig. 13.10 Characteristic curves of the potential energies of ground state and excited state of a luminescence centre (sketched drawing). The shaded areas indicate the vibrational wavefunction amplitudes. Further explanations are given in the text.

The energy absorbed by the centre can go into (electronic) excitation and into vibrational energy with respect to neighbouring nuclei. The vibrational wavefunction amplitudes corresponding to different excitation energies are also sketched in fig. 13.10. The transitions for absorption and decay hence must respect electronic and vibrational energy changes ('vibronic' transitions[5]) and are the more likely the larger the overlap between the involved vibrational wavefunction amplitudes (Frank–Condon principle). The centre's decay back to the ground state then takes place in the new lattice configuration via the transition B→D, the luminescence transition. It does not directly go into the ground state minimum A such that the distance BD in fig. 13.10 is smaller than AC with correspondingly smaller wavelength (Stokes shift). This avoids re-absorption of the emitted light by luminescence centres. Finally, return from D to A proceeds by thermal dissipation of the excess energy.

Quenching. Besides transitions by photon radiation, transitions can also be radiationless (quench centres). This occurs with higher probability if in fig. 13.10 point F can be reached by excitation, where the two potential curves approach so closely that a non-radiative transition (e.g. a thermal transition) becomes possible.

Trapping. An electron can also be held in an energy level for some time (*trapping centre*), until it returns into the conduction band by thermal excitation or by a radiationless transition into the valence band. The energy of the impinging radiation or particle will therefore only in part be converted into photons with wavelengths in the visible or UV range.

Excitons. Excited electrons can also be bound to a hole in the valence band, or an e/h pair can remain bound after excitation. This state is called an *exciton* (Frenkel 1931) and corresponds to an (electron) energy level lying slightly below the conduction band. By comparatively small energy input excitons can easily be converted into an electron in the conduction band and a hole in the valence band. Conversely, conduction band electrons and valence band holes can recombine into exciton states. Like electrons and holes, excitons can also move freely in a solid. They can also decay radiatively.

Localisation. Besides *traps*, caused by lattice defects, locally fixed centres can be created by a strong local interaction, for example between a hole and the lattice which is locally polarised by the presence of the hole. This interaction locally fixes the hole (*self-trapped hole*, V_k centre). Excitons can be also localised (*self-trapped exciton*).

Luminescence transition. The ground states and excited states of the centres can be easily reached energetically by electrons from the conduction band and by holes from the valance band such that e/h recombination can take place via the centre. This leads to radiation when de-excitation from the excited state to the ground state occurs radiatively. If the luminescence centre is neutral, simultaneous capture of an electron from the CB and a hole from the VB is needed, or alternatively the capture of an exciton. Exciton states also act as luminescence centres since they can recombine. More efficient, however, is the coupling of an exciton to a luminescence centre, whereby energy and/or charge transfer from the exciton to the centre takes place.

Chronology of the scintillation process. Figure 13.11 illustrates in a simplified way (e.g. only one core band is drawn) the chronology of a self-activated, intrinsic scintillator, where the scintillation occurs via activator centres, for example an ion, an

[5]Simultaneous changes in the vibrational and electronic energy states.

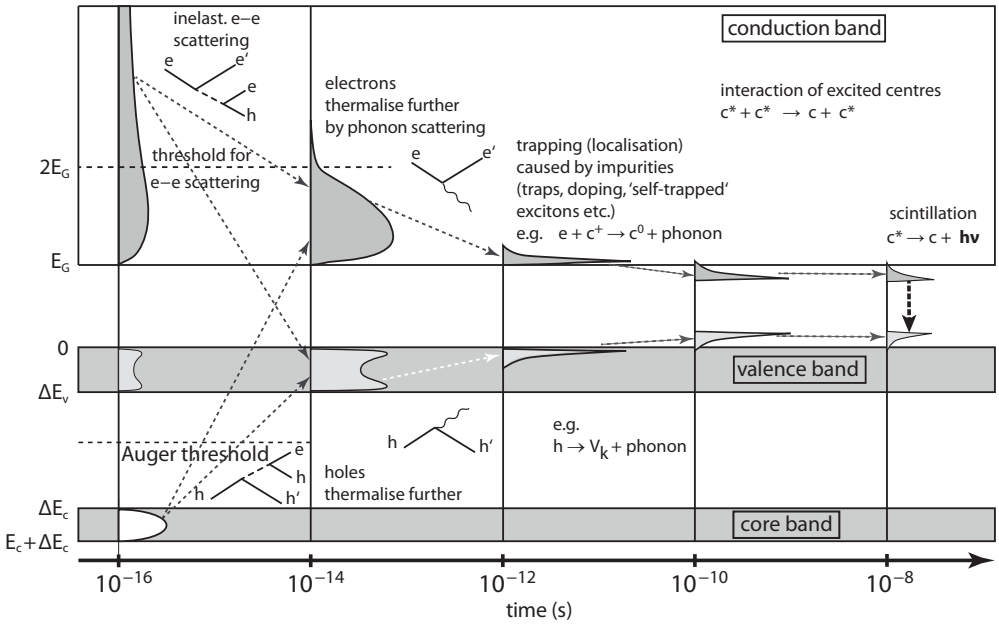


Fig. 13.11 Simplified time evolution of the scintillation process (here for intrinsic scintillations). For simplification only one atomic core band is drawn. The letter *c* in description labels trapping centres, V_k denotes the *self-trapped hole centre*. The dashed arrows indicate the population change of the energy bands caused by the contributing processes; see also explanations in the text (adapted from [647] with kind permission of Springer Science+Business Media).

anionic complex or an exciton state. The created holes normally are 'deep' holes, the electrons are 'hot' electrons, meaning that their energetic locations are far away from the respective band edges, and can—for the holes—even be created within the atom's core bands.

Electrons and holes can move quasi freely in their respective bands. Once (hot) electrons and (deep) holes are created as primary excitations by incident ionisation they scatter inelastically (electrons) or by Auger[6] processes (holes) (see section 3.5.3) with the lattice atoms, thereby creating a large number of secondary excitations with electrons in the conduction band and holes in the valence or core bands, respectively. This happens on a time-scale of 10^{-16}–10^{-14} s. The charge carriers then thermalise to lower energies, the electrons by inelastic e–e scattering and the holes by a succession of Auger processes, in which the energy lost to a lattice atom is transferred into electron excitation energy. Further thermalisation takes place by scattering with the lattice phonons on a time-scale up to 10^{-12} s after which the charge carriers are cooled down to the bottom of the conduction band (electrons), respectively, the top of the valence band (holes).

Timewise the next step after thermalisation is localisation of the excitations (i.e. of the respective charge carriers) in traps (defects, impurities, etc.) on a time-scale of 10^{-10}–10^{-8} s. The energy levels (ground state and excited state) of the centres lie

[6]Since hole movement in valence or core bands physically is electron 'hopping', the energy transferred in an Auger process to an electron can move that electron into the conduction band.

in between the bands near the band edges. Out of the excited states of the centres, finally, the radiative transition of scintillation (10^{-8} s) takes place.

Contrary to Cherenkov or transition radiation, for scintillation light there is a whole chain of different and complex processes responsible for its creation characterised by different time constants. Not mentioned here are still more complex processes which are caused by induced centres (induced by secondary generation and excitation of luminescence centres, thermal activation or by electron tunneling, see [647]). A classification in terms of the most important processes responsible for scintillation in crystals is given further below.

Doping with activator centres. In undoped crystals light emission by exciton generation and decay via a luminescence centre often is an inefficient process. To increase the probability for visible photon emission in the decay of the exciton state, crystals are often doped with activator centres. For NaI and CsI, thallium (Tl) is most often used, for rare earth crystals usually cerium (Ce) is used. The doping concentration typically amounts to about 0.1% molar fraction.

Time constants determining the pulse shape. Also in inorganic crystals the pulse shape of the detected scintillation light is governed by several time constants [647]: (i) the time span of the generation of (hot) electrons and holes ($\tau_1 = 10^{-18}$–10^{-9} s), (ii) the time for thermalisation by cooling processes through lattice interactions ($\tau_2 = 10^{-16}$–10^{-12} s), (iii) the transfer time of the electrons and holes or an exciton to reach a luminescence centre ($\tau_3 = 10^{-12}$–10^{-8} s), and (iv) the decay time of the luminescence centre ($\tau_4 > 10^{-10}$ s).

An important role in determining the pulse shape is played by the generation of the mentioned secondary centres. If we call the characteristic time needed for the generation of primary and secondary luminescence centres plus their interaction τ_{int} then, for the case that

$$\tau_{int} \approx \tau_3 \ll \tau_4 \,,$$

direct scintillation with a very fast rise time and a single exponential decay time dominates. If, hovever,

$$\tau_{int} \gg \tau_3 \qquad \text{and} \qquad \tau_{int} \gg \tau_4 \,,$$

which often is the case in real crystals, direct scintillation is accompanied by a delayed decay of a secondary luminescence centre (phosphorescence) with time constants in the > 100 ns to µs range. A description by simple exponential decay laws has limitations due to the interaction of the luminescence centres with each other and with charge carriers. More parameters are needed. It is nevertheless common practice to describe the pulse shape of crystal scintillators by a sum of exponentials [647].

This complexity is not found for scintillation processes in liquids or gases, respectively vapours, since to first order there is no interaction between the particles in the medium. Here the decay time of the luminescence centre governs the scintillation process.

Light yield. Capture of the exciton state by quenching or trapping centres leads to losses in the light yield. Other causes for light losses also exist (see e.g. [62]). Light yield reduction depends on the relative concentrations of crystal impurities and activator centres, but also for example on the temperature. Also, high ionisation densities saturate the luminescence yield (ionisation quenching).

After these considerations, the light yield L can be specified depending on the absorbed energy E_{abs} and the average energy $w_{e/h}$ required to create an e/h pair. Introducing further characterising quantities [647] we find

$$L = \frac{E_{abs}}{w_{e/h}} S Q \,, \tag{13.12}$$

where S is the energy transfer efficiency for the transfer from the (thermalised) e/h pairs onto the luminescence centres and Q is the quantum yield (probability for radiative decay) of the centre. Since $w_{e/h}$ is proportional to the band gap E_G it is advantageous to choose a material with a small band gap. However, this also increases the probability that the ground levels or excited levels of the centres are very close to the band edges and can thus be easily ionised instead of decaying via luminescence.

The maximum achievable theoretical yield has been computed by Dorenbos (see [647]) to be 140 000 photons per MeV deposited energy for an ideal crystal, doped with Ce^{3+} (e.g. $LaBr_3$:Ce^{3+}), assuming a small band gap, into which the optical transition just fits.

13.3.2 Scintillation crystals: classification and comparison

In table 13.3 some often used inorganic scintillation crystals are compiled, listing some characteristic properties which are important for their usage in experiments. A typical plastic scintillator is included in the table as a reference standard.

Most important and, depending on the application, most decisive characteristics for a choice are above all density, light yield, signal decay time and radiation hardness, the latter in particular for experiments at hadron colliders like LHC. While plastic scintillators have short rise and decay times in the region of nanoseconds, for inorganic crystals both are typically (much) longer; for exceptions see below. Particularly crystals with high light yield like NaI(Tl) and CsI(Tl) have decay times in the range of several hundred nanoseconds to microseconds.

In particle physics inorganic crystals are usually used for energy measurement of high energetic photons or electrons having energies larger than about 10 MeV and hence develop an electromagnetic shower in the crystal (see also chapter 15). For higher energies above some GeV the light yield no longer imposes a major limitation for good energy measurement. Instead high density and atomic number (small radiation length X_0) are important in order to efficiently absorb the shower within a reasonable crystal depth.

Classification according to scintillation mechanisms. In order to further classify scintillating crystals, one can distinguish according to the underlying scintillation mechanism following [647]. Here coupling and energy transfer between lattice and luminescence centre are essential. Of particular importance is the location of the energy levels of the centres within the band gap. Their position and hence their distance to the band edges depend on the way an ion is built into the lattice and can thus for different lattice hosts lead to different scintillation properties.

In the simplest radiation process without additional luminescence centres, electrons and holes, when being energetically close to the band edges, recombine after thermalisation. The chronology proceeds as described on page 516. After cooling down of electrons and holes the luminescent transition is found after interim creation of an exciton state or by using another centre in the lattice (e.g. an ion) which features easily

Table 13.3 Properties of often used scintillation crystals in comparison to a typical plastic scintillator [647, 746, 762].

Material	ρ (g/cm^3)	X_0 (cm)	R_M (cm)	λ_{max} (nm)	Photons per MeV	τ_{dec} (ns)	Remark
plast. scint.	1.03	42.5	∼9.6	423	10 000	≈2	easy handling
NaI(Tl)	3.67	2.59	4.13	410	43 000	245	hygroscopic
CsI(Tl)	4.51	1.86	3.57	550	52 000	1220	mildly hygrosc.
BaF_2	4.89	2.03	3.10	220	1 430	0.8	fast component
				310	9 950	620	slow component
$LaBr_3$(Ce)	5.29	1.88	2.85	356	61 000	17–35	hygroscopic
CeF_3	6.16	1.77	2.40	330	4 500	30	rad. tol. $>10^3$ Gy
GSO	6.71	1.38	2.23	430	9 000	56	rad. tol. $>10^3$ Gy
BGO	7.13	1.12	2.26	480	8 200	300	
LYSO(Ce)	7.10	1.14	2.07	420	33 000	40	rad. tol. $\sim 10^4$ Gy
LSO(Ce)	7.40	1.14	2.07	402	27 000	41	rad. tol. $>10^3$ Gy
$PbWO_4$	8.30	0.89	2.00	425	130	30	rad. tol. $>10^3$ Gy

X_0 = radiation length; R_M = Molière radius; λ_{max} = wavelength at the maximum of the emission spectrum; τ_{dec} = decay time constant of the emission. The quoted radiation tolerance refers primarily to γ radiation. Abbreviations: BGO = $Bi_3Ge_4O_{12}$, GSO = Gd_2SiO_5, LSO = Lu_2SiO_5, LYSO = $Lu_{1.8}Y_{0.2}SiO_5$.

accessible energy levels close to the band edges. The luminescence centres can also be excited directly through the ionisation process, emitting photons when de-exciting.

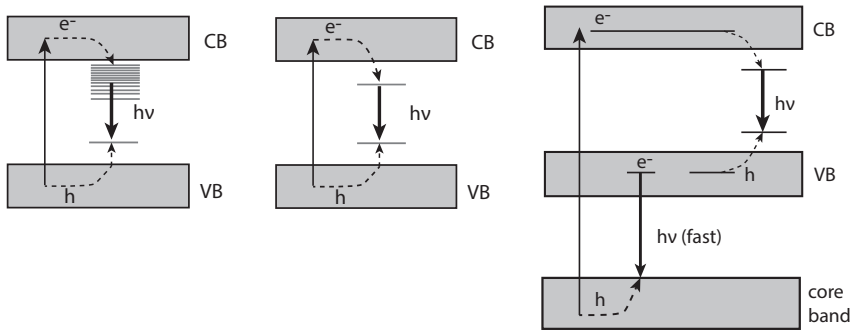
Under certain configurations of valence and lower lying atom core bands, however, thermalisation of the holes by Auger processes cannot take place. In this case direct recombination of a deep hole from a core band with an electron from the valence band is more likely, leading to fast (typical ≈ 1 ns) and energetic light emission, usually in the UV range due to the energy distance of the bands. This is known as *cross luminescence* [568] or *core-valence luminescence* [829, 828] (see fig. 13.12(c)). Normally such crystals also have a slow scintillation component from one of the mechanisms not restricted by Auger suppression.

Following [647] we classify inorganic scintillators according to the underlying scintillation mechanism:

– self-activated (by the interaction process) intrinsic scintillation: fig. 13.12(a);
– scintillation by (external) activator centres (for example by doping): fig. 13.12(b);
– scintillation by *cross luminescence*: fig. 13.12(c).

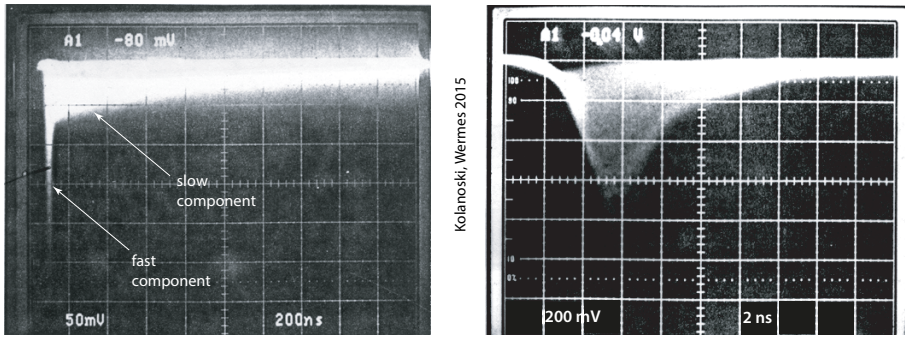
If combined with further physico-chemical properties, for example doping with specific ions, one can distinguish two important classes of inorganic scintillator crystals: halides (with F, Cl, Br, I) and oxide compounds (e.g. with WO_4, SiO_5 or Ge_4O_{12} groups). A detailed compilation can be found in [647]; the most important crystals for particle physics experiments are listed in table 13.3.

Fluorides generally have a large band gap ($E_{gap} > 7$ eV), a condition for possible *cross luminescence*. A well-known representative of this class is BaF_2, a scintillator crystal having a fast ($\tau_{dec} = 0.8$ ns) and a slow ($\tau_{dec} = 620$ ns) light component. The fast component is attractive for experiments that have to cope with high rates. Unfortunately, the fast component emits in the UV range and dedicated detection



(a) Self-activated. (b) Externally activated. (c) Cross luminescence.

Fig. 13.12 Classification of scintillation crystals according to their scintillation mechanism (see text): (a) intrinsic activation (for example via exciton excitation), (b) extrinsic activation (for example by doping), (c) (fast) radiation by *cross luminescence*.



(a) Slow and fast component. (b) Fast component only.

Fig. 13.13 Oscilloscope signals from a ^{137}Cs radioactive soure (662 keV γ), measured using a BaF_2 crystal: (a) slow and fast components (scope setting: 200 ns per grid spacing); (b) fast component (scope setting: 2 ns per grid spacing). Note that the rise time of the pulse is governed by the photomultiplier rise time.

arrangements must be foreseen, for example a PMT with a UV transparent quartz window. Figure 13.13 shows oscilloscope pictures of BaF_2 pulses: in fig. 13.13(a) the fast (peak at the very left) and the slow component can be seen; in fig. 13.13(b) only the fast component is shown on a nanosecond scale.

From the self-activated crystals CeF_3 sticks out, which is a good candidate for calorimetric measurements in experiments. However, its radiation length $X_0 = 1.77\,\text{cm}$ is still rather large. For this reason other crystals have been preferred for the LHC experiments CMS and ALICE. A scintillator crystal, especially developed for CMS is lead tungstate ($PbWO_4$) featuring high density, short radiation length and good radiation resistance at very short signal duration (see also section 15.5.2 on page 613). The low light yield can be tolerated for the comparatively high photon energies occurring in LHC collisions (more than several GeV). Oxide group crystals have another advantage: different to halides (especially fluorides) they are mechanically more stable, chemically inert, and non-hygroscopic.

Crystals doped with rare earth elements, especially cerium (Ce), feature a high light yield compared to oxides. A good example is cerium-doped (5–10%) lanthanum bromide, $LaBr_3$ (Ce) [669]. This scintillator combines high light yield (61 000 photons per MeV) with a fast signal (≈20 ns). However, the material is hygroscopic.

Lead glass. Lead glass (SiO_2 + PbO) is often used in similar applications as scintillating crystals, for example as a homogeneous electromagnetic calorimeter (see section 15.5.2 on page 614 and fig. 15.24). Lead glass is no scintillator, however, but its light output is rather based on Cherenkov light (chapter 11) emitted by the electrons and positrons created in the electromagnetic shower evolving in the lead glass material. The light output is typically lower by 2–3 orders of magnitude than that of scintillating crystals, leading to a correspondingly lower energy resolution (see table 15.5).

13.3.3 Radiation damage in inorganic crystals

All known inorganic scintillators suffer from radiation damage [1022, 1023], induced by γ radiation and also by charged particles or neutrons. Three possible radiation-induced damage effects can be considered: (1) influence on the scintillation mechanism itself (light creation), (2) radiation-induced phosphorescence, called 'afterglow', and (3) absorption of light by radiation-induced defects (colour centres). While the third and most important effect enhances the light absorption and hence reduces the light output in a scintillator or wavelength shifter, and the second one increases dark current and hence causes noise in photodetectors, there is no experimental evidence that supports a damage-induced deterioration of the scintillation mechanism itself (see Zhu in [488], p. 536).

Crystal-type dependent, the colour centres can have different origins. They can be electrons in anion- or holes in cation-vacancies, or interstitial–anion ions. It is possible that colour-centre formation can recover spontaneously upon temperature application or even at room temperature, a process called colour-centre annihilation. In this case the damage effect would be dose-rate dependent. If recovery under temperature does not take place or the recovery is too slow, the colour-centre density increases continuously with irradiation until all defect sites are filled. In this case, the corresponding radiation damage effect is dose dependent, but not dose-rate dependent. Light yield measurement as a function of dose rate can therefore be used for damage characterisation [488].

The temperature dependence is used for thermal annealing by applying high temperatures to damaged crystals (>200°C, crystal dependent). Alternatively, 'optical bleaching' can also be applied, which is the illumination of the crystal with light of different wavelengths. While both methods reduce the density of the colour centres and thus improve the optical transparency, a spatial variability in transparency is easily introduced in the annealing process which influences the homogeneity of the light output.

For high radiation environments like at LHC, $PbWO_4$ has been chosen for the CMS electromagnetic calorimeter (see section 15.5.2). Still, the light output of crystals at large rapidity was reduced by 70 % after two years of operation at still medium luminosities [1023]. Radiation damage in $PbWO_4$ crystals manifests itself predominantly in light absorption due to radiation-induced colour-centre formation. Besides a loss in light output, a dose rate dependence is observed leading to an equilibrium in light output during irradiation periods for a certain dose rate, caused by the mentioned

colour-centre dynamics, where colour-centre formation (damage) is as fast as annihilation (recovery), so that their density is constant. Attenuation increase has also been observed in hadron irradiation and has likely been linked to highly ionising heavy fission fragments from elements with $Z > 71$ [356].

LSO (Lu_2SiO_5:Ce) and especially LYSO ($Lu_{2(1-x)}Y_{2x}SiO_5$:Ce) crystals (see table 13.3), containing only light elements below the fission threshold, have hence come into focus [1023]. They show radiation hardness to 10^4 Gy against gamma rays and are also tolerant to damage from neutrons and charged hadrons, making them a preferred material for calorimeter applications in severe radiation environments.

13.3.4 Scintillating gases and liquids

Certain high-purity inert gases can be used as scintillation detector media, for example noble gases or (sometimes) nitrogen, with xenon currently receiving the largest attention.

The scintillation mechanism is comparatively simple and is attributed to the decay of excited dimers (called *excimers*) from the excited state to the ground state. Inert gases usually have complete outer shells and can bind only when an electron is excited from the shell. Such excimers (e.g. Xe_2^*) can be produced by direct excitation of an atom ($Xe^* + Xe \rightarrow Xe_2^* \rightarrow 2\,Xe + h\nu$) or by electron–ion recombination ($Xe^+ + e^- \rightarrow Xe_2^* + \text{heat}$; $Xe_2^* \rightarrow 2\,Xe + h\nu$). The lifetime of the excited states is therefore short, in the range of nanoseconds.

The emitted light has the largest yield in the UV range where standard PMTs are not sensitive. Wavelength shifting, achieved for example by wall coating, is hence mandatory, shifting the peak wavelength into the blue-green band. For not too high ionisation densities the light yield depends linearly on the deposited energy (dE/dx), see (13.1). The yield per deposited energy does not change much between the gaseous and the condensed phase (see [115] and references therein). To increase the ionisation density in gases and hence the light yield, high pressure operation (up to 200 bar) or operation as a liquid is preferred.

Table 13.4 lists scintillating inert gases and their characteristics. From the table it is evident that He and Ne have boiling temperatures far below the liquid nitrogen cooling range and also emit in the far UV where no transparent window materials exist. Ar and Kr are principally suited scintillators, but Xe stands out because it features high Z, high density and small radiation length X_0 together with high light yield at more easily accessible wavelengths (tables 13.4 and 3.3).

Liquid noble gas detectors [115] are used for example in experiments searching for Dark Matter represented by WIMPs[7] (see section 16.7.2) as for example the XENON100 [114] and LUX [69] underground experiments.

13.4 Light collection and readout techniques

Scintillation detectors need secondary photon detectors to detect the generated scintillation light. Photodetectors are described in detail in chapter 10. Contrary to Cherenkov detectors, scintillators do not require single photon detection. Nevertheless, different readout techniques are used depending on the amount of energy deposited in the scintillator, the resolution requirements and the available space in an experiment.

[7] WIMP = weakly interacting massive particle.

Table 13.4 Compilation of characteristic properties of scintillating nobel gases/liquids using various sources [547, 115, 206, 616]. The liquid density is quoted at the boiling point. For the decay times τ_{dec} only the shortest typical time constants [115] are listed. Note that decay times as well as peak emission depend on pressure and (much) on the gas purity. Values for N_2 and CF_4 are from [649]. NaI(Tl) is included as a reference.

Gas/liquid	Z	Boiling point at 1 bar (K)	ρ (liquid) (g/cm^3)	$\lambda^{\text{em}}_{\text{peak}}$ (nm)	Photons per MeV	τ_{dec} (typ.) (ns)
			gases (1 atm, 20° C)			
N_2	7	77	1.17×10^{-3}	390	100	2.5
CF_4	8.6	145.3	3.93×10^{-3}	300/630	1200	6
Helium	2	4.2	1.66×10^{-4}	78	1100	< 20
Neon	10	27.1	8.39×10^{-4}	∼80		1.2
Argon	18	87.3	1.66×10^{-3}	127	18 200	6
			liquids			
Helium	2	4.2	0.13	80	15 000	10
Neon	10	27.1	1.21	78	30 000	15
Argon	18	87.3	1.40	127	40 000	6
Krypton	36	119.8	2.41	147	25 000	3
Xenon	54	165.0	3.06	175	46 000	3
			solid (for reference)			
NaI(Tl)	46.5		3.67	410	43 000	245

A typical scintillator-detector system, used either with plastic scintillators or with inorganic crystals, consists of (fig. 13.14):

- the scintillator,
- a light guide where necessary,
- a conversion unit transforming the scintillation light into electrical charge carriers (e.g. electrons in a photocathode of a photomultiplier or electron/hole carriers in a photodiode),
- one or several amplifier stages (secondary electron multiplier or/and an electronic amplifier if applicable).

In readout systems with PMTs, but also for avalanche photodiodes and variants (section 10.3.2) and for SiPMs (*silicon photomultiplier*, section 10.5), the conversion and amplification function of the system are merged. A light guide can be omitted if a direct, large area coupling of scintillator and photoconversion unit is possible.

The *light collection efficiency* of a scintillator system is determined by the fraction of light that enters the detecting surface (photocathode of a PMT or photodiode surface). The conversion into electrons or electron–hole pairs, respectively, is characterised by the *quantum efficiency*, QE (see section 10.2). The subsequent amplification elements should provide high and noiseless amplification, for example of the electrons in the SEM (secondary electron multiplier = dynode system, see section 10.2.1) of a PMT.

In the following we introduce two different systems for scintillator coupling and readout which are often chosen in experiments. One is typical for readout with a

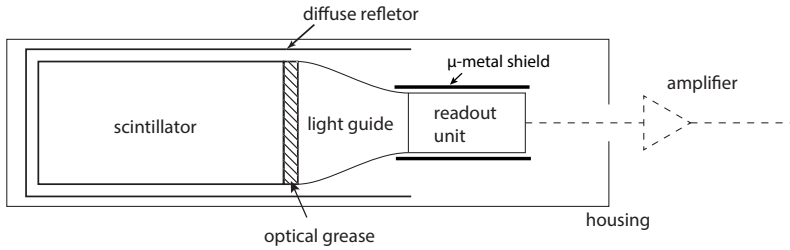


Fig. 13.14 A typical scintillator-detector set-up including light guide, light conversion and readout unit, often realised by a PMT. This combination is often used for plastic scintillators, but also appears for crystals. When using a PMT for readout further amplification is often not necessary. To shield against external magnetic fields (even weak ones) the sensible dynode system of the PMT is often housed in a μ-metal cylinder (see section 13.1 on page 503).

photomultiplier system, the other typical for the readout when semiconductor photodetectors are used.

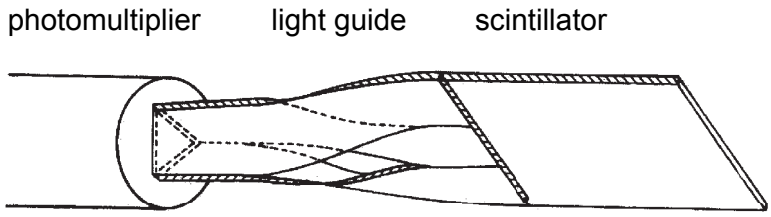
13.4.1 The system: scintillator–light guide–photomultiplier

A system of scintillator, light guide, PMT is used particularly often in combinations with planar-shaped plastic scintillators (see fig. 13.15). The scintillation light is fed by the light guide onto the photocathode of a photomultiplier, where it is amplified in the SEM of the PMT. Such a system is often wrapped with a reflector foil (see fig. 13.14), either specular (e.g. Al foil) or diffuse (e.g. a foil of magnesium oxide), with diffuse reflection generally yielding slightly better results [616]. For better contact between the surfaces of scintillator and light guide and between light guide and PMT, respectively, *optical grease* or *optical glue* is used, where 'optical' here refers to the fact that the refractive index of the contact glue is near that of the connected parts to minimise losses.

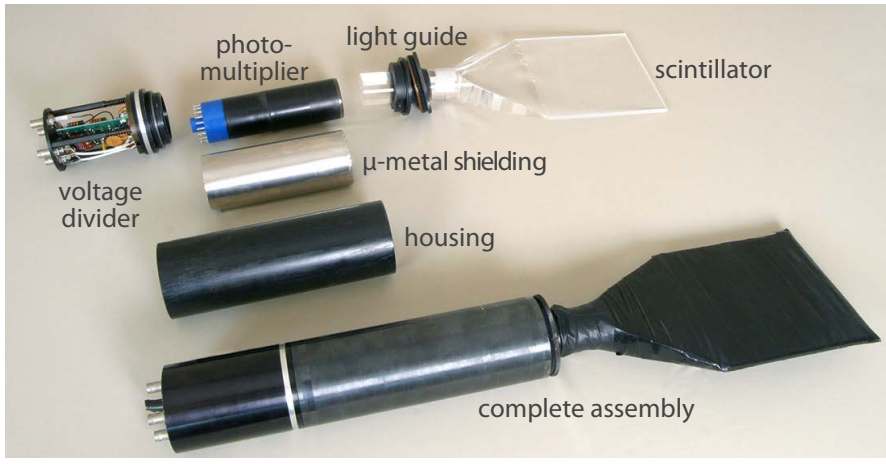
The combinations scintillator–photomultiplier and scintillator–lightguide–photomultiplier, respectively, are often used in experiments, last but not least because of their ease of handling compared to other arrangements providing fast particle detection with moderate spatial resolution. Figure 13.15(a) shows a special light-guide geometry (explanations further below) and fig. 13.15(b) shows the individual components of the assembly. Photomultipliers as detectors for scintillation are discussed in detail in section 10.2.1. The photocathode sensitivity of the PMT should match the emission spectrum of the scintillation light. Peak emission of CsI(Tl) is at 540 nm (fig. 10.17 on page 423), whereas the mentioned fast component of BaF_2 ($\tau = 0.8$ ns) emits at 220 nm in the deep ultraviolet (see also table 13.3). For BaF_2, therefore, special photomultipliers with quartz entrance windows having better transparency for UV light and with special photocathodes must be used to detect the fast BaF_2 light component with sufficient quantum efficiency.

Light guide. In order to guide the light to the PMT, a dedicated light guide can be more efficient than direct coupling between the scintillator and the PMT,

– if the geometry of the light exit face of the scintillator can only be mapped with difficulty onto the dimensions of the photocathode of the PMT, and/or
– if, for example for space reasons or the presence of a magnetic field, the photomultiplier cannot be directly coupled onto the scintillator.



(a) Scintillator–light guide–PMT system.

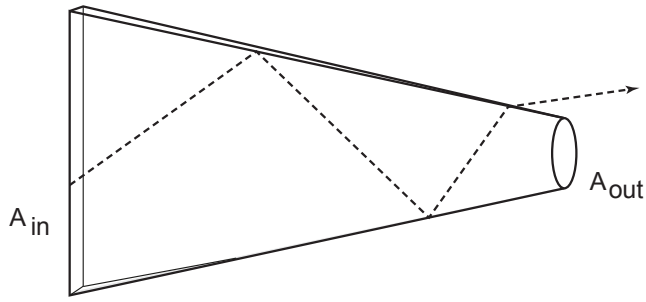


(b) Components and complete assembly.

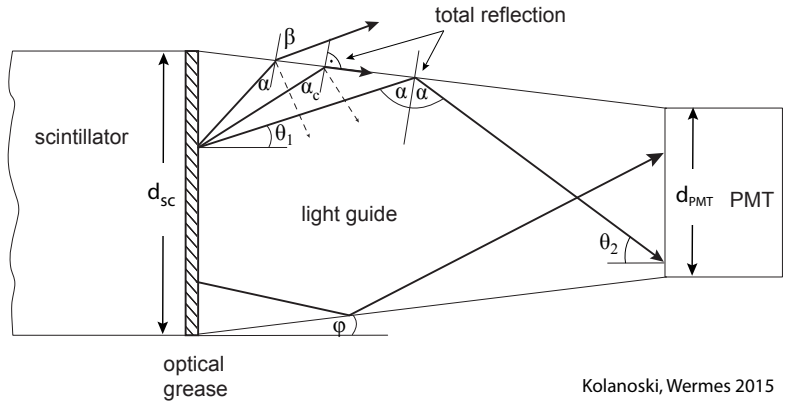
Fig. 13.15 Scintillation-detector system. (a) Set-up with scintillator, light guide, and photomultiplier [616]. By means of so-called 'adiabatic' shaping of the light guide with only gradually changing forms, light losses are minimised. (b) Components and complete assembly (source: DESY).

Complicated geometrical circumstances in a detector often require complex shapes of the light guide, for example if additionally focusing or bending of the light is also necessary. The scintillation light should arrive at the PMT entrance face with as little losses as possible. However, losses are unavoidable, as is shown in fig. 13.16(a) for a simple light guide called 'fish tail' due to its characteristic shape.

A lossless mapping of the light intensity from the scintillator exit face onto a smaller entrance face at the PMT end, would violate the 'Liouville theorem' [695] (see also e.g. [467]), which demands that the phase space volume is conserved. The relationships can be illustrated considering the light path in two dimensions with projected cross sections, as pictured in fig. 13.16(b). The phase space in the case of light propagation can be specified as the product of the maximum transverse coordinate extension $x_T^{\max}$ and maximum divergence ($\sin\theta^{\max}$), each at the input and at the output of the light guide, where the maximum transverse extensions are given by the respective projected cross section widths d_{sc} and d_{PMT}:



(a) Light guide in the form of a *fish tail*.



(b) Light guiding by total reflection.

Fig. 13.16 Illustration of losses in light guides. (a) Entrance and exit face for a light guide in the form of a fish tail. Multiple reflections to reach the output cross section result in losses due to the tapering shape of the light guide. (b) Light guiding by total reflection (see text).

$$d_{\rm sc}\,\sin\theta_1^{\rm max} = d_{\rm PMT}\,\sin\theta_2^{\rm max} \quad \Rightarrow \quad \sin\theta_1^{\rm max} = \frac{d_{\rm PMT}}{d_{\rm sc}}\sin\theta_2^{\rm max}\,, \qquad (13.13)$$

where $\theta_{1,2}^{\rm max}$ are the maximum angles at the entrance and exit face of the light guide, respectively, which must respect the maximum angle criterion for total reflection inside the light guide. Guiding non-parallel light by total reflection from a larger input to a smaller output cross section (area) without losses is not possible. The fraction $f_{\rm det}$ of the detected light is always at most equal to the ratio of light input area A_{in} to output area A_{out} (for a proof see [445]):

$$f_{\rm det} \leq \frac{A_{out}}{A_{in}} \qquad (13.14)$$

such that the effective readout cross section of a system is given by the coupling area to the photomultiplier.

In a light guide the light is mainly conducted by total reflection (fig. 13.16(b)). Wrapping with a reflecting foil can somewhat enhance the yield. From a medium with refractive index n exiting into air (n_0) the light entering the boundary with angle α to the vertical is refracted towards larger angles β with only little reflection, until a refraction angle of 90° is reached and total reflection sets in:

$$\sin \alpha_c = n_0/n \,, \tag{13.15}$$

with α_c being the critical angle. Light rays entering the boundary with $\alpha < \alpha_c$ are only partially reflected under the same angle as the incident angle. For small tapering (angle φ in fig. 13.16(b)) the light reflection angle is roughly maintained at subsequent reflections and the corresponding light rays will eventually be lost, whereas total reflection is lossless. Thus the maximally allowed photon entrance angle into the PMT corresponding to α_c is:

$$\theta_2^{\max} = \frac{\pi}{2} + \varphi - \alpha_c \,, \tag{13.16}$$

where φ is the tapering angle of the light guide. Hence we get using (13.15) and (13.16) with $n_0 \approx 1$ and $\varphi \ll \alpha_c$:

$$\begin{aligned} \sin \theta_2^{\max} &= \sin\left(\frac{\pi}{2} + \varphi - \alpha_c\right) = \cos\left(\alpha_c - \varphi\right) \\ &= \sqrt{1 - \sin^2\left(\alpha_c - \varphi\right)} \approx \sqrt{1 - \sin^2 \alpha_c} \approx \sqrt{1 - \frac{1}{n^2}} \approx 0.75 \\ \Rightarrow \qquad \theta_2^{\max} &\approx 48^\circ \end{aligned} \tag{13.17}$$

where a typical light guide made of plexiglas ($n \approx 1.5$) has been assumed. In order to keep the losses due to not reaching the critical angle (i.e. small α, large θ_1) small, the index of refraction should be as large as possible, while maintaining some level of 'matching' of the refractive indices of light guide and PMT, so that the light can efficiently enter the latter. The angular range in α below the critical angle α_c is called the *loss cone*.

The condition for total reflection restricts the available phase space for light transport. Equation (13.16) yields a maximum $\theta_2^{\max}$ for a given α_c which relates to a maximum $\theta_1^{\max}$ via the Liouville condition of (13.13). Assuming isotropic light emission into the forward solid angle (2π), the fraction f of the light which can maximally be extracted at the end of the light guide (see fig. 13.16(b)) can thus be estimated with (13.17) and (13.13) to be

$$f = \frac{\Delta\Omega}{2\pi} = \frac{1}{2\pi}\int_{\theta=0}^{\theta_1^{\max}} d\Omega = 1 - \sqrt{1 - \sin^2 \theta_1^{\max}} = 1 - \sqrt{1 - \left(\frac{d_{\mathrm{PMT}}}{d_{\mathrm{sc}}}\right)^2 \left(1 - \frac{1}{n^2}\right)} \,. \tag{13.18}$$

For $n = 1.5$ and $d_{\mathrm{PMT}}/d_{\mathrm{sc}} = 0.5$ one obtains $f = 7.2\,\%$.

Best results for more complex light guides such as the one in fig. 13.15(a), are obtained by so-called 'adiabatic' light guiding,[8] in which only gradual (small angles) and step-less shape changing without sharp bends or kinks occurs. More considerations for light directing can be found in [593] as well as in [652] and references therein.

13.4.2 The system: scintillator–(WLS plate)–photodiode

Photodiodes and avalanche photodiodes (described in sections 10.3.1 and 10.3.2) are well suited as photodetectors for scintillation light, in particular if the input light intensity is sufficiently high. An example is scintillation light from photons in the gamma

[8] Here 'adiabatic' means 'without loss of light' ($\delta Q = 0$).

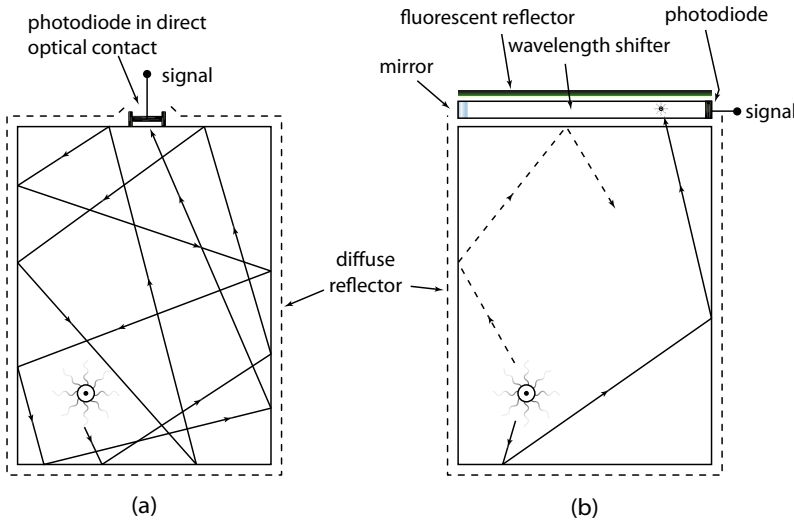


Fig. 13.17 Two alternative arrangements of photodiodes in scintillator readout. (a) Direct attachment of the photodiode on the scintillator crystal; (b) Enlargement of the light collecting surface by employing a wavelength-shifter – photodiode configuration.

energy range from about 300 keV up to GeV by their total absorption in inorganic crystal scintillators. Besides simplicity and price the immunity of the photodiode against external magnetic fields is a big advantage in scintillator readout.

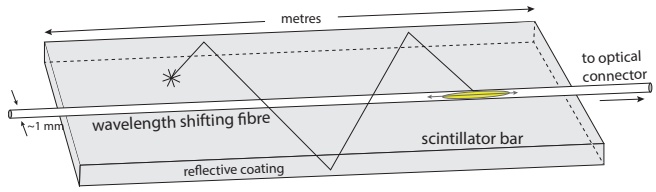
The output signals of the photodiode must, however, be amplified by an external amplifier to obtain a well-recordable signal. The amplification electronics largely determines the noise properties of the system, which are usually dominated by the noise contributions which are governed by the size of the diode capacitance (thermal and 1/f noise; see section 17.10.3.3).

The larger the collection area of a photodiode the larger is the light yield. However, large area photodiodes also feature a larger capacitance C_D which is for the most part proportional to the diode's surface. Since noise is directly proportional to C_D and the capacitance roughly scales with the diode's surface (at least for large area diodes), the signal-to-noise ratio becomes area independent (Groom's theorem [478]):

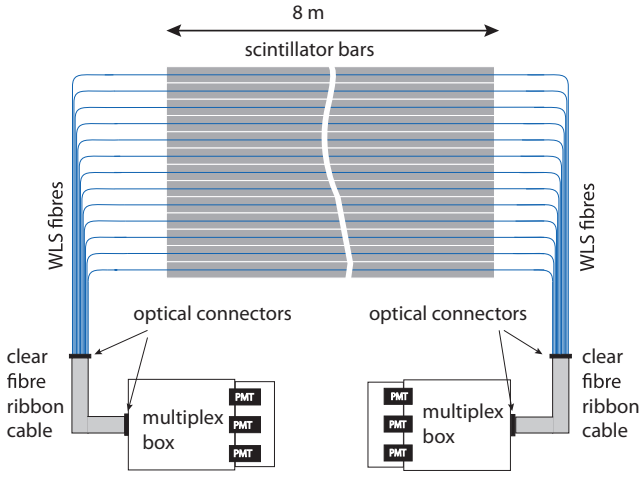
$$\frac{S}{N} \sim \frac{\text{light yield}}{\text{noise contribution}} \simeq \frac{\text{diode area}}{\text{capacitance}} \approx \text{const.}$$

As shown in fig. 13.17 (a) the photodiode can either be directly placed on the crystal or, alternatively, it can be used for readout of the scintillation light via a wavelength shifter (fig. 13.17(b)).

The properties of wavelength shifters are described in section 13.2.2. In the example of fig. 13.17(b) one exploits the fact that the WLS emission is isotropic in all directions. By using a cost-efficient small area photodiode, which is placed at the edge of the WLS plate, light can be read out which is collected in an area (WLS surface) considerably larger than the area of the diode itself. Light losses in the WLS, but lower noise figures of a smaller photodiode, are to be compared against an alternative large area photodiode without WLS plate.



(a) WLS fibre readout of a scintillator bar.



(b) Scintillator array with WLS fibre readout.

Fig. 13.18 Readout of scintillators via wavelength-shifting fibres. (a) Principle: scintillation light is reflected at the coated walls of a scintillator bar. Some fraction of the light enters the WLS fibre where it is absorbed and isotropically re-emitted within the fibre. (b) Example of an arrangement of scintillators with WLS readout as realised in the MINOS neutrino oscillation experiment [90]. Figure is not to scale.

13.4.3 The system: scintillator–WLS fibre–PMT/SiPM

Modern techniques for efficient light collection of scintillators arranged in large arrays in timing or calorimetric applications employ *wavelength-shifting optical fibres*, embedded for example in a plastic scintillator bar. Typical fibre diameters are of the order of 1 mm. They can be several metres long and are fed into multi-channel PMTs or SiPMs (see e.g. [250]). The principle is shown in fig. 13.18, together with an arrangement as employed by the neutrino oscillation experiment MINOS [90]. The light-transmitting fibre contains a wavelength-shifting agent which isotropically re-emits absorbed scintillation light. A good example of a calorimeter application is the ATLAS tile calorimeter, a iron/scintillator sampling calorimeter with 3 mm thick plastic scintillator tiles which are read out by wavelength-shifting fibres (see chapter 15 on page 656).

13.5 Scintillators as particle detectors

While organic plastic or liquid scintillators are most often used in time critical applications because of their short scintillation time constants, inorganic crystals are predominantly employed for energy measurement of (final states with) photons owing to their good energy resolution.

13.5.1 Fast plastic scintillators

Assembly of plastic scintillators is comparatively simple and they provide fast signals in the sub-nanosecond range. Therefore they are well suited and often used for timing

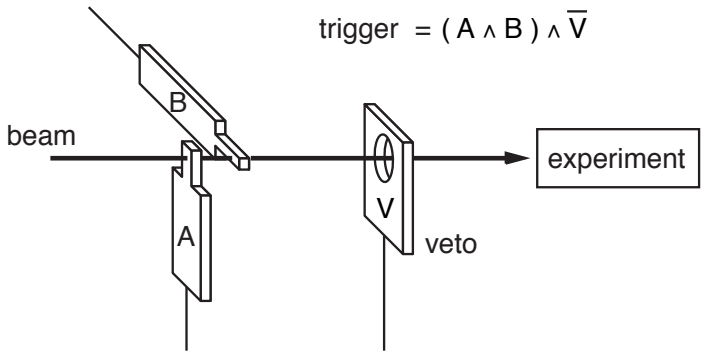


Fig. 13.19 Plastic scintillator counters used in a beam-test set-up for triggering.

measurements or for supplying time critical signals for time marks or time windows like triggers or gates.

The application of plastic scintillators for particle identification by time-of-flight (TOF) measurement is described in detail in section 14.2.1. For the generation of fast trigger or gate signals plastic scintillators are often employed in fixed-target experiments or in set-ups for detector test in accelerator beams (so-called beam tests). An example is shown in fig. 7.25(a) on page 213. The plastic counters are often arranged as 'finger counters' to select a well-defined portion of the incoming particle beam or as 'veto counters' (scintillators with a central hole for beam passage) to separate background particles not directly coming from the straight beam. The logic operation of the signals from fig. 13.19 is

$$\text{trigger} = (A \wedge B) \wedge \overline{V}\,.$$

The TOF systems in collider experiments described in section 14.2.1 are also suited for triggering if different counter configurations (e.g. counters 180° opposite to each other trigger on back-to-back particles) are combined as coincidence signals and used as inputs to the trigger hardware.

13.5.2 Scintillation detectors for photon energy measurements

While organic plastic scintillators are usually used to provide a fast particle signal (section 13.5.1), inorganic crystal scintillators are often used for energy measurement of gamma quanta or of electrons and positrons. The application range starts at about 10 keV in nuclear gamma spectroscopy reaching up to more than 100 GeV at high energy accelerators. At the low energies within this range, mainly photoeffect (section 3.5.3) and Compton scattering (section 3.5.4) play a role for photoabsorption, with pair creation (section 3.5.5) starting at energies above 1 MeV and becoming dominant at high energies, such that the former and any other processes can be neglected (see also section 15.2).

13.5.2.1 Photon spectra in inorganic crystals

The energy spectrum measured in a detector also depends—apart from the gamma energy—on the geometry and the size of the scintillator crystal and on the question whether energy escaping from the crystal is lost or can be measured by neighbouring detectors. In an individual and isolated crystal detector, whose dimensions are

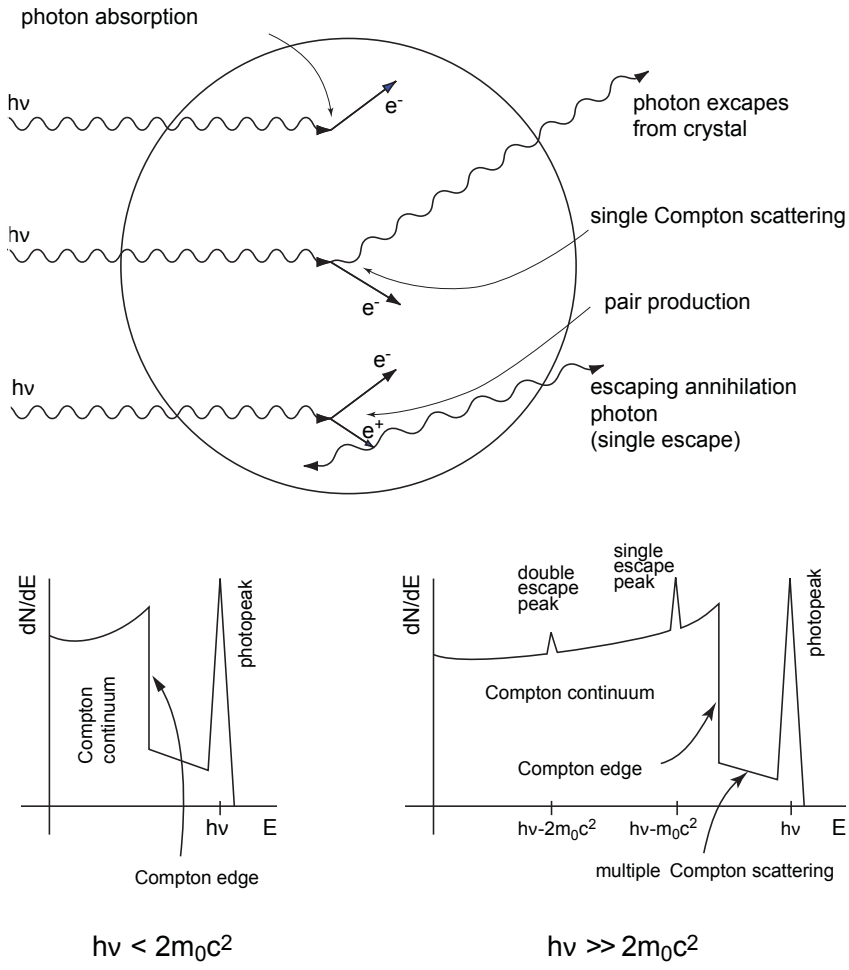


Fig. 13.20 Relationship between photon spectra and photon absorption processes in an inorganic scintillation crystal as the detector (adapted from [616]). For an explanation see text.

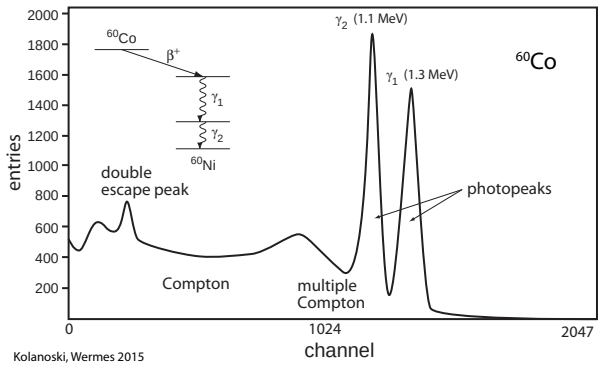
small compared to the photon absorption length, the following processes can occur (fig. 13.20):

– The photon is absorbed by photoeffect, thereby emitting an electron from the atomic shell (normally from the K-shell) which receives the energy of the photon reduced by the binding energy: $E_e = E_\gamma - E_B$. The created electron ionises the scintillation crystal, thereby generating scintillation light. If the electron remains in the detector and also the binding energy ends up in the signal (e.g. by absorption of the transition quantum, normally from the L- to the K-shell) or is negligible, then this process generates the so-called 'photopeak'.
– If the photon interacts instead by the Compton effect and the scattered photon leaves the crystal without further interaction, then the energy transferred to the electron according to the Compton kinematics (eq. (3.125) in section 3.5.4) is measured in the

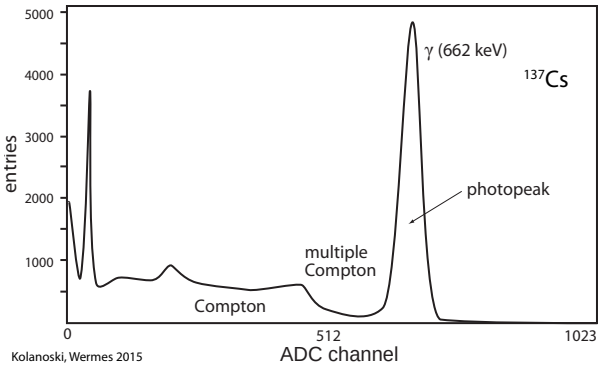
detector, which is always smaller than the energy of the photopeak and is maximal for photon backscattering ($\theta = 180°$). The so-called 'Compton continuum' is formed with the sharp Compton edge in the measured spectrum.

– At energies above 1 MeV pair creation with subsequent processes is also possible. Often pair creation occurs with subsequent annihilation of the positron with a crystal electron, a system with total energy of about $E = 2m_e c^2$, since the kinetic energies of the partners can be neglected in comparison. The annihilation is into two photons of which one or even both can escape the detector. Both events create peaks in the spectrum at the photopeak energy, less the single ($h\nu - m_e c^2$) or double ($h\nu - 2m_e c^2$) electron mass, causing the *single escape* or *double escape* peaks, respectively (see fig. 13.20).

If the scintillation crystal is large enough such that all reaction products (electrons, positrons and photons) are finally absorbed in the crystal and hence scintillation light is generated proportional to the total energy, one always measures the *full energy peak* corresponding—as the photopeak mentioned above—to the full energy of the incoming original photon. In praxis the mentioned peaks and the Compton edge appear together with further processes showing up in the energy spectrum: multiple Compton processes create a continuum in the energy spectrum between Compton edge and photopeak. Further energy losses from the original energy occur if individual photons or electrons



(a) ^{60}Co spectrum.



(b) ^{137}Cs-Spectrum.

Fig. 13.21 Spectra of two different radioactive γ sources (^{60}Co, 1.1 MeV and 1.3 MeV; ^{137}Cs, 662 keV), measured with a CsI(Tl) detector [638]. The scales on the abscissae are different by a factor two.

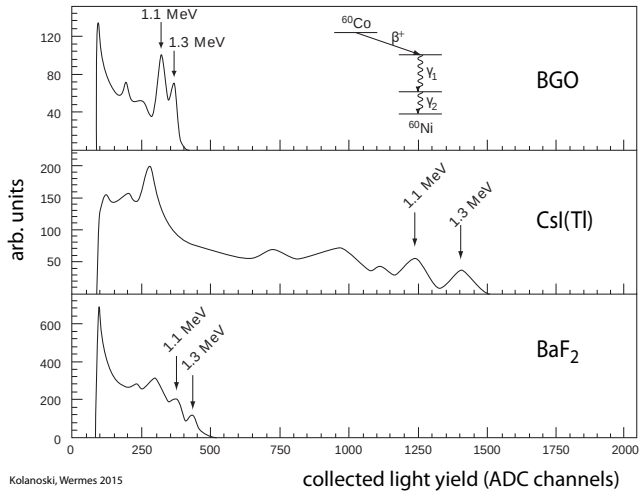


Fig. 13.22 ^{60}Co spectra measured with three different $1 \times 1 \times 1\,\mathrm{cm}^3$ scintillator crystals using the same measurement set-up [638]. The two peaks from the γ transition of ^{60}Ni at 1.1 MeV and 1.3 MeV are marked. Shown is the measured light intensity without correcting for any difference in the emission spectra.

escape the crystal. In addition back scattering at the walls reflect photons back in the crystal.

Figure 13.21 shows spectra of γ quanta from two different radioactive sources measured with a $1 \times 1 \times 1\,\mathrm{cm}^3$ CsI(Tl) crystal. In fig. 13.21(a) the two photon lines from ^{60}Co (1.3 MeV and 1.1 MeV, respectively) are detected in a scintillating crystal. In fig. 13.21(b) the spectrum of ^{137}Cs ($E_\gamma = 662\,\mathrm{keV}$) can be seen. The spectra are measured with the same set-up and can hence be directly compared. The abscissae are different by a factor two.

In fig. 13.22 spectra of the same ^{60}Co source, measured using three different scintillation crystals, read out by a photomultiplier with the same readout chain, are directly compared. The probability for photoeffect ($\propto Z^5$) compared to Compton effect ($\propto Z$) and pair creation ($\propto Z^2$) depend on the atomic number (see section 3.5). The light yields for the three scintillator materials differ from each other. Bismuth germanate ($Bi_4Ge_3O_{12}$), called BGO ($Z_{\mathrm{eff}} = 62.5$), and BaF_2 ($Z_{\mathrm{eff}} = 45.8$) (see table 13.3) feature a smaller peak yield compared to CsI(Tl) ($Z_{\mathrm{eff}} = 54$) due to their smaller light yield per MeV energy deposit. BGO also shows a larger photoeffect to Compton ratio because of its higher Z value. BaF_2 is interesting for many applications because of its very fast scintillation component at 220 nm ($\tau = 0.8\,\mathrm{ns}$), but it features a lower light yield and has lower Z in comparison to BGO and CsI.

13.5.2.2 Energy resolution of scintillating crystals

Considerations on the energy resolution of totally absorbing detectors are discussed in detail in chapter 15, in sections 15.4.3 and especially in section 15.5.5 for electromagnetic calorimeters, which also include crystal calorimeters. Therefore here we only discuss some aspects which are specific for scintillating crystals.

The generation of scintillation photons is subject to statistical fluctuations (see also section 17.10.2). If there are no other dominating noise sources the energy resolution is mainly governed by the number of generated scintillation photons. Since N_γ is proportional to E the relative energy resolution becomes

$$\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{N_\gamma}} \propto \frac{1}{\sqrt{E}}\,. \tag{13.19}$$

For detectors based on the generation of statistically fluctuating quanta one finds parameterisations of the energy resolution in the form (see also eq. (15.33)):

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\,. \tag{13.20}$$

The stochastic term $a/\sqrt{E}$ describes the number fluctuations of primary quanta (here the scintillation photons), the second term b/E parameterises the energy independent noise contribution of the electronics, which hence contributes to the relative energy resolution σ_E/E proportional to $1/E$, and a constant term c takes account of other effects influencing the resolution such as calibration errors or a non-homogeneous response behaviour.

For scintillating crystals with large light yield such as NaI(Tl) or CsI(Tl) with about 50 000 photons per MeV of deposited energy, the uncertainty from statistical fluctuations in the number of photons is normally small in comparison to systematic effects and to the noise of the readout electronics. For these crystals one therefore usually finds a resolution parametrisation which falls off more weakly than $1/\sqrt{E}$ and includes systematic measurement contributions from calibrations and electronics:

$$\frac{\sigma_E}{E} \propto \frac{a}{E^{1/4}}\,. \tag{13.21}$$

As examples we compare BGO with about 8200 photons per MeV and CsI(Tl) with about 52 000 scintillation photons per MeV (see table 13.3) with each other. The energy resolutions achieved in experiments with crystal calorimeters for photons in the energy range from >10 MeV to $>$ GeV are [746]:

$$\text{BGO:} \qquad \frac{\sigma_E}{E} = \frac{2\%}{\sqrt{E/\text{GeV}}} \oplus 0.7\% \qquad \text{(L3)}$$

$$\text{CsI(Tl):} \qquad \frac{\sigma_E}{E} = \frac{2.3\%}{(E/\text{GeV})^{1/4}} \oplus 1.4\% \qquad \text{(BaBar)}$$

The trend of these resolutions as a function of energy is compared in fig. 13.23. At low energies the resolution is governed by the quantum fluctuations (first term), whereas at high energies the constant term dominates.

13.5.2.3 Crystal detectors for multi-photon final states

Inorganic crystals with sufficient absorption depth are suited for measurement of the energy of photons provided that their energy is completely absorbed in the crystal. Depending on the energy of the photon the absorption process involves only a few successive interaction processes, either for low energies by photoeffect, Compton effect, pair creation, as shown in the previous section or at higher energies by a succession of pair creation and bremsstrahlung processes leading to the formation of an electromagnetic shower (see section 15.2.1). Examples of crystal calorimeters in collider experiments are given in section 15.5. Large detectors whose main component are scintillating crystals are for example the Crystal Ball detector (NaI(Tl) crystals) [530,216], the Crystal Barrel detector (CsI(Tl) crystals) [67] or the BaF_2 calorimeter TAPS [740]. We choose here as a particularly representative example the Crystal Ball detector which first studied the charmonium ($c\bar{c}$ bound states) spectrum at the SPEAR storage ring at SLAC/Stanford (1978–1982), and from 1982 onwards the bottomonium ($b\bar{b}$ bound

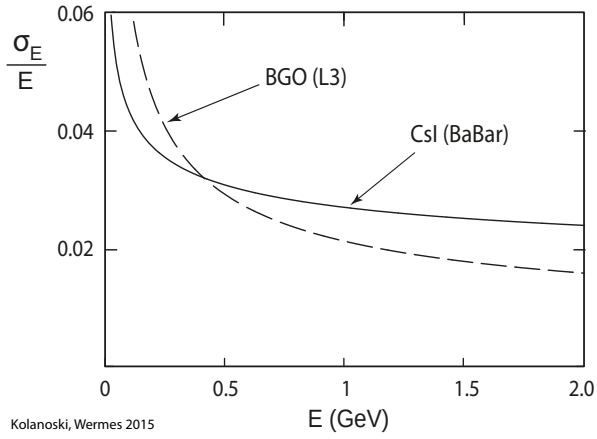


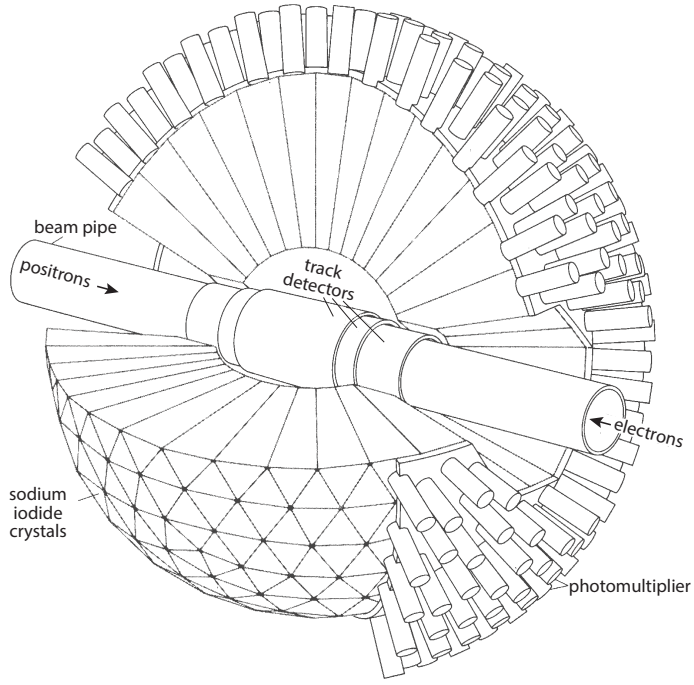
Fig. 13.23 Energy resolution as a function of photon energy for BGO and CsI(Tl) resolutions as quoted in the text.

states) at the DORIS storage ring at DESY/Hamburg. Later it was used for photon spectroscopy in synchrotron experiments at the Brookhaven National Laboratory [728] and in Mainz [729].

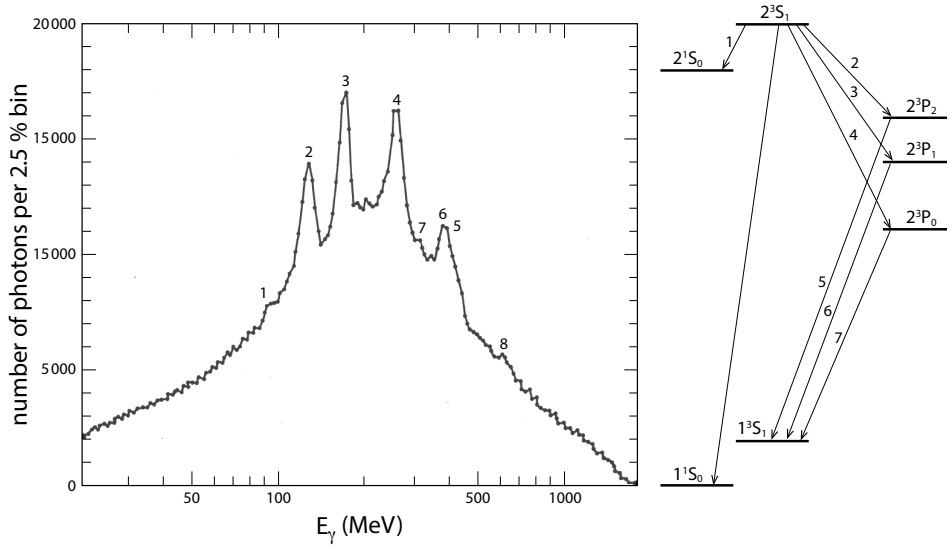
The Crystal Ball detector. In the *Crystal Ball detector* [530,216] a crystal sphere (fig. 13.24(a)) is formed by spherically arranging 672 NaI(Tl) crystals (table 13.3) which together with special crystals in the regions around the beam pipe cover 98% of the solid angle for photon detection. A typical energy threshold is around 20 MeV to suppress background not coming from e^+e^- beam collisions. The crystals were 16 X_0 deep. The resolution measured in the experiment is $\sigma_E/E = 2.8\%\ E/\mathrm{GeV}^{-1/4}$ [749]. Around 1980, when the Crystal Ball detector came into operation, its energy resolution for electromagnetically interacting particles was better by an order of magnitude than that of comparable detectors at the storage rings of that time. Therefore it was an excellent instrument for the detection of final states consisting of several photons as is the case for decays of heavy quark–antiquark states (charmonium, bottomonium) which could be measured with good resolution. The resolution not only allows for identification of particles decaying to two photons as $\pi^0/\eta \to \gamma\gamma$, but also for the identification of the transition photons between different states. Figure 13.24(b) shows the inclusive photon spectrum, measured at a centre of mass energy of 3.686 GeV, which corresponds to the mass of the charmonium state ψ' (spectroscopic notation 2^3S_1 in fig. 13.24(b)). One can recognise in the spectrum the transition lines between the different charmonium levels belonging to different angular excitations.

In the 1980s the development of CsI(Tl) crystals, having a still higher light yield than NaI(Tl) (see table 13.3), had progressed sufficiently to take over the role of NaI(Tl) for detectors. The Crystal Barrel detector [67] at the LEAR storage ring at CERN and later at the ELSA accelerator in Bonn is an example for photon energies in the 100 MeV to GeV range. These crystals (as well as BGO and others) have the big advantage to be much less hygroscopic than NaI(Tl) and to possess smaller radiation lengths X_0 and Molière radii (table 13.3).

For experiments operating at still higher energies, for instance at LEP, with correspondingly higher average photon energies, crystals with a larger density were preferred such as BGO for L3 [50] or lead glass for OPAL [63]. At even higher energies (LHC) the role of the light yield for the crystal choice becomes less important but instead absorption depth (high density, high Z) and radiation tolerance play the major role.



(a) Crystal Ball detector.



(b) Inclusive photon spectrum.

Fig. 13.24 (a) Arrangement of 672 NaI(Tl) crystals in the Crystal Ball experiment (inner ⊘ 25 cm, outer ⊘ 132 cm) for the detection of photon final states; (b) measured photon energy spectrum in which the γ transitions of the charmonium system are clearly visible. Adapted from [215, 216] with kind permission of the SLAC National Accelerator Laboratory.

This has led for example to the dedicated development of $PbWO_4$ crystals used in CMS [298] and ALICE [10] and of LYSO crystals for upgrades (see table 13.3).

13.6 Scintillating fibres

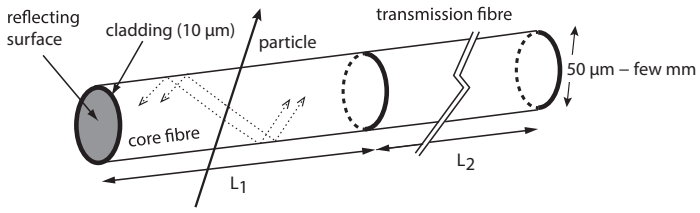
Scintillators, finely drawn in fibres and then bundled and arranged in layers, find applications in experiments as tracking detectors and for calorimeter readout. The fibres can have diameters significantly less than 100 µm, but typical diameters for tracking applications are between 250 µm and 1 mm. A trade-off must be chosen between spatial resolution and light yield. Fibre trackers can be alternatives to gas or semiconductor trackers when their main benefit, easy and robust assembly, outweighs other performance criteria as there are: spatial resolution (typically 80–200 µm), material thickness (typical are 10–15 layers with 0.25 % X_0 per mm thickness) and radiation tolerance. The hit resolution can be improved by staggering fibres, for example in double layers, but at a price of increasing the material budget. A detailed review on scintillating fibre (SciFi) properties is given in [839].

13.6.1 Fibre materials and properties

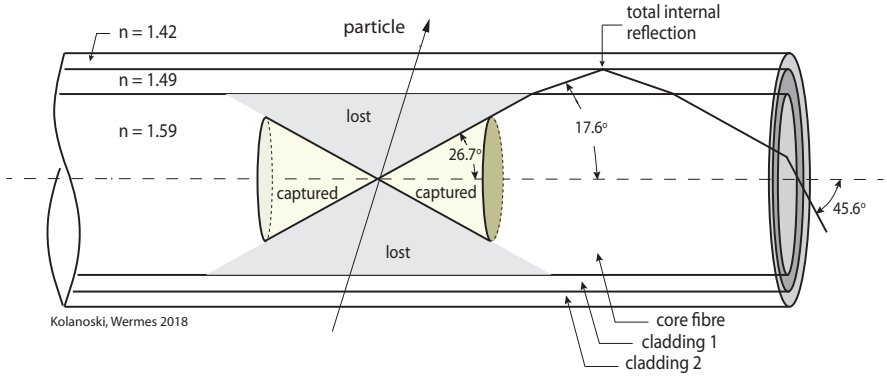
Fibres can be made of scintillating glasses or plastic scintillators; capillaries filled with liquid scintillator can also be considered. Typical dimensions are depicted in fig. 13.25. Often a scintillating fibre is connected to a transmission fibre for light transmission to the photosensor. The physics and technical challenges when using scintillating fibres as tracking detectors are as follows:

- For normal incidence of a particle into a fibre a sufficient amount of light must be generated that stays in the fibre by total reflection and can be detected at the fibre end.
- Self-absorption of light in the fibre should be low such that for long trackers the transmission lengths can reach several metres.
- The generation point of the scintillation light should be as close as possible to the location of the energy deposition by the particle (dE/dx) and should not be delocalised (the choice of the scintillation dye is important, see page 509).
- Optical cross talk between neighbouring fibres should be vanishingly small.
- Optical coupling from the scintillating fibre into a subsequent transmission fibre should be without substantial losses.
- The fibre thickness must be optimised regarding spatial resolution (small diameter) and light yield (large diameter, more photons and less reflections).
- At the end of the fibre the photons must be detected with high quantum efficiency (and if applicable with high rate capability).
- The position of the fibres within the tracking system must be precisely known.

Many organic scintillator materials contain several dye additions as wavelength shifters (ternary or quaternary systems, see page 507 in section 13.2.2.1). The transmission length between emission and re-absorption from the first to the second dye typically lies in the order of 200 µm. While this is no problem for large volume scintillators devoid of good spatial resolution, in scintillating fibres it leads to the so-called 'non-locality' problem. The scintillation light can couple into neighbouring fibres thus reducing the spatial resolution as well as the detection efficiency. Suitable dye compositions have been found that do not show the non-locality problem. In essence,



(a) Fibre geometry.



(b) Captured light.

Fig. 13.25 (a) Structure of a scintillating fibre glued to a readout fibre for light transmission; typical geometrical dimensions are indicated. (b) Illustration of the captured light fraction for a typical fibre with two cladding layers.

one has to relinquish the second dye addition which provides an energy transfer resulting in a large Stokes shift (see fig. 13.5), but usually also features large transfer lengths. Instead the light of the first dye should be used onto which the transfer takes place via fast Förster transitions coming together with very short transfer lengths (see section 13.2.2.1). As a penalty smaller Stokes shifts result, thus reducing the fibre's light attenuation length. A familiar example is PMP (1-phenyl-3-mesityl-2-pyrazoline) which is successfully used even in fibres with small diameter ($< 50\,\mu m$) [839].

Fibres using organic scintillator material are preferred over inorganic scintillating glass fibres regarding speed, material thickness and light yield. Plastic scintillators are simpler in production and more robust than capillaries filled with liquid scintillator, but capillaries can have higher light yield and can provide better radiation tolerance because the scintillating liquid can be exchanged (see e.g. [104]). For tracking detectors in particle physics experiments (see section 13.6.2) long fibres with diameters in the sub-millimetre range, balancing photon yield and spatial resolution, are preferred.

Light from tracks crossing the fibre is propagated by total reflection if the incidence angle with respect to the fibre's surface normal exceeds the critical angle defined by

$$\sin\,\theta_{\text{crit}} = \frac{n_{\text{clad}}}{n_{\text{core}}}\,, \tag{13.22}$$

resulting in an acceptance cone yielding the fraction of light kept in the fibre by total reflection (trapped fraction) as:

$$f_{\text{trap}} = \frac{1}{4\pi} \int_{0^\circ}^{90^\circ - \theta_{\text{crit}}} 2\pi \sin\theta \, d\theta = \frac{1}{2} \left(1 - \frac{n_{\text{clad}}}{n_{\text{core}}} \right) , \tag{13.23}$$

as illustrated in fig. 13.25(b). Optimal yield for total reflection would in principle be obtained for a fibre–air interface with ($n_{\text{air}} = 1$). However, surface imperfections like cracks, or the attachment of a fibre to neighbouring fibres or to mounting surfaces or glue, cause large yield losses such that a cladding of the fibre core is necessary [839]. The cladding should have an index of refraction as low as possible.

A typical scintillating fibre consists of the base material (e.g. polystyrene, PS) furnished with PMP dye which emits in the violet (430 nm) with a short decay time (< 1 ns). The index of refraction is $n = 1.59$. For the cladding an often used material is PMMA (polymethyl acrylate, $n = 1.49$). Improved performance has been achieved by double-cladding, where the outer layer has an even lower index of refraction, for example a fluorine-acrylate polymer ($n = 1.42$) [578]. The PMMA layer then serves as a mechanical interface to which the inner and outer layer surfaces can be well bonded. For single cladding and round fibres with $n_{\text{clad1}} = 1.49$ and $n_{\text{core}} = 1.59$, the trapped fraction is 3.1%. A second cladding with $n_{\text{clad2}} = 1.42$ renders a shallower incidence angle such that the acceptance fraction becomes 5.3% [578]. Note that the obtained efficiency is hence determined by the ratio of refractive indices of the core and the outer cladding.

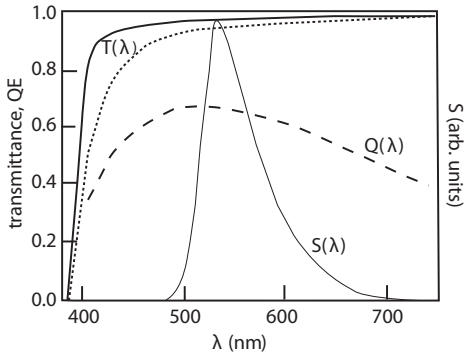
The light is attenuated over the fibre length following an exponential law $I(x) = I_0 \exp(-x/\Lambda_{\text{att}})$. The attenuation length, Λ_{att}, has wavelength-dependent contributions from different phenomena such as Rayleigh scattering, absorption by dyes as well as radiation-induced effects creating additional scattering and absorption centres in the base plastic. Radiation also degrades the abundance of scintillating fluors. Mitigation is obtained by adding radiation robust fluors with large Stokes shift and good quantum efficiency in a base plastic scintillator of good optical quality (typically polystyrene). The attenuation is generally stronger at shorter wavelengths (blue, UV). The emission spectrum therefore shifts towards larger wavelengths with time and irradiation and the photosensor selection must be adapted to this effect. Typical values obtained are $\Lambda_{\text{att}} = 3$–5 m.

The mean number of photoelectrons detected in the readout element is given by equation eq. (13.5) on page 13.5, which can approximately be expressed as

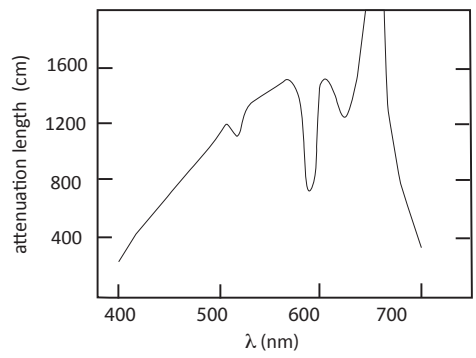
$$\langle N_{pe} \rangle \approx \langle S \rangle \langle T \rangle \langle Q \rangle \, \Delta\lambda \, , \tag{13.24}$$

where the averages are taken of S, the source strength (number of produced photons per wavelength), T the light transfer efficiency and Q the quantum efficiency of the photodetector. Figure 13.26(a) shows the spectral dependences of the fluorescence emission, the transmission and a typical quantum efficiency. In fig. 13.26(b) the attenuation length is plotted as a function of the wavelength for a typical fibre.

For a numerical estimate [839] we take a PS fibre with 830 µm active diameter d and a cladding thickness of 30 µm. The fibre core has a diameter of 770 µm. A particle traverses on average a fibre core distance $\Delta x = \pi/4 \, d = 605$ µm for straight incidence and deposits on average $\langle dE/dx \rangle = 2$ MeV/cm, that is, 120 keV on the path through the fibre core. For an excitation in PS an energy of 4.8 eV is needed, the fluorescence light yield upon de-excitation amounts to about 4%. Hence about 10^3 photons with a peak wavelength of about 530 nm are emitted according to a spectrum, as shown in fig. 13.26(a). The transfer efficiency $\langle T \rangle$ can be estimated as follows, if one assumes



(a) Source strength S, transmission T, and quantum efficiency Q.



(b) Attenuation length.

Fig. 13.26 (a) Illustration of typical spectral relations in scintillating fibres (e.g. polystyrene with PTP and 3HF). Fluorescence spectrum S (thin solid line, arbitrary units), transmission curve T before (thick solid line) and after radiation damage (1 kGy) and annealing (dotted line), typical quantum efficiency Q of a photodiode detector (dashed). (b) Attenuation length (example) as a function of the wavelength for a fibre with double layer cladding (drawn with input from [839]).

an arrangement with a scintillating fibre of length L_1 followed by an optical fibre of length L_2 (see fig. 13.25):

$$\langle T \rangle = \frac{\Delta\Omega}{4\pi} \left[A(x) + R\, A\,(2L_1 - x) \right] T_1\, A_2(L_2)\, T_2\,, \tag{13.25}$$

where x is the position of the ionisation. $T_{1,2}$ are the transmissions of scintillating and optical fibre, respectively, R is a non-zero reflection factor if only one side is mirror coated and $A(x)$ describes the absorption in the fibres. Using the following values [839]

$L_1 = 3\,\mathrm{m}$ $\qquad$ $L_2 = 8\mathrm{m}$
$A_1(x = 3\mathrm{m}) = A_1(2L_1 - x = 3\mathrm{m}) = 0.55$ $\qquad$ $A_2(5\mathrm{m}) = 0.45$
$R = 0.9$ $\qquad$ $T_1 = T_2 = 0.9$
$\Delta\Omega/4\pi = 0.05$

and a quantum efficiency of 70% for the photosensor (silicon photodiode or APD, see table 10.2), one obtains $\langle T \rangle = 1.5\%$ and $N_{pe} \approx 10$ for 1000 primary photons produced. This rough estimate makes it clear that minimisation of the different loss contributions is important.

13.6.2 SciFi detectors

In particle physics experiments scintillating fibres are used for example as tracking detectors, including 'active targets', and for readout of sampling calorimeters (see e.g. figs. 15.27 and 15.29(b) in chapter 15). Designed as bundles with typically 10^5–10^6 fibres they are spatially sensitive track detectors. The parallel development of multi-channel photodetectors like multi-anode PMTs, CCD cameras or SiPMs with individual pixels as (single) photon counting sensors (see section 10.5) have made SciFi trackers possible.

Active target. In active targets the target material (the fibres) and the detector form one unit. They can be formed from coherent arrays of scintillating glass fibres,

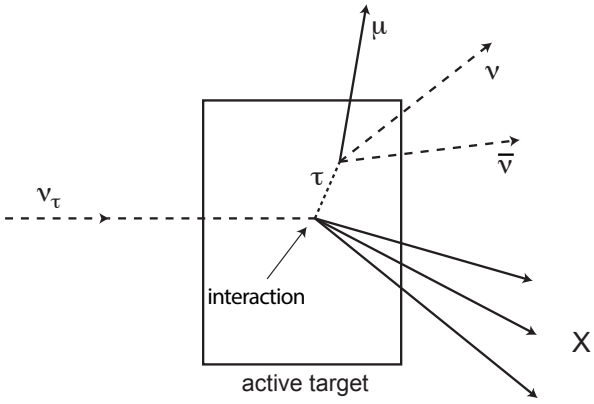
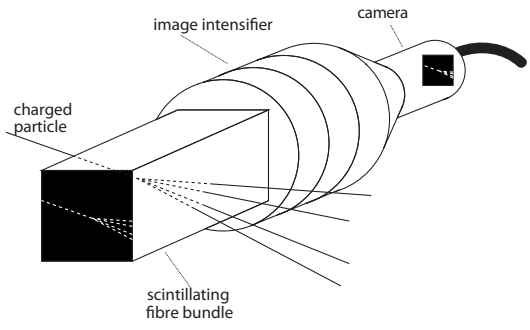
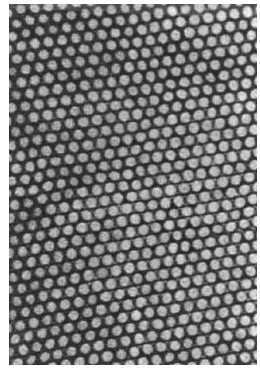


Fig. 13.27 Illustration of ν_τ detection in an 'active target'. Here the decay $\tau \to \mu\nu\bar{\nu}$ is assumed.



(a) Fibre bundle detector.



(b) Bundle of capillaries.

Fig. 13.28 A scintillating fibre bundle as a tracking detector (schematic). The homogeneously arranged fibres are captured by an image intensifier (see footnote 9 on page 543) and read out by a camera (e.g. CCD or other). (b) Capillary bundle with 20 μm diameter fibres and wall thickness of 2 μm, filled with liquid scintillator.

plastic fibres, or capillaries filled with liquid. A possible application is τ-neutrino detection, where τ leptons are sought for in a target hit by a ν_τ. If exposed to a ν_μ beam, as done for example by the CHORUS experiment [380] and the OPERA [39] experiment, such a reaction indicates a transformation of a ν_μ into a ν_τ during flight:

$$\nu_\mu \longrightarrow \nu_\tau + N \to \tau + X \quad . \\ \hookrightarrow \mu\bar{\nu}\nu$$

Here the τ lepton is identified by detection of its decay vertex with its distance from the production point lying in the sub-millimetre to millimetre range, as illustrated in fig. 13.27.

In several experiments, e.g. in CHORUS [380,104] and DONUT [621], the produced tracks from the τ decay are detected by SciFi trackers behind the target. CHORUS in its design phase had also discussed an active target [839,103,104] sketched in fig. 13.28 and shown here for a general illustration. One end of the fibre can be mirror coated

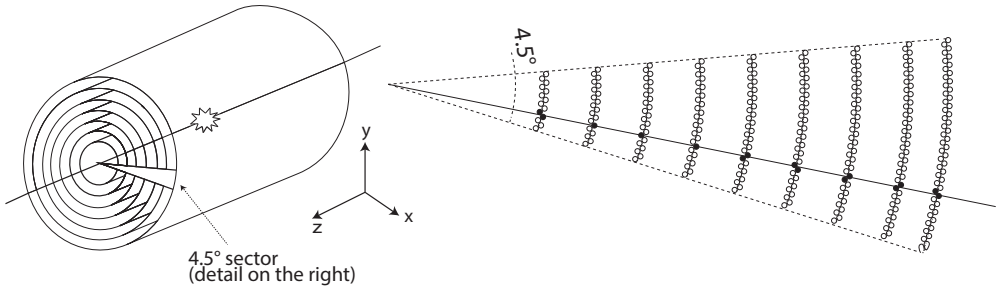
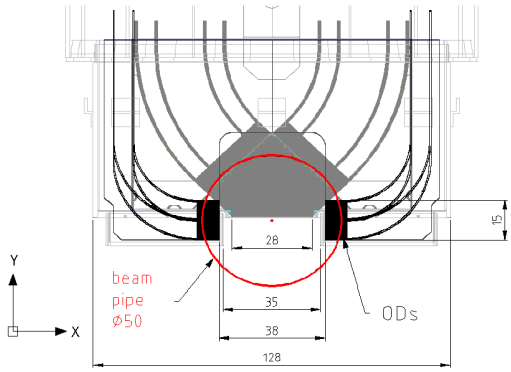
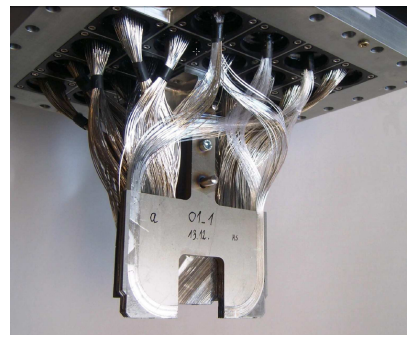


Fig. 13.29 Use of scintillating fibres as a tracking detector in a collider experiment (D0) (from [899] with kind permission of Elsevier).



(a) Forward spectrometer ALFA built with fibre detectors.



(b) Photo showing the fibre arrangement.

Fig. 13.30 Use of scintillating fibres (0.5×0.5 mm^2, squared) as a tracking detector in extreme forward direction of the ATLAS detector at the LHC. Very small angle forward spectrometer ALFA [694]. The fibre detector has 10 planes with two fibre faces (grey) each, turned by 45° with respect to the vertical of the beam axis (y axis). The planes can be moved to less than one millimetre close to the proton beam. The circle represents the beam pipe. The fibres are moved out (right picture) and read out by multi-anode PMTs (source: CERN/ATLAS ALFA Collaboration).

while the other is attached to an image intensifier,[9] or another camera system which has an entrance window suitable for the fibre bundle.

SciFi tracker. Scintillating fibres have been used as cylindrical trackers in collider experiments, for example in the experiments UA(2) [437], D0 [256,14,899] and L3 [50]. Considering advantages and disadvantages compared to other tracking detectors (especially semiconductor or gaseous micropattern detectors, see sections 8.6.2 and 7.9) one has to balance the benefits

– flexibility in size and shape (planar, cylindrical)

[9] An image intensifier is an electron tube which amplifies low light levels below the visibility threshold from various wavelengths into visible light at a single wavelength by photon to electron conversion, electron amplification in a high field and then back conversion into photons.

– (low) mass ($X_0 = 42.4\,\text{cm} \Rightarrow 1\,\text{mm fibre} = 0.25\%\ X_0$)
– (almost) 100% active and uniform material
– fast signals ($\tau_{dec} \sim$ ns)

against the less beneficial characteristics:

– only medium resolution ($\mathcal{O}(50\text{–}100)\mu\text{m}$),
– small signals (few photoelectrons),
– limited radiation tolerance.

Figure 13.29 shows the central fibre tracker (CFT) of the D0 experiment [256] at the $p\bar{p}$ collider Tevatron. About 90 000 double clad fibres, between 1.66 m and 2.52 m long, are arranged in eight concentric double layer cylinders at a distance between 20 cm and 52 cm around the interaction point. The fibre diameter is 835 µm. The achieved space resolution is in the order of 100 µm. The fibres are read out with a precursor variant of SiPMs discussed in section 10.5 (called VLPCs [971], visible light photon counters). They are based on 1 mm^2 Si pixels operating as avalanche diodes in linear mode (see fig 10.20 on page 426) with typical amplifications of 40 000 and a quantum efficiency of around 80% yielding even single photon detection capability. Due to the small energy distance between impurity and conduction bands of the VLPC silicon bulk high dark count rates were prevented by operating them at cryogenic temperatures ($T = 9\,\text{K}$). Yields of 10 photoelectrons per fibre and hit efficiencies above 99% were achieved in full operation [899].

Owing to their mechanical robustness and above all because of their insensitivity against electromagnetic interference, fibre trackers are attractive in extreme forward regions of collider detectors near the beam line, usually placed inside so-called 'Roman Pots', dedicated mechanical housings with handling tools to allow for operation of detectors close to the beam. An example is the precise detection of elastically scattered protons, for example for luminosity determination, at the LHC. Figure 13.30 shows a schematic and a photograph of the forward spectrometer ALFA which measures protons scattered under very small angles in ATLAS (LHC) at a distance of 240 m from the interaction point with a point resolution of about 20 µm.

The LHC experiment LHCb replaces its tracker by scintillating fibre planes with a total area of 360 m^2 [578] in six-layer mats using 250 µm diameter fibres read out by SiPMs (see section 10.5). The SciFi tracker offers lower hit occupancy and faster response than the current outer tracker consisting of straw tubes and a smaller material budget for both outer and inner (Si microstrip detectors) tracker parts. Despite shielding, the detector is exposed to an ionising radiation fluence of about $6 \times 10^{11}\ \text{n}_{\text{eq}}/\text{cm}^{-2}$ corresponding to 35 kGy. The damage mainly affects the SiPMs leading to shot noise which produces noise count rates in excess of 1 GHz. To cope with these radiation conditions the SiPMs are operated at $-40\,^{\circ}\text{C}$ and extensive annealing periods, in which the damage is partly cured, are foreseen [578].

14
Particle identification

14.1 Overview

14.1.1 Identity of a particle

By 'particle identification' we understand the determination of quantities that allow us to infer the identity of a particle. A particle's identity is principally defined by its mass, its lifetime and its quantum numbers, such as charge, spin, parity and so on. To classify a newly discovered particle one aims to determine as many of these quantum numbers as possible. Two historical examples are shown in fig. 14.1. To identify already known particles in a specific reaction (an interaction or a decay) it often suffices to measure one property, for example the mass.

Particles so short-lived that they decay instantly after their creation, without interaction with the detector, can only be identified indirectly through their decay products. This fact renders it a fortiori important that long-lived particles such as photons, electrons, muons, pions, protons, and kaons, that are either stable or possess decay lengths larger than the physical dimensions of the detector, are identified by detectors. Deducing the identity of an (already known) particle from the 'signal' of one or several detector components is what we understand under the term 'particle identification' in this chapter. Their identity is deduced from the detector's 'response' occurring when the particle interacts with the detector medium, such as a shower development initiated by an electron or photon, specific energy loss or bremsstrahlung of a charged particle, or the penetration property of a muon. Beyond this, particles with lifetimes long enough that they travel a measurable distance before they decay can also be identified.

14.1.2 Methods of particle identification

The most important methods employed in accelerator and astroparticle experiments to identify a particle's characteristic properties can be summarised as follows:

(a) Charged particles: by measuring momentum p and velocity β the mass can be deduced:

$$m = \frac{p}{\gamma\beta}\,. \tag{14.1}$$

(a) Ω^- generation and decay in a bubble chamber (b) Discovery of the τ-lepton

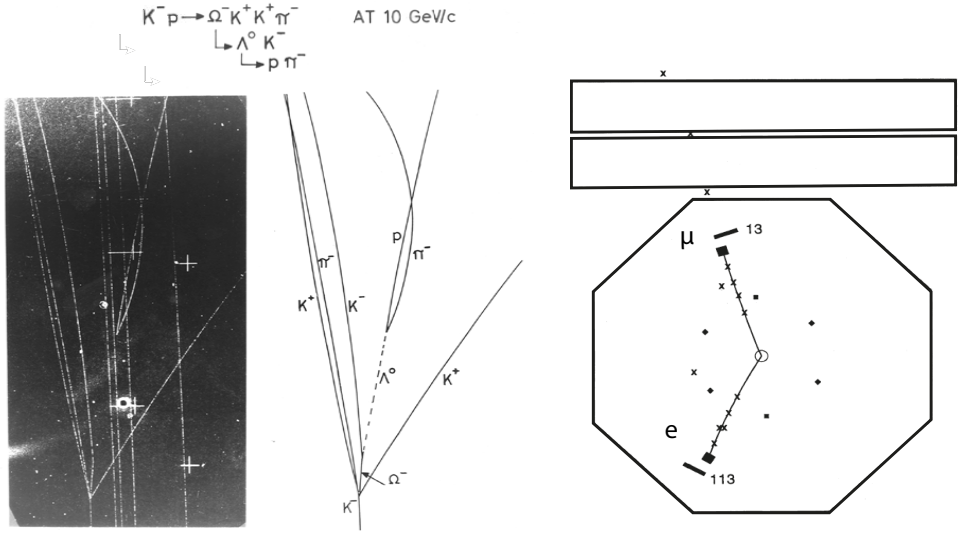


Fig. 14.1 Two historical examples for identification of a new particle. (a) The Ω^- baryon [156] is identified in a bubble chamber by its long lifetime, its kinematically reconstructed mass and associated strangeness production (source: CERN [284]). (b) τ-lepton pairs were first identified in e^+e^- collisions by observing a muon and an electron track in a cylindrical proportional chamber [774, 773]. The momenta of the tracks are unbalanced, indicating unidentified particles (neutrinos) taking part in the event. The upper track is identified as a muon since it traverses the iron yoke of the detector; the bottom track is an electron because 113 units of energy are deposited in the shower counter, to be compared to the muon energy of only 13 energy units. Taken from SLAC Report SLAC-PUB-5937 with kind permission by the SLAC National Accelerator Laboratory.

The momentum can be measured in a magnetic field (section 9.3) and the Lorentz variables β, γ or $\beta\gamma$ can be determined by measuring:
- time of flight, (section 14.2.1),
- specific energy loss by ionisation dE/dx (section 3.2.1),
- Cherenkov radiation (chapter 11),
- transition radiation (chapter 12).

(b) Masses (and sometimes also other quantum numbers) of decaying particles can be determined by kinematic reconstruction of the decay products:

$$m = \sqrt{\left(\sum_i p_i\right)^2}\,, \tag{14.2}$$

where p_i are the 4-momenta of the decay products.

(c) Electrons and photons create electromagnetic showers. For electrons (and also for positrons) the measured shower energy E is about equal to the (magnetically determined) momentum p, such that $E/p \approx 1$. Photons, on the other hand, do not leave a trace in the ionisation detector.

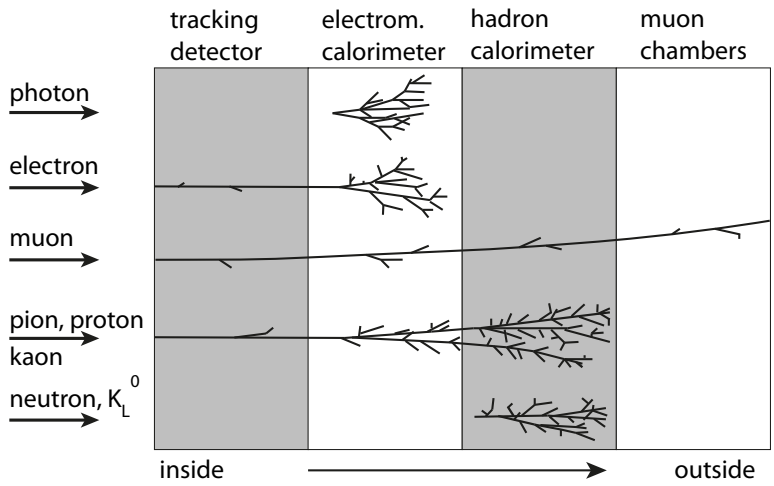


Fig. 14.2 Schematic illustration of particle identification in a detector at a high energy accelerator/collider like, for example, the LHC (adapted from [665]).

(d) For muons their capability to penetrate thick absorber materials is a characteristic.
(e) Hadrons can (at sufficiently high energies) be recognised via hadron showers.
(f) Particles with lifetimes in the order of 10^{-10} s like K_S^0 and Λ have a characteristic 'V-shaped' decay topology after a flight distance of several tens of centimetres (see e.g. Λ^0 in fig. 14.1(a)).
(g) Particles decaying on a time-scale of picoseconds (e.g. weakly decaying heavy bottom or charm quarks or τ leptons) can be identified by measuring the flight distance from production to decay by measuring the traces of the decay products.

The experimental techniques employed in particle identification change with the energy of the particles to be identified. Methods, employed very successfully for low-energy particles, as for example time-of-flight or dE/dx methods, are not suitable at high energies, for example for those particles created in LHC proton–proton (pp) collisions, or they are too complex and elaborate to build (e.g. Cherenkov detectors) when full solid angle coverage is aimed for. For this reason the LHC pp-collision experiments ATLAS and CMS restrict themselves to identifying

– jets as signatures of quarks or gluons,
– leptons (electrons, muons),
– photons,
– secondary vertices as signatures of heavy quarks or heavy leptons.

The other two LHC detectors, LHCb and ALICE, which do not aim for full angle coverage, employ Cherenkov detectors for particle identification.

A typical sequence of detectors in high energy physics experiments with their capabilities to distinguish different particle species is illustrated in fig. 14.2.

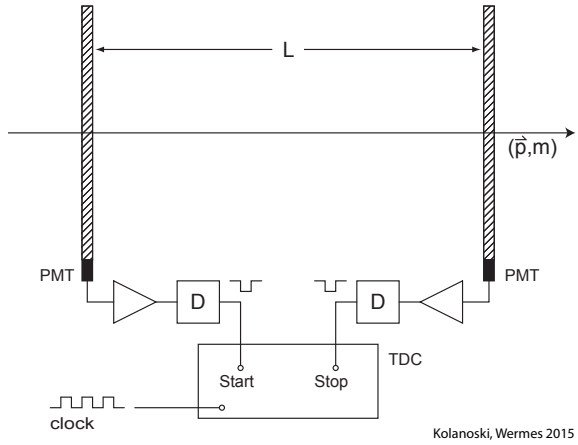


Fig. 14.3 Principle of the time-of-flight measurement (TOF), in this example realised by plastic scintillators with PMT readout.

14.2 Identification of long-lived charged particles

In this section we describe the methods to identify long-lived ('stable') particles. The physics foundations and the detection techniques are described in respective chapters: chapter 3 (dE/dx measurement); chapter 11 (Cherenkov radiation); chapter 12 (transition radiation).

14.2.1 Time-of-flight measurements

The flight time of a particle between two detectors (TOF, *time-of-flight*) directly yields the velocity of the particle. For example, the flight time between two detectors at a distance of 2 m for π, K, p, each with momentum $p = 500\,\mathrm{MeV}/c$ is 6.2 ns, 8.5 ns, 12.7 ns, repectively (light needs 6.7 ns).

The principle of TOF measurement is shown in fig. 14.3 and a more detailed review can be found in [614]. A particle with momentum p and mass m traverses two detectors positioned at a distance L from each other. One of the detectors generates the start, the other the stop signal. The time difference between the signals is digitised by a TDC (*time-to-digital converter*, section 17.7.3) with a resolution given by the TDC clock. From the flight duration

$$\Delta t = \frac{L}{\beta c} = \frac{L}{c}\sqrt{\frac{p^2+m^2}{p^2}} \tag{14.3}$$

with m, p given in units of eV, one obtains the velocity $\beta = \dfrac{L}{\Delta t\, c}$ and, knowing the momentum, with $(\beta = p/m\gamma)$ the squared mass:

$$m^2_{\mathrm{TOF}} \;=\; p^2\left[\left(\frac{\Delta t\, c}{L}\right)^2 - 1\right]. \tag{14.4}$$

Due to the finite measurement resolution the squared mass from the TOF measurement m^2_{TOF} can in principle also become negative. Hence it is not reasonable to extract the root.

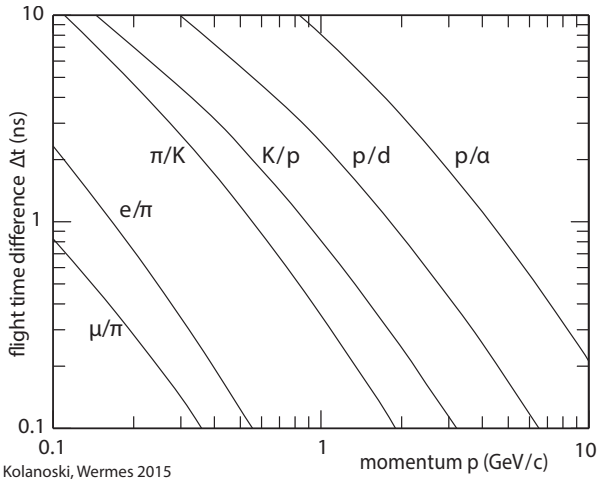


Fig. 14.4 Flight-time differences for various pairs of particle species at a flight distance of $L = 1\,\mathrm{m}$.

For the difference in flight times of two particles of equal momenta but different masses $m_2 > m_1$ one obtains with (14.3) in relativistic approximation ($p^2 \gg m^2$):

$$\frac{\Delta t_2 - \Delta t_1}{L} \approx \frac{1}{2c\,p^2}\left(m_2^2 - m_1^2\right) = 1667\,\frac{m_2^2 - m_1^2}{p^2}\,\frac{\mathrm{ps}}{\mathrm{m}}\,. \tag{14.5}$$

In fig. 14.4 the flight-time differences for different particle species are shown for a length $L = 1\,\mathrm{m}$. For 1 GeV pions ($m = 139.6\,\mathrm{MeV}$) the comparison to kaons ($m = 494\,\mathrm{MeV}$) yields a flight-time difference of about 1.9 ns over a distance of $L = 5\,\mathrm{m}$.

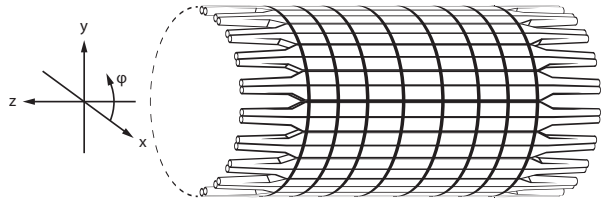
The mass resolution achievable with the TOF method has contributions from the uncertainty on momentum p, flight time Δt and distance L added in quadrature ($\oplus$):

$$\frac{\sigma_m}{m} = \frac{\sigma_p}{p} \oplus \gamma^2 \left(\frac{\sigma_{\Delta t}}{\Delta t} \oplus \frac{\sigma_L}{L}\right)\,, \tag{14.6}$$

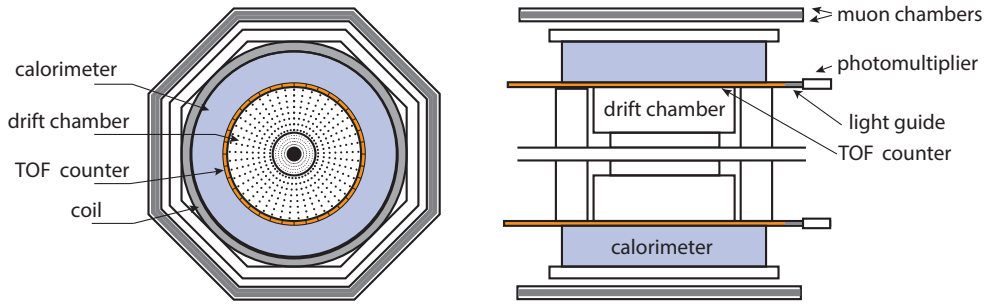
where $\gamma = (1-\beta^2)^{-\frac{1}{2}}$ with $\beta = L^2/c^2\Delta t^2$. Typically, p is known with an accuracy in the order of a per cent and $\sigma_L/L \approx 10^{-3}$, such that the mass resolution is predominantly given by the time measurement, $\frac{\sigma_m}{m} \approx \gamma^2 \frac{\sqrt{2}\sigma_t}{\Delta t}$, where σ_t is the time resolution of the system.

Typical Δt values are in the order of nanoseconds. In order to reach mass resolutions of 10% or better, time resolutions below 100 ps must be aimed for. Detector types fast enough to achieve such resolutions are for example scintillators, usually plastic scintillators, or thin, gas-filled detectors with fast amplification, like RPCs (*resistive plate chambers*, see section 7.7.3).

The time resolution technically achievable in timing detectors limits the usage of the TOF technique for particle identification. Particle identification by TOF is often used when predominantly low momentum particles ($\sim \mathrm{GeV}/c$) are the products of a reaction as, for example, in experiments at colliders with centre-of-mass energies smaller than about 10 GeV like CESR, PEPB, KEKB, but also in heavy ion experiments at RHIC and LHC. For high energy collider experiments at LEP, HERA, Tevatron, or LHC (except heavy ion experiments), where generally larger particle momenta are of interest, the TOF technique has mostly not been used.



(a) Cylindrical counter arrangement.



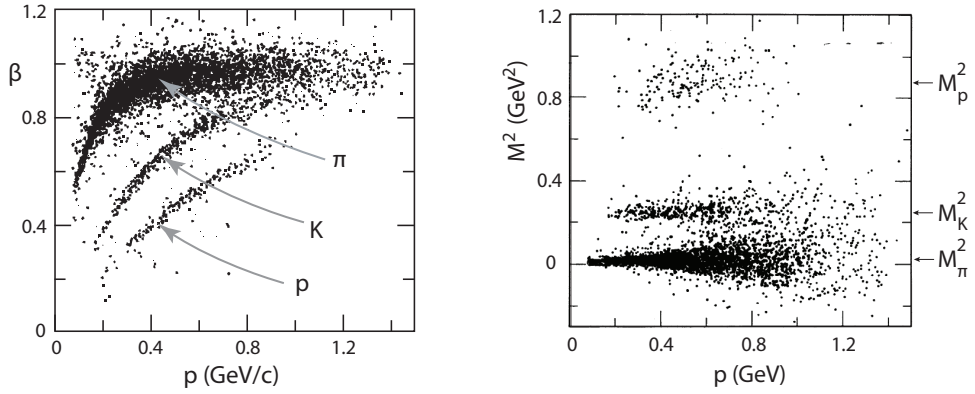
(b) Detector front and side views.

Fig. 14.5 Example of a time-of-flight counter system of a collider experiment (Mark III detector [185] at the 4 GeV storage ring SPEAR at SLAC). The TOF counters are arranged in between the central drift chamber and the calorimeter. Adapted from [257] with kind permission of Elsevier.

At colliders, the TOF detectors are arranged cylindrically as well as in 'endcaps', at a distance of 1–2 m from the interaction point parallel to the beam line, as shown for a typical application in Fig. 14.5. The TOF start signal is given by a so-called *beam pick-up* signal that is induced on electrodes attached to the beam pipe when beam bunches pass by, thus having a fixed time correlation with the collisions. The stop signal is the TOF system signal produced from one of the scintillation counters or another detector capable of precise time marking. The resolution typically achieved in large systems lies between 80 ps and 200 ps. Hence, for typical distances to the interaction point of 1–2 m, the TOF method is only useful for particle identification for momenta $\lesssim 2$ GeV.

Figure 14.6(a) shows the particle velocity β calculated from the time differences measured by the TOF system of fig. 14.5 as a function of momentum, and fig. 14.6(b) shows the squared mass calculated from (14.4) [257]: protons, kaons and pions can be separated with 2σ significance for momenta up to 1.2 GeV/c.

The mass resolution σ_m/m in (14.6) is governed by the time resolution σ_t. The latter contains contributions from intrinsic fluctuations in the signal process σ_{gen} (e.g. the light generation process in the case of scintillation detectors), geometrical light travelling differences caused by different entrance points and angles of particles into the detector σ_{tr}, from signal time fluctuations in the readout system (e.g. fluctuations in photoelectron arrival times in a photomultiplier σ_{PMT}), as well as time fluctuations σ_{el} in the electronics [489]:



(a) β–p plot. (b) TOF squared mass.

Fig. 14.6 π, K, p separation by TOF measurement (source: Mark III experiment, SLAC): two-dimensional representations of (a) the particle velocity β and (b) the squared mass from the TOF measurement as a function of the particle momentum [257].

$$\sigma_t = \sqrt{\frac{\sigma_{\text{gen}}^2 \; + \; \sigma_{\text{tr}}^2 + \; \sigma_{\text{PMT}}^2}{N_{\text{eff}}} \; + \sigma_{\text{el}}^2} \; . \tag{14.7}$$

N_{eff} here is the number of photons, respectively photoelectrons, effectively contributing to the time critical signal.

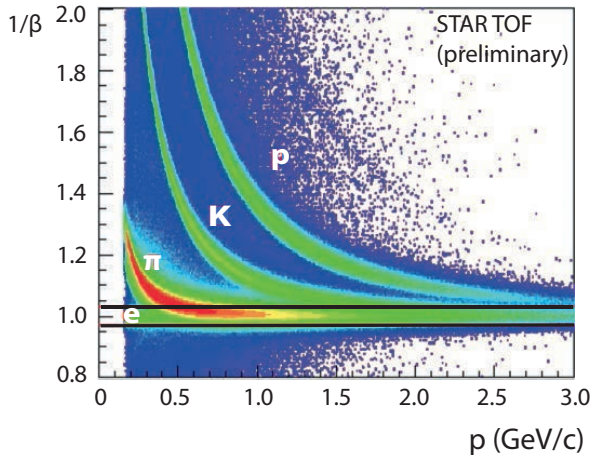
From (14.4) we can with $\Delta t \approx \frac{L}{c}\,(1+\frac{1}{2}\,m^2/p^2)$ quantify the 'separation power'; this is the number of standard deviations by which the (TOF) measurement distributions of two masses are separated. For two particles of the same momentum p and different masses $m_2 > m_1$ the n-sigma separation power is:

$$n_\sigma(\text{TOF}) = \frac{m_2^2 - m_1^2}{\sigma_{m^2}} \approx \frac{\Delta t_2 - \Delta t_1}{\sigma_{\Delta t}} \approx \frac{L}{2\sqrt{2}\,c\,p^2\,\sigma_t}\,\left(m_2^2 - m_1^2\right) \; , \tag{14.8}$$

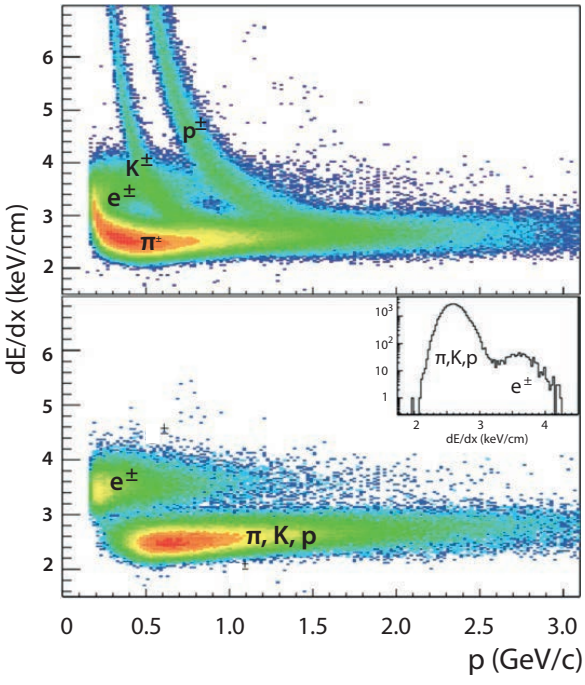
where $\sigma_{\Delta t} = \sqrt{2}\,\sigma_t$ is the resolution of the time-difference measurement with individual resolution σ_t. For a large TOF system with many counters, as in fig. 14.5(a), further uncertainties contribute, for example dispersions in how the time digitising clock is distributed to all detector units, as well as inter-calibration errors. In a collider experiment σ_{t_1} is usually the uncertainty in the knowledge of the collision time t_1.

In typical TOF detector systems, as the one in fig. 14.5(a), 2–3 m long plastic scintillators are used that are relatively thick (3–5 cm) in order to provide high light yield per traversing particle. In order to achieve time resolutions of 120–180 ps the counters are read out on both sides R, L and the time of the particle passage through the counter (TOF stop signal) is obtained by electronic averaging (*mean timing*) $t_{\text{stop}} = (t_R + t_L)/2$. Thus the dependence on the entrance point of the particle into the counter is eliminated to a large extent.

Fast signals and precise time marks can also be achieved by gas-filled detectors using thin gas volumes and high resistance electrodes (see section 7.7.3). In the STAR experiment at the ion storage ring RHIC for part of the central barrel detector *multigap resistive plate chambers* (MRPCs, see section 7.7.3 and fig 7.17) are used to precisely



(a) TOF measurement.



(b) dE/dx measurement.

Fig. 14.7 Particle separation in the STAR experiment (from [997], with kind permission of IOP Publishing). (a) Measurement of the time of flight using MRPC chambers; (b) measurement of the specific ionisation. In the bottom plot of (b) a cut in the flight time of $|1 - 1/\beta| < 0.03$ has been applied (line pair in (a)).

measure the time mark of particles, for example from (d+Au) collisions, over a distance of about 5.4 m [1007]. The time resolution is about 120 ps. Figure 14.7(a) shows the measured velocity, represented as $1/\beta$, as a function of the particle momentum. A clear separation of protons, kaons and pions up to a momentum of about 1 GeV/c is possible, whereas electrons can hardly be distinguished from pions. The latter can be achieved, however, by a simultaneous measurement of the specific ionisation [1007] (see below and fig. 14.7(b)).

The ALICE experiment at the LHC has a TOF detector system using MRPCs,

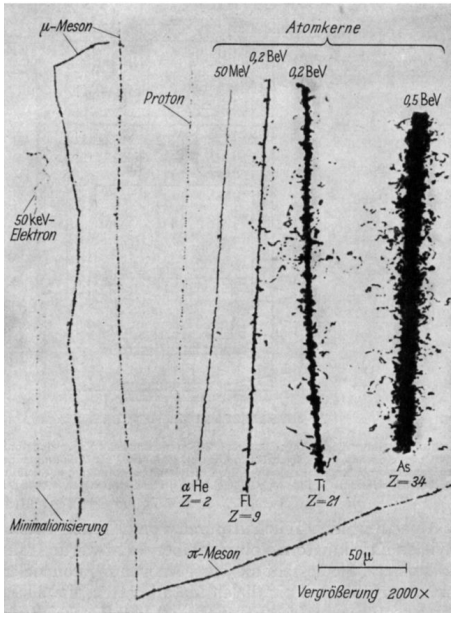


Fig. 14.8 Energy loss by ionisation: tracks of particles with different masses and energies in a photoplate. The energy labels are energies per nucleon using the out-dated unit BeV corresponding to today's GeV. Source: [410], with kind permission of Springer Science+Business Media. The labelling is left as in [410] (in German). The magnification is ×2000. The track on the very left is from a minimum-ionising electron.

arranged in two stacks of six planes each separated by 250 μm gaps. In-system time resolutions of about 85 ps are obtained at an intrinsic resolution of 50 ps [80]. So-called 'instrinsic' resolutions are understood as resolutions which have been cleaned from all other effects deteriorating the measurement, like electronics contributions, geometrical effects, calibration inhomogeneities, etc.

14.2.2 Measurement of specific ionisation (dE/dx)

In chapter 3 we described the mean energy loss dE/dx of a charged particle by ionisation of a traversed medium, expressed by the Bethe–Bloch formula (3.25). The amount of energy loss as a function of $\beta\gamma$ is approximately the same for all particles of the same absolute value of charge (fig. 3.5). The mass dependence of the momentum for a given $\beta\gamma$ ($p = \beta\gamma\, m$) and hence also the mass dependence of the average energy loss for a given momentum can be used for particle identification (see also eq. (3.34) on page 34).

For the energy loss measurement one can use detectors in which the measured signal is proportional to the amount of energy lost in the detector medium: for example light generation in scintillators, the blackening of photographic plates in emulsions (fig. 14.8), or charge generation and separation in gaseous or semiconductor detectors. A measurement of the average energy loss $\langle dE/dx \rangle$ or of the most probable energy loss (MPV, see section 3.2.3) is used to derive the velocity β and—if the momentum is known—to determine the masses of the long lived particles (e, μ, π, K, p, d, α):

$$m^2_{dE/dx} = p^2(1/\beta^2_{dE/dx} - 1)\,, \tag{14.9}$$

where $\beta_{dE/dx}$ denotes the velocity extracted from $\langle dE/dx \rangle$ employing the Bethe–Bloch formula (3.25).

The energy loss ΔE in a medium of thickness Δx approximately follows a Landau distribution (fig. 14.9, see also section 3.2.3), an asymmetric distribution with long

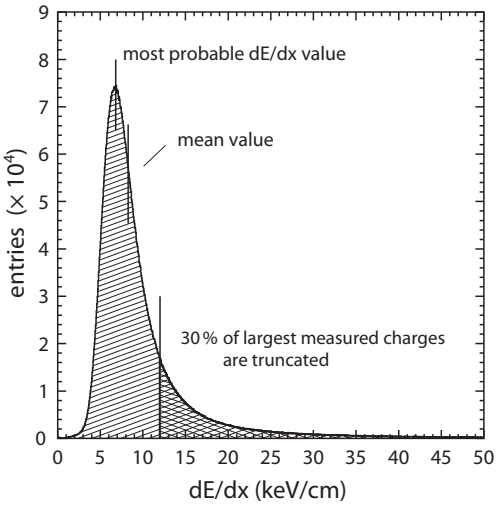


Fig. 14.9 Distribution of measured dE/dx values of minimum-ionising pions ($p = 400$–$800\,\mathrm{MeV}/c$) [499]. Marked in addition to the most probable dE/dx value at the maximum of the distribution (at $6.8\,\mathrm{keV/cm}$) is the average value as well as the value above which 30% of the largest energy loss entries are located. Truncating a certain percentage of the highest values (often 30%) and calculating the average only using the remaining entries (*truncated mean*) yields a mean measurement that is statistically more stable and is situated near the most probable value.

tails towards higher energy loss values. In building the average $\langle dE/dx \rangle$ from a finite number of measurements the spontaneously occurring large ΔE values lead to larger than Gaussian fluctuations and uncertainties.

To measure the average $\langle dE/dx \rangle$ of the distribution with good resolution, one must, besides having a sufficiently large number of statistically independent dE/dx measurements, also minimise the influence of the tail of the Landau distribution towards large energy loss values. This can be achieved for example by adopting the *truncated mean* method described in fig. 14.9. For individual particle tracks in high energy physics experiments multiple measurements of the energy loss per path length interval can be used for this purpose, for instance by charge or pulse height measurements at several wires along a track in a drift chamber, as explained in fig. 14.10. The pulse heights measured at 50 equidistant wires are individual dE/dx measurements (fig. 14.10(a)) which are Landau distributed (fig. 14.10(b)) and can be used to apply the truncated mean method. The quantity

$$\left\langle \frac{dE}{dx} \right\rangle_{\alpha} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{dE}{dx} \right)_i \tag{14.10}$$

with $\frac{dE}{dx}_{(i)} \leq \frac{dE}{dx}_{(i+1)}$ for $i = 1, \ldots N$ and $N = \alpha N_{tot}$, $\alpha \in [0.5, 0.85]$, follows an almost perfect Gaussian distribution [665] with width $\sigma_{dE/dx}$ being the dE/dx resolution.

The achievable dE/dx resolution primarily depends on the statistical fluctuation of the mean value of the deposited charge carriers, $\sigma_{dE/dx} \propto 1/\sqrt{N_e}$, and on the number of independent measurements N_{meas}. For gases the former leads to a dependence on the gas pressure, $\sigma_{dE/dx} \propto 1/\sqrt{P}$. An empirical formula obtained from fits to measured data yields [967, 1012]:

$$\frac{\sigma_{\mathrm{dE/dx}}}{\left\langle \frac{dE}{dx} \right\rangle} = 0.41\, N_{meas}^{-0.43} \left(\frac{\Delta x}{\mathrm{m}} \frac{P}{\mathrm{bar}} \right)^{-0.32} . \tag{14.11}$$

The deviation from a strict $N_{meas}^{-1/2}$ behaviour shows that measurements at different points along a track are not completely independent of each other. The formula also

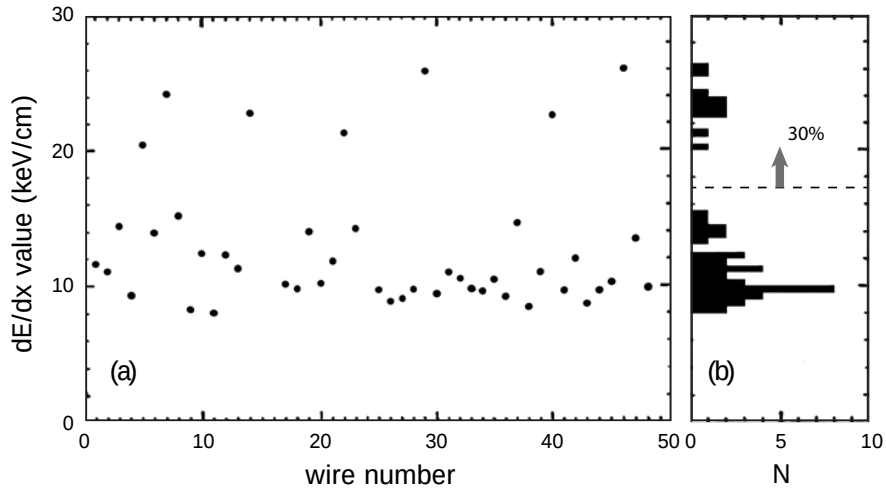


Fig. 14.10 Measurement of the specific ionisation caused by a particle obtained from 50 individual measurements (50 wires of a chamber) [963]: (a) measurements for every wire; (b) projection of the measurements as a histogram with a mark of the 30% cut line.

shows that for a fixed detector length $L = \Delta x\, N_{meas}$ one achieves a better resolution the larger the number of measurements[1] (rather than increasing Δx) and the higher the gas pressure. For a dE/dx measurement at high momenta one therefore must optimise the gas pressure (respecting technical limits) against the decrease in signal (energy loss) which is due to the density effect in the region of the relativistic rise (see also section 3.2.1). Typical resolutions for gaseous detectors are in the range between 3% and 12% [508].

The separation power for two particles 1 and 2 can as usual be given in units of the resolution $\sigma_{dE/dx}$:

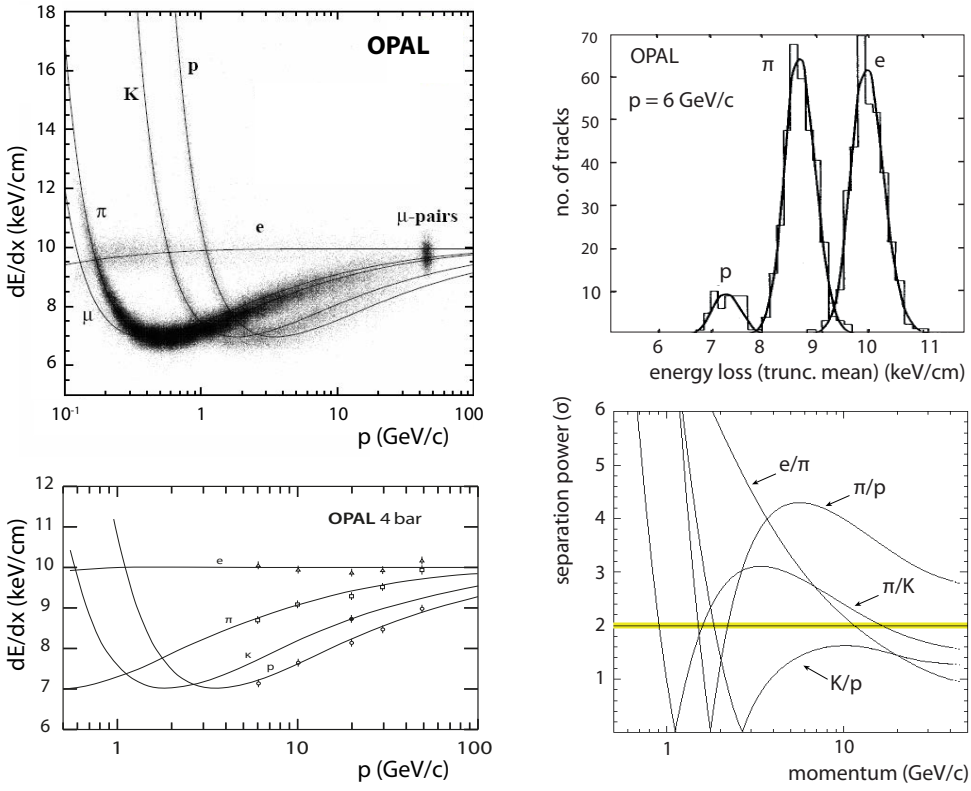
$$n_\sigma(dE/dx) = \frac{m_2^2 - m_1^2}{\sigma_{m^2}} \approx \frac{|\langle \frac{dE}{dx}\rangle(1) - \langle \frac{dE}{dx}\rangle(2)|}{\sigma_{dE/dx}}\,, \tag{14.12}$$

where the dE/dx mean values are given by the Bethe–Bloch formula (3.33) and $\sigma_{dE/dx}$ denotes the dE/dx resolution at a given β. In practice, instead of the mean value $\langle dE/dx\rangle$ a dE/dx estimator as in (14.10) is used.

Because of the characteristic form of the Bethe–Bloch curve (fig. 3.5) good separation (especially for π/K) is usually obtained at low momenta in the so-called $1/\beta^2$ region or at high momenta in the region of the 'relativistic rise' (see section 3.2.1.2 on page 33) whereas in the region of medium momenta separation power is reduced.

Figure 14.11 shows dE/dx measurements taken with the 'jet chamber' of the OPAL detector [63] as a function of momentum of particle tracks measured with the same chamber [509]. The OPAL jet chamber (see section 7.10 and table 7.6) was operated at a pressure of 4 bar. Up to 159 dE/dx measurements are taken at certain points (wires) along a track. The solid lines in fig. 14.11 (a) and (b) show the theoretically expected

[1] The number of possible measurements is limited by the minimum detectable signal.



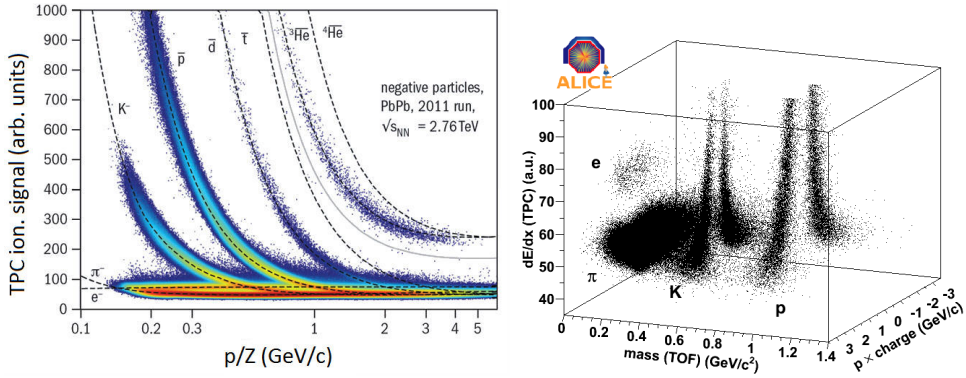
(a) dE/dx measurement. (b) Separation power.

Fig. 14.11 Ionisation measurements obtained with the gas-filled central jet chamber of the OPAL detector at high pressure (4 bar) [63]. (a) Individual measurements of $\langle dE/dx \rangle$ (points) of charged particle tracks as a function of momentum (top). The theoretical expectation (Bethe–Bloch formula) is given by the solid lines and is again shown separately below for better illustration together with data points at specific momenta (from [251] with kind permission of Elsevier). (b) Separation power of ionisation measurement (see also [251, 509]) shown (top) as dE/dx histogram for protons, pions, and electrons at 6 GeV/c momentum and (bottom) as separation distance curves in units of the resolution σ_{dEdx}. The shaded horizontal line indicates 2σ-separation.

dE/dx mean values; the points are individual dE/dx measurements. Their density is given by the abundance of a particular particle species in the produced collisions.

The curves in fig. 14.11(b) illustrate the separation power to distinguish e/π, π/K, π/p and K/p, in the top figure as a histogram of measured signal heights and in the bottom figure as differences of the respective mean values shown in units of the resolution σ. Good separation between different particle species is possible over a large momentum range [509]. The separation power (14.12) is momentum dependent. It is well above 2σ (yellow line in fig 14.11(b), bottom) at low and high momenta away from the minimum of the Bethe–Bloch, curve as explained above. Separation of π, K, and p for momenta between 1 and 3 GeV/c is hardly possible.

At the LHC experiments the dE/dx technique is employed for particle identification in the ALICE experiment at low momenta ($\lesssim 2$ GeV/c) and also at momenta in the



(a) dE/dx measurement. (b) dE/dx and TOF measurement.

Fig. 14.12 Particle identification with the ALICE TPC [85]. (a) Measurement of the ionisation signal as a function of momentum [26] (source: CERN 2013, ALICE Collaboration). (b) Measured dE/dx versus reconstructed particle mass (from TOF measurement) and momentum × charge in 2-dimensional representation, demonstrating that also electrons can be separated from π, K and p (from [665] with kind permission of Elsevier).

range of the relativistic rise. The central *time projection chamber* (TPC [85], see also section 7.10.10) delivers on electrode pad rows up to 159 ionisation measurements per particle track on small readout electrodes (typical 4×7.5 mm^2), thus achieving a total dE/dx resolution of about 5% [85, 665].

Figure 14.12(a) demonstrates the separation power of the TPC measurement for GeV/c momenta in Pb–Pb collisions, where deuterium, tritium and helium nuclei are also identified. Together with the TOF measurement, electrons can also be separated from pions [665], as shown in fig. 14.12(b) in a two-dimensional representation.

At low momenta semiconductor detectors can also provide good dE/dx particle separation. This is demonstrated in fig. 14.13 for the Si pixel detector of the ATLAS experiment [8] at momenta $\lesssim 1$ GeV/c. The pixel detector is described in section 8.7. The individual dE/dx measurements are obtained from three detector layers with 250 µm thick silicon pixel sensors. For solids the relativistic plateau for the (restricted) energy loss (fig. 3.13 on page 44) is only minimally ($\lesssim 10\%$) higher than the minimum of the curve due to the density effect. Therefore particle separation using semiconductor detectors is only possible in the $1/\beta^2$ region of the Bethe–Bloch curve (3.25), that is, at low momenta.

14.2.3 Particle identification with Cherenkov detectors

Cherenkov detectors as well as the basic principles of Cherenkov radiation are described in detail in chapter 11. Cherenkov radiation is generated when particles in a medium are faster than the velocity of light in that medium. From every point along the particle's track in the medium Cherenkov light is emitted on a cone with opening angle θ_c according to the Cherenkov relation (11.1),

$$\cos\theta_c = \frac{1}{\beta\, n}\,. \tag{14.13}$$

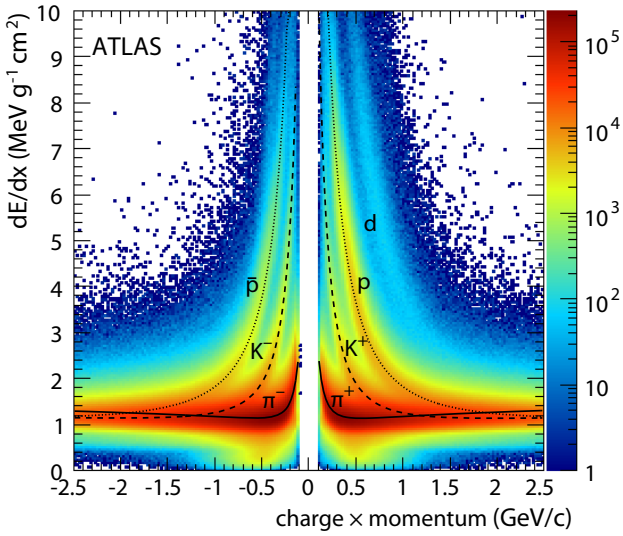


Fig. 14.13 Identification of charged particles at low momenta in the pixel detector of the ATLAS experiment by measurement of the specific ionisation dE/dx (adapted from [8] with kind permission of Elsevier).

This allows a measurement of the particle velocity β. Depending on whether Cherenkov radiation is detected in an imaging way as Cherenkov rings or if only the sheer emission of Cherenkov intensity is used for detection, we distinguish imaging RICH- or DIRC detectors from threshold Cherenkov detectors (consult chapter 11 where examples for the use of Cherenkov detectors in experiments are also given). A particular feature of Cherenkov radiation is its prompt and directed emission. This is exploited in the design of detectors for particle identification, for example by using coincident track information in the reconstruction of Cherenkov rings.

In modern accelerator experiments most often imaging detectors (RICH) are used. The mass of a particle is determined from eq. (11.27) in chapter 11:

$$m^2_{\text{cherenkov}} = \frac{p^2}{(\beta\gamma)^2} = p^2 \left(n^2 \cos^2\theta_c - 1\right) . \tag{14.14}$$

Cherenkov radiation generated in a radiator is optically focused on a photon detector where it is detected as a Cherenkov ring. The achievable mass resolution has been derived in chapter 11, equation (11.31):

$$\frac{\sigma_m}{m} = \frac{1}{2}\frac{\sigma_{m^2}}{m^2} = \gamma^2 \frac{\sigma_\beta}{\beta} = \left(\frac{p}{m}\right)^2 \frac{1}{\beta^2} \left(k_R \oplus \frac{1}{n}\frac{dn}{d\lambda}\Delta\lambda\right) , \tag{14.15}$$

where k_R is the RICH constant from (11.32) on page 461, $dn/d\lambda$ the dispersion of the refractive index and $\Delta\lambda$ is the accepted wavelength range. The $\oplus$ sign denotes quadratic addition. The total resolution is mainly composed of the resolution with which the opening angle of the Cherenkov rings can be measured by the imaging system and the photon detector:

$$\sigma_{\theta_c} = \frac{\sigma_{\theta_i}}{\sqrt{N}} \qquad \text{or} \qquad \sigma_{\sin^2\theta_c} = \frac{\sigma_{\sin^2\theta_i}}{\sqrt{N}} \tag{14.16}$$

as well as chromatic ring dispersion as the refractiver index n in (14.13) is wavelength dependent. Other contributions may enter as well (see below). In (14.16) σ_{θ_c} is the an-

gular error for N detected photons, respectively photoelectrons, and σ_{θ_i} is the angular resolution for one individual Cherenkov photon.

The achievable separation power between two particles with charge $z = 1$ and with $m_2 > m_1$, expressed in resolution units, which here contain all contributions to the angular uncertainty including chromatic uncertainties, can be obtained evaluating (11.33) with (11.31) and (11.32) in chapter 11:

$$\begin{aligned} n_\sigma(\text{Cherenkov}) &= \frac{m_2^2 - m_1^2}{\sigma_{m^2}} \\ &= \frac{\sin^2\theta_c(1) - \sin^2\theta_c(2)}{\sigma_{\sin^2\theta_c}} = \frac{\sin^2\theta_c(1) - \sin^2\theta_c(2)}{\sigma_{\sin^2\theta_i}/\sqrt{N}} \\ &= \frac{m_2^2 - m_1^2}{2p^2\,\sigma_{\theta_i}}\,\frac{\beta}{n}\,\sqrt{N_0 L} \approx \frac{m_2^2 - m_1^2}{2p^2\,\sigma_{\theta_c}}\,\frac{1}{\sqrt{n^2-1}}\,. \end{aligned} \tag{14.17}$$

From the first to the second line one arrives using

$$\sin^2\theta_c = 1 - \frac{1}{\beta^2 n^2} = 1 - \frac{p^2 + m^2}{p^2\,n^2} = 1 - \frac{1}{n^2} - \frac{m^2}{p^2\,n^2} \tag{14.18}$$

and by replacing $\sigma_{m^2} = p^2 n^2 \sigma_{\sin^2\theta_c}$. In the last line $N = N_0 L \sin^2\theta_c$ of (11.17) has been used, which is then approximated in the last expression by assuming $\beta \to 1$ and $n \sin\theta_c = \sqrt{n^2 - 1}$ after (11.5). In (14.17) the error on the Cherenkov angle σ_{θ_c} is the average of the individual photon errors σ_{θ_i} following (14.16) and includes all possible measurement contributions. These are, in addition to the position measurement uncertainty and the chromatic dispersion discussed in section 11.6.2, also other contributions like geometry error and detector alignment, multiple scattering of the particle track, noise and background hits in the photon detector, reconstruction error in track and ring reconstruction [665]. The single photon resolutions σ_{θ_i} contain contributions from the photon detector's spatial resolution σ_{θ_r}, usually determined by the electrode granularity (pixels), from the smearing of the emission point of the radiation $\sigma_{\theta_{em}}$, in particular for proximity focusing (fig. 11.18 on page 462), and finally from chromatic aberrations σ_{θ_n} (see section 11.6.2 on page 463):

$$\sigma_{\theta_i} = \sqrt{\sigma_{\theta_r}^2 + \sigma_{\theta_{em}}^2 + \sigma_{\theta_n}^2}\,. \tag{14.19}$$

As examples for particle identification and separation by Cherenkov radiation in modern large-scale particle and heavy ion physics experiments we discuss the imaging Cherenkov detectors of the LHC experiments ALICE and LHCb in which two modern techniques for the detection of Cherenkov photons are employed: wire chambers with CsI-coated photosensitive cathodes (ALICE HMPID, see also section 11.6, page 458) and (hybrid) photomultipliers with bialkali photocathode coating (LHCb RICH, see also section 10.4, page 427).

Particle identification using RICH detectors in the ALICE experiment. The RICH detector HMPID (*High Momentum Particle IDentification*) in ALICE provides K/π (resp. K/p) separation power for momenta $< 3\,\text{GeV}/c$ (resp. $< 5\,\text{GeV}/c$) in a limited solid angle range, with separation power beyond that of the dE/dx technique employed in the TPC. Light nuclei can also be identified. The detector consists of seven about $1.5 \times 1.5\,\text{m}^2$ large RICH modules that project the Cherenkov rings cre-

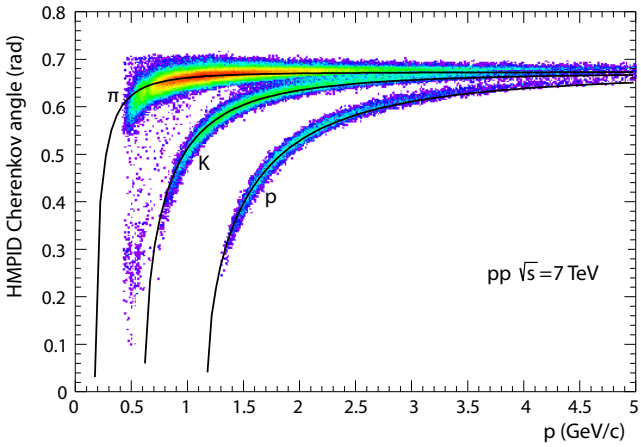


Fig. 14.14 Separation of pions, kaons, and protons with the RICH detector HMPID of the ALICE experiment as a function of momentum [27] (source: CERN).

Table 14.1 Characteristics of the two RICH detectors of the LHCb experiment at LHC [51]: refractive index, asymptotic Cherenkov angle, angular resolution, number of measured photoelectrons, momentum range for particle identification [656].

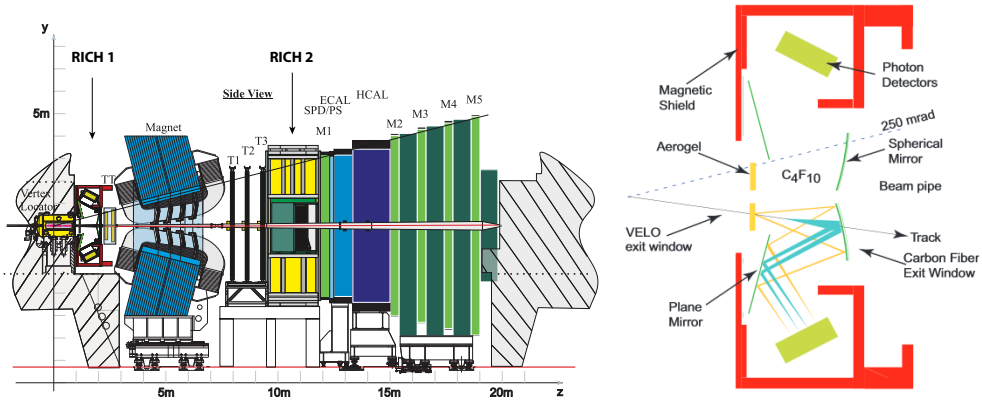
Radiator	n	θ_∞	σ_θ	$N^{\text{pe}}_{\text{meas}}$	Momentum range
aerogel, RICH-1	1.03	13.9°	±0.3°	≈5	some GeV
C_4F_{10}, RICH-1	1.0015	3.03°	±0.1°	≈25	10 GeV to 65 GeV
CF_4, RICH-2	1.00053	1.81°	±0.04°	≈20	15 GeV to >100 GeV

ated in a liquid radiator (C_6F_{14}, n=1.283) via proximity focusing onto in a gas-filled multiwire chamber with CsI cathode coating. The technique is described in chapter 11 on page 458.

The achieved resolution for single photons is $\sigma_{\theta_i} \approx 12\,\text{mrad}$, yielding a Cherenkov angle resolution of $\sigma_{\theta_c} \approx 3.5\,\text{mrad}$ for perpendicularly incident tracks with $\beta \approx 1$ [665]. Figure 14.14 demonstrates the separation power for π/K/p obtained solely by the RICH detectors.

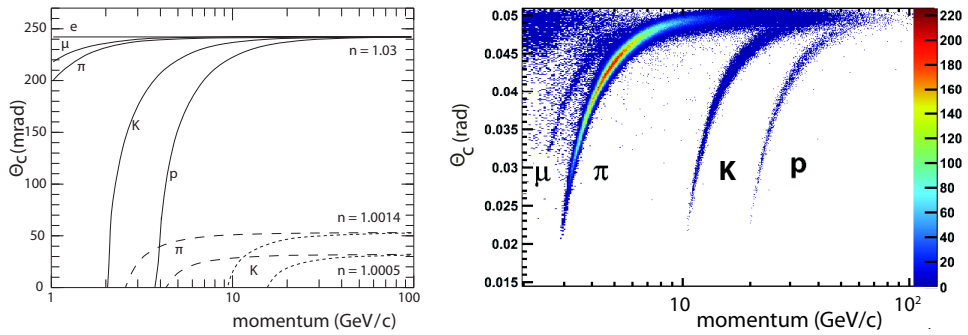
Particle identification using RICH detectors in the LHCb experiment. At LHC centre-of-mass energies the momentum spectrum of baryons and mesons comprises a large range reaching up to about 100 GeV/c and above. Identification of these hadrons is particularly important for the LHCb experiment, whose goals are detailed studies of B-meson decays for which in particular π/K separation at high momenta is crucial. For this reason the LHCb detector [87] features two gas RICH detectors for particle separation in two different momentum ranges [51]. In addition an aerogel radiator is included in one of the detectors (RICH-1) to cover low momenta.

A cross section of the LHCb detector is shown in fig. 14.15(a). RICH-1 for momenta between 2 and 40 GeV/c has a perfluorobutane radiator (C_4F_{10}, $n = 1.0015$) and an additional aerogel radiator with $n = 1.03$ (fig. 14.15(b)); RICH-2 for momenta between 15 and 100 GeV/c is filled with tetrafluoromethane (CF_4, $n = 1.00053$). These gases feature little chromatic dispersion ($\sigma_n/(n-1) \approx 2$–3%, see also section 11.6 on page 463. The Cherenkov light is read out in both systems by hybrid photomultipliers (HPD, see section 10.4), based on pixelated ($2.5\,times\,2.5$ mm^2) photodiodes. The LHCb RICH detectors are summarised in table 14.1.



(a) LHCb experiment. (b) RICH-1.

Fig. 14.15 (a) The LHCb experiment at LHC with two RICH Cherenkov systems for particle identification. (b) RICH-1 has two radiators, aerogel and C_4F_{10}, and is read out by hybrid photodetectors (HPD) (source: CERN/LHCb Collaboration).



(a) θ_c of different radiators. (b) Achieved separation with RICH-1.

Fig. 14.16 Particle identification with the RICH detectors in LHCb [656]. Shown is the Cherenkov angle as a function of particle momentum. In (a) calculated angles for pions and kaons are shown for three different indices of refraction; for $n = 1.03$ (aerogel) the angles for e, μ and p are also drawn. In (b) the separation achieved with RICH-1 only is demonstrated [51].

The Cherenkov angle for different refractive indices used for particle separation in LHCb is displayed in fig. 14.16(a) as a function of momentum. The achievable particle identification with the LHCb RICH system is shown in fig. 14.16(b) (RICH-1 only). Separation of p, K, π, and even μ (at small momenta) is possible over a large momentum range by simultaneous measurements of the Cherenkov angle (table 14.1) and of the particle momentum in LHCb's magnetic spectrometer.

14.2.4 Electron identification with transition radiation detectors

Relativistic particles traversing the boundary between two media with different electric permittivities ϵ_1, ϵ_2, radiate photons with energies up to the high energy X-ray regime. The fundamentals of transition radiation (TR) and its use for particle detectors are

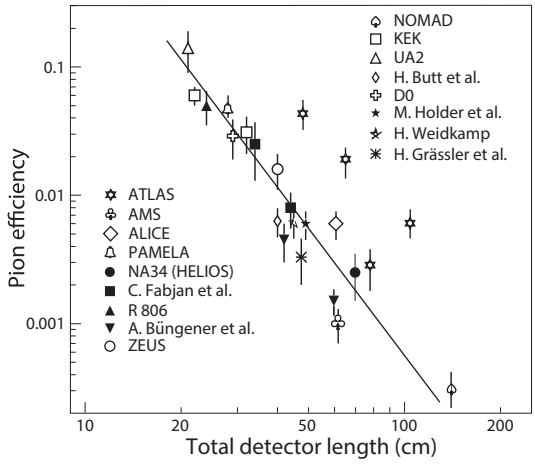


Fig. 14.17 Quality of electron-pion separation power with TR detectors: pion detection probability for a detection efficiency of 90% for electrons as a function of detector length for different experiments [746].

discussed in chapter 12. The TR intensity increases with the Lorentz factor γ of the particle up to a saturation value γ_s (see section 12.2.3 on page 492). The radiation is emitted in the forward direction according to a distribution given by (12.7) on page 483 with the most probable emission angle being proportional to $1/\gamma$ (fig. 12.3):

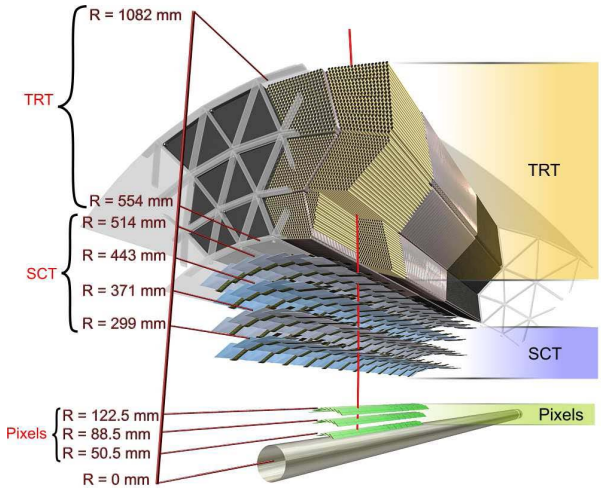
$$\text{intensity} \propto \gamma\,, \qquad \theta_{\text{mpv}} \propto \frac{1}{\gamma}\,. \tag{14.20}$$

Due to its γ dependence transition radiation is utilised in particle physics mainly to separate high energetic electron from other particles like pions, protons and kaons. Electrons have a γ factor almost 300 times larger than pions at the same momentum. The usage range of transition radiation can also be taken from fig. 14.22(b) on page 567.

The TR intensity reaches a maximum depending on the number and thicknesses of the media used in the radiator (see fig. 12.10 on page 493). This choice also selects a photon frequency range which depends on γ (fig. 12.8(a) on page 490). Furthermore, for a significant detection of transition radiation accompanying a charged particle track a minimum TR intensity is necessary which in addition is modulated by interference maxima and minima from the superposition of radiation from many boundaries. For this reason transition radiation features a threshold behaviour (in γ) for detectability (see chapter 12).

The transition radiation detectors in use are constructed from alternating media with different ϵ, usually either foil stacks or foam/fibre materials. While with foil stacks a constant thickness-to-distance ratio can be tuned, for foams and fibres only an average pore size, respectively an average fibre interspace, can be selected via the material's density. Examples are shown in fig. 12.11 on page 495. On average, increased TR is obtained by (partially) constructive interference and incoherent addition of radiation emitted from the many, very similar transition boundaries (see section 12.3). The more boundaries that exist along a track, the more photons are created up to saturation. In order to reduce self-absorption in subsequent radiator material, arrangements with several stages can be used, each having a photon detector. The separation power for e/π separation then depends on the total length of the detector (see fig. 14.22(b)). The experimental verification is shown in fig. 14.17.

The calculation of the TR separation power must account for the fact that TR



(a) ATLAS inner detector (barrel).

(b) Detail of the TRT detector (endcap).

Fig. 14.18 (a) Detail of the barrel region of the ATLAS inner detector [4]. The *transition radiation tracker* (TRT) covers the outer part. (b) Detail of an endcap module: Layers of drift tubes alternating with radiator layers composed of polypropylene fibre foils serve as tracking detector and transition radiation detector at the same time. Source: CERN/ATLAS Collaboration.

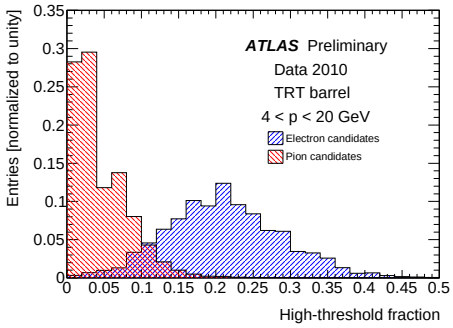
signals have a threshold behaviour and only intensity above threshold can be used for the separation of two particle species. Simplified, a similar definition as in (14.12) can be written as

$$n_\sigma(\mathrm{TR}) = \frac{|\langle \mathrm{N}_\gamma \rangle(\mathrm{e}) - \langle \mathrm{N}_\gamma \rangle(\pi)|}{\sqrt{\langle \mathrm{N}_\gamma \rangle(\mathrm{e})}}\,, \tag{14.21}$$

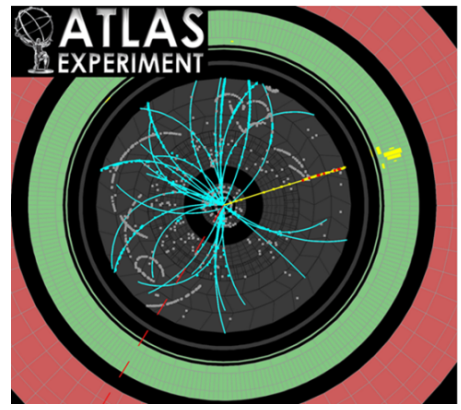
where $\langle N_\gamma \rangle$ is the detected number of TR photons for electrons (e) or pions (π), respectively. In typical applications and to first approximation only electrons generate X-ray photons, such that the separation power directly depends on the TR intensity for electrons. Larger detector length L, distributed over several stages read out individually, increases the number of photons and hence the separation power against pions (and even more so against heavier hadrons) at the same momentum. Saturation in the number of TR photons, starting at a γ value somewhat below 10^4 must be considered (see fig. 12.8(b) in section 12.2.2).

For LHC reactions the detection of high energetic charged leptons is very important. For electron identification beyond electromagnetic calorimetry, TR detectors offer a combination of electron identification and detection of the track of the electron candidate. In the following we therefore describe two detector concepts used at LHC: the TRT detector of ATLAS and the TRD detector of the ALICE experiment.

The transition radiation tracker of ATLAS. The so-called *transition radiation tracker* (TRT) [4] in ATLAS serves two tasks at the same time. It is part of the tracking detector in the outer part of the ATLAS inner detector (fig. 14.18) and at the same time provides the possibility to separate electrons from hadrons (above all from pions) by recognising TR signals in a momentum range between 1 GeV/c and 200 GeV/c.



(a) Hit distribution ATLAS TRT.



(b) Electron identification in the $W \to e\nu$ decay.

Fig. 14.19 (a) Distribution of hits above the higher TR threshold for tracks in the ATLAS TRT [125]; (b) event display of a collision (Run 152409, Event 5966801), in which the decay of a W boson, $W \to e\nu$, is observed. The decay electron can clearly be detected: electron in yellow with a shower in the electromagnetic calorimeter (green); the TR hits (high thresholds) are visible as red points along the track in the TRT; the (invisible) neutrino is shown as the red dashed line (source: CERN/ATLAS Collaboration).

The ATLAS TRT detector is described in section 12.3 on page 497. It consists of cylindrically (barrel) or radially (endcap) arranged drift tubes, between which the TR radiator is integrated as a fibre radiator (barrel) or as a foil stack (endcap), respectively. In the central region of the ATLAS TRT always at least 36 drift tubes are traversed by a particle track. The TR signals, being additional to 'normal' ionisation signals are electronically distinguished by defining two thresholds (high/low) as shown in fig. 12.14.

Figure 14.19(a) shows the distribution of the fraction of hits in the ATLAS TRT above threshold. For electrons high threshold signals occur much more frequently than for pions. At a 90% electron detection efficiency pions are suppressed by a factor of 20 to 50 [125]. An event display of an LHC collision is shown in fig. 14.19(b) in which a W boson is created decaying into an electron and a neutrino. The electron track has seven TR hits.

The transition radiation detector of the ALICE experiment. Different to the TR application in ATLAS, where electron identification at high momenta up to above 200 GeV/c is the goal, the heavy ion experiment ALICE aims for electron identification in the low momentum regime (from about 2 to 10 GeV/c) with very strong pion suppression (factor 100). Below 1 GeV/c electron identification in ALICE is achieved using dE/dx measurements in the central TPC [10, 751], as illustrated in fig. 14.12. The principle of a chamber unit is shown in fig. 14.20. The entire TRD is arranged in 18 modules consisting of 540 drift chambers total, filled with a photon absorbing gas mixture of 85% Xe and 15% CO_2. The drift chamber detects signal charges generated by TR photons in addition to charges created by ionisation of the charged particle track. The chamber reads out signals induced on the pads of the cathode plane, resolving them both in space and in time. The chamber is operated at moderate gas gain to avoid space charge developing in the gas volume by ions drifting back to the

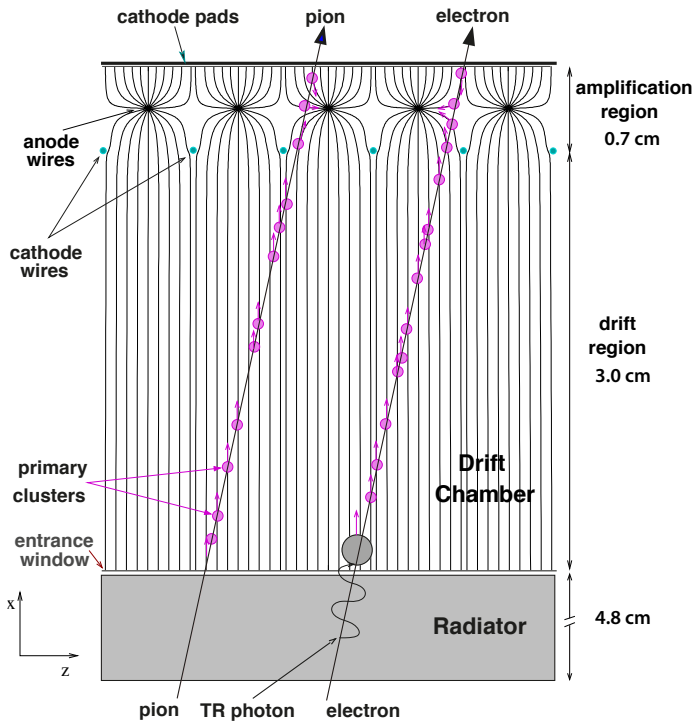
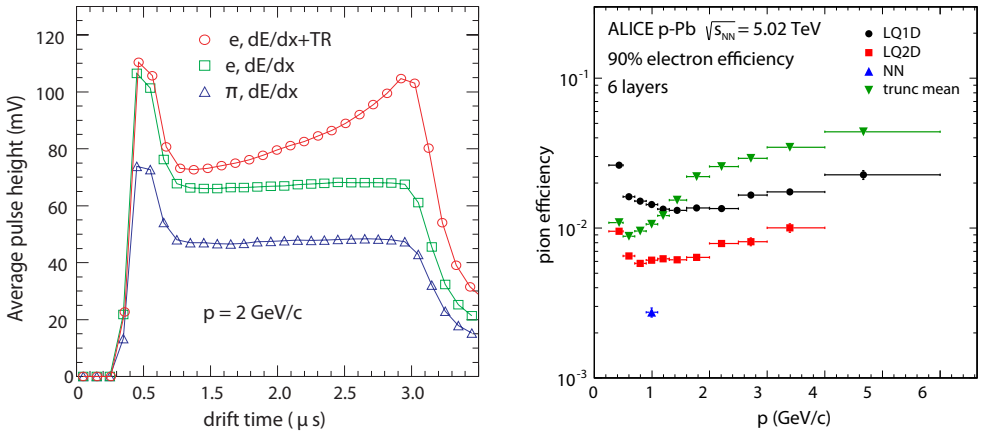


Fig. 14.20 Function principle of a module of the ALICE TRD detector, consisting of a TR radiator combined with a driftchamber (from [10], source: CERN/ALICE Collaboration).

cathode. Since the TR photon absorption mainly takes place at the side facing the radiator (bottom side in fig. 14.20), the transit time of TR charges is comparatively large, causing a peak in the time spectrum at large drift times. Electrons hence differ from pions at small momenta ($\sim$ GeV/c) by larger ionisation signals (GeV electrons are far in the relativistic rise, different to pions, see e.g. fig. 14.11(a)) and by additional TR contributions resulting in average pulse height spectra, as shown in fig. 14.21(a). In the GeV/c momentum range pion suppression factors of about 80–150 are achieved, depending on the chosen analysis technique [27] (fig. 14.21(b)).

14.2.5 Characteristic detector length

The particle identification methods described in the previous sections, which all rest on eq. (14.1), require larger characteristic detector lengths with increasing particle momentum, respectively velocity. The characteristic detector lengths (e.g. the distance between two timing detectors for a TOF measurement or the radiator length for Cherenkov measurements) required for a certain separation (e.g. 2σ or 3σ) are drawn in fig. 14.22 as a function of the particle momentum, estimated using typical application ranges and resolutions. Considered are π/K separation in fig. 14.22(a) and e/π separation in fig. 14.22(b). For the TOF measurement, for example, the functional dependence follows directly from (14.8). For an identification measurement using Cherenkov radiation the detector length is buried in the number of detected photons for the needed resolution, which increases proportional to the radiator length (eq. (14.17)). In the figure, lengths for different radiator types (liquid, aerogel, gas) for typical resolutions are shown, characterising the large range of particle identification by means of Cherenkov radiation. The constant chromatic contribution to the resolution (sec-



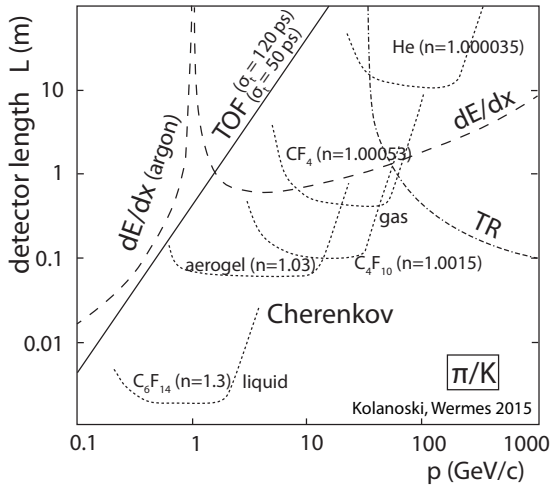
(a) Time spectra. (b) Pion suppression.

Fig. 14.21 (a) Time spectra (simulation, pulse height as a function of drift time) for typical tracks in the TRD chamber of the ALICE experiment [10]. Shown are ionisation signals (dE/dx) for pions and electrons individually as well as the sum of dE/dx and TR signals for electrons. Early times correspond to ionisations near the wire, late drift times to ionisations near the TR radiator. In the amplification region (early times), the signal is larger, because the ionisation from both sides of the anode wires contributes to the same time interval. (b) Inverse suppression factor (*pion efficiency*) for pions at a 90% detection efficiency for electrons as a function of momentum, determined employing different identification techniques indicated by the different symbols: *truncated mean* (green triangle), Likelihood methods (black and red), and neural net (blue) [27]. Source: CERN/ALICE Collaboration.

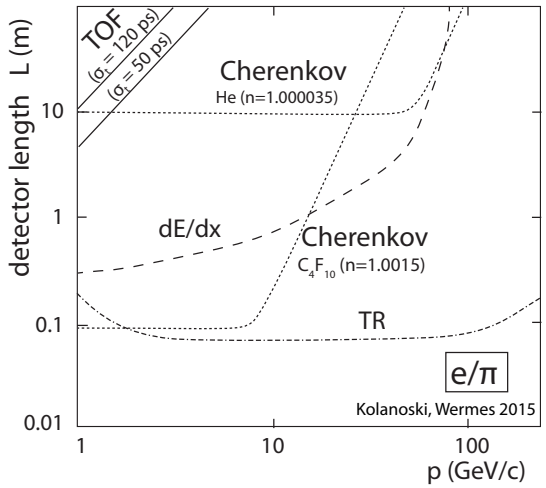
tion 11.6.2) is a lower limit for the generally achievable resolution leading to a lower plateau in the curves for Cherenkov radiation in fig. 14.22.

While for transition radiation the intensity of one interface is proportional to the Lorentz factor γ, the interference pattern resulting from two or more interfaces as well as absorption in the radiator introduce an effective γ threshold and a saturation around a certain value γ_s according to (12.30) in section 12.2.3. The detectable number of photons is given by (12.15). The separation power follows from (14.21). A sufficient number of photons is achieved by stacks of foils (here 100 foils), whose number is, however, limited by self-absorption in the stack (section 12.2.3). Using several stack-/detector stations the detector becomes longer and the separation power increases. Due to saturation beyond γ_s the separation power (fig. 14.22(b)) reaches a lower limit in length and gets worse again when competing particles (here pions) begin to emit transition radiation . For momenta typical for today's particle experiments TR is exclusively used for e/π separation. Only at very high momenta it can in principle also be used for π/K separation. In astroparticle physics experiments measurements of γ are also attempted at high energies (see section 16.2.1 on page 663).

For the dE/dx method a gas volume (argon, 1 bar) of a typical multiwire chamber is assumed. The ionisation signal is measured and the separation power is calculated using equation (14.12).



(a) π/K separation



(b) e/π separation

Fig. 14.22 Schematic views of the magnet spectrometer detector length for particle separation (3σ): (a) π/K and (b) e/π separation as a function of momentum for various PID methods described in the text and for typical application assumptions: time of flight measurement (TOF, solid lines) with time resolutions of $\sigma_t = 120$ ps and 50 ps; ionisation measurement (dE/dx, dashed lines), argon filled chamber at 1 bar with 3% charge resolution. For Cherenkov measurements (dotted lines) for different radiators the following angular resolutions have been assumed: C_6F_{14} ($n = 1.3$, $\sigma_{\theta_c} = 20$ mrad, $\sigma_{\theta_n} = 12$ mrad); aerogel ($n = 1.03$, $\sigma_{\theta_c} = 2$ mrad, $\sigma_{\theta_n} = 1$ mrad); and for the gas radiators C_4F_{10} ($n = 1.0015$, $\sigma_{\theta_c} = 0.5$ mrad), CF_4 ($n = 1.00053$, $\sigma_{\theta_c} = 5$ mrad), He ($n = 1.000035$, $\sigma_{\theta_c} = 0.2$ mrad), each with chromatic dispersion according to (11.46) with $\sigma_n/(n-1) = 3\%$. For transition radiation (dashed-dotted lines) TR stations with 100 foils and a TR photon detection energy threshold of 1 keV are assumed as well as a saturation in the number of TR photons starting at $\gamma \approx 8 \times 10^3$ [357].

14.3 Identification of muons

Electrons and muons play a very important role in high energy physics experiments because they provide a 'signature' for heavy quarks and leptons and/or exotic particles. Furthermore, their detector signals are well suitable for triggering an experiment (see chapter 18). Many neutral particles decay into lepton pairs. Prominent examples are decays of Z^0 bosons or quark–antiquark states like $\rho(770)$, J/ψ and Υ into e^+e^- or $\mu^+\mu^-$ pairs. $W^\pm$ bosons and heavy quarks decaying semileptonically emit leptons under comparatively large angles relative to the flight direction of the mother particle and are hence a signature for them.

Electrons—depending on their energy—can be identified for example by methods described in the previous section and in addition by detector systems containing a tracking system and an electromagnetic calorimeter (fig. 14.2 and section 15.5). For muons the techniques of section 14.2 are less suited to distinguish them from long-lived

hadrons, especially from pions because the mass difference is only $34\,\mathrm{MeV}/c^2$. However, muons do not feel the strong interaction and also do not develop electromagnetic showers as bremsstrahlung is much suppressed compared to electrons ($\propto 1/\mathrm{m}^2$). As a consequence they are able to pass thick and massive materials that absorb hadrons (hadron absorbers). Such absorbers are automatically present in experiments that contain a hadron calorimeter. Charged tracks that exit the hadron absorber very likely are muons. Hadrons can cause background by 'punching through' the absorber with some (small) probability.

Other possibilities to identify muons under certain experimental conditions are described in other sections of this book, for example: by employing the shape of Cherenkov rings in neutrino reactions on page 473 in section 11.6.3, by using energy loss measurements at high energies in section 3.3.4 as well as by exploiting their strong penetration capability of for example rock or water/ice to detect them in underground astroparticle detectors for the identification of this penetrating air shower component (see section 16.6.5.2 and fig. 16.32).

Having a lifetime of 2.2 µs GeV-muons fly many kilometres before decaying. In order to identify them in high energy physics experiments one exploits the fact that—up to momenta of several hundred GeV/c—their predominant interaction process is energy loss by ionisation. The critical energy E_c^μ defined in (3.92) on page 63, at which the ionisation energy loss equals the loss by radiation, in contrast to the definition for electrons also includes pair creation and photonuclear contributions in the radiation process. It is defined in eq. (3.96) in section 3.3.4 and is listed in table 3.4; it can be approximated [746] for solids by

$$E_c^\mu \approx \frac{5.7\,\mathrm{TeV}}{(Z+1.47)^{0.838}}\,. \tag{14.22}$$

The range of muons varies strongly with momentum (see fig. 3.18). For muons with $10\,\mathrm{GeV}/c$ momentum the range in iron is about 7 m, for $100\,\mathrm{MeV}/c$ muons the range is only 5.5 cm. Starting at a threshold momentum, which depends on the absorber thickness and typically lies in the region of $1\,\mathrm{GeV}/c$, muons are no longer absorbed, even in thick absorbers as for example in calorimeters. Their identification thus can be obtained by detecting a charged particle's track before and after a hadron absorber. The muon momentum is often determined from a combined measurement of the track curvature in front and behind the absorber after correcting for the energy loss in the absorber (see figs. 3.15 and 3.16). This method needs a minimum depth of the calorimeter of about 10 hadronic interaction lengths λ_a (about 16.8 cm in iron).

The probability that hadron remnants escape the calorimeter (hadron *punch-through*) and thus mimic a muon in the muon spectrometer, decreases exponentially with absorber thickness according to (3.144) on page 86. The hadron remnants can be debris of a hadron shower or secondary muons from pion or kaon decays. The latter is dominant for thick absorbers ($\gtrsim 10\lambda_a$). To determine the background in a measurement the punch-through probability must be weighted with the a priori abundance of hadrons relative to muons. In typical LHC reactions hadrons are more abundant than muons by orders of magnitude (depending on final state and event selection). Nevertheless, the contribution of tracks falsely identified as muons due to punch-through can be suppressed, for example in CMS [300] to a level of less than 1% for muons with transverse momenta larger than $20\,\mathrm{GeV}/c$ and when strict criteria on the quality of the muon tracks are imposed.

In many particle physics experiments dedicated muon spectrometers are installed as independent detector units, for example in the LHC experiments ATLAS [4] and CMS [298] (see also [513] and chapter 19 in [488]).

As described in section 9.1 in eqs. (9.5) and (9.8), the trajectory of a singly charged particle in a homogeneous magnetic field $\vec{B}$ projected onto a plane perpendicular to the flight direction is a circle with radius

$$R = \frac{p_T}{eB} = \frac{1}{0.3}\,\frac{p_T}{B}\,, \tag{14.23}$$

where p_T in GeV/c is the component of the muon momentum perpendicular to $\vec{B}$, R has to be given in metres and B in tesla. The direction of the curvature determines the sign of the charge. The method of measurement and the obtainable precision are described in chapter 9. Because large volumes are often to be covered by detectors for muon tracking, usually large area, gas-filled chambers (chapter 7) are employed for position sensitive detection of muon tracks.

The magnet spectrometers of the LHC experiments use solenoid magnets that generate a B-field parallel to the accelerator beam. Muons from beam collisions are bent on a circle in a plane perpendicular to the beam ('r–ϕ' plane) (fig. 14.23). In addition ATLAS has toroid magnets in the exterior part of the detector (section 9.2.3). In the barrel region the toroid has eight large coils (fig. 9.4(b)) creating a circular (toroidal) magnetic field around the inner part of the ATLAS detector. Muons, having penetrated the calorimeter material, are bent a second time, now in a plane containing the LHC beam ('r–z' plane, fig. 14.23(b)).

The muon chambers of CMS are placed in the return yoke of the CMS magnet, the field of which causes track bending in the opposite direction to the bending inside the coil such that altogether the muons follow an S-shaped trajectory (fig. 14.23(a)). In ATLAS the return of the solenoid field takes place inside the hadron calorimeter. With the muon spectrometers (together with the inner tracking detectors) the following momentum resolutions are obtained [352, 299]:

$$\begin{aligned} \text{ATLAS:} \quad & \frac{\sigma_{p_T}}{p_T} \approx 3.5\%\ (10\%) \quad && \text{at 50 GeV (1 TeV)}\,, \\ \text{CMS:} \quad & \frac{\sigma_{p_T}}{p_T} \approx 2.3\%\ (10\%) \quad && \text{at 50 GeV (1 TeV)}\,. \end{aligned}$$

14.4 Separation of electrons, muons and hadrons in calorimeters

Particles can also be distinguished by exploiting the characteristic shapes of showers generated by different particles in calorimeters.

The extension of electromagnetic showers (section 15.2), in longitudinal and in transverse direction, is described by the radiation length X_0 (see eq. (3.88) on page 61 and for the transverse dimension eq. (15.21) on page 590). The extension of hadron showers (section 15.3) is given by the nuclear absorption length λ_a defined in (3.145) on page 86. Since for all absorber materials with $Z \gtrsim 6$ the absorption length λ_a is significantly larger than the radiation length X_0 (see table 15.3 on page 599), the extension of hadronic showers is also larger than that of electromagnetic showers, in depth and in width. Furthermore, an electromagnetic shower usually starts less deep in an absorber than a shower triggered by a hadron. Examples for both types of showers

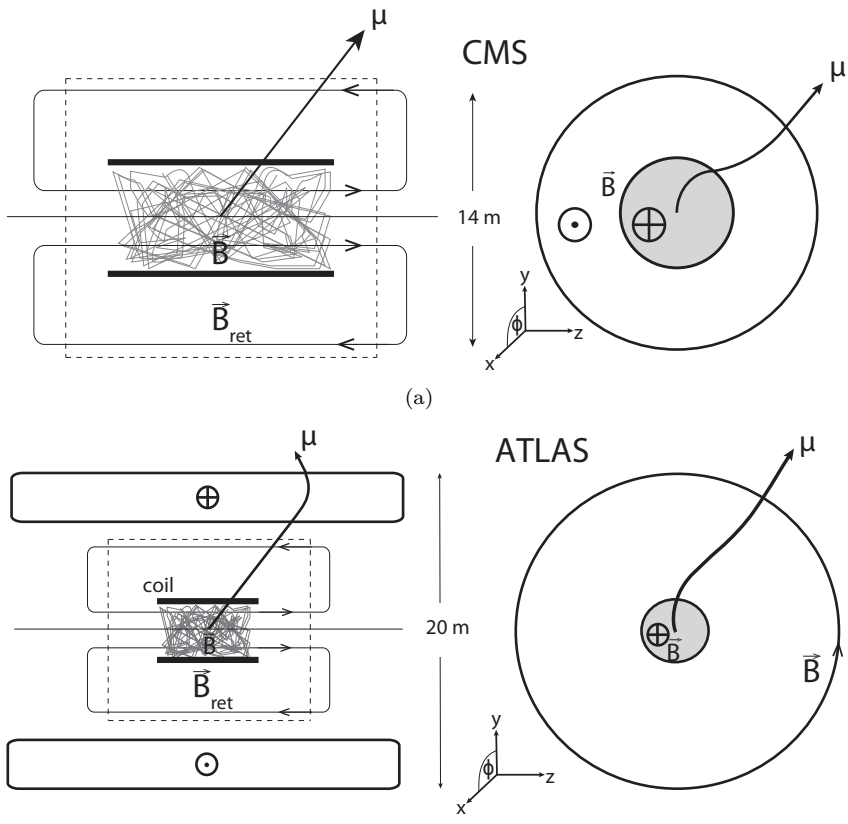


Fig. 14.23 Schematic views of the magnet spectrometers of the LHC experiments CMS (a) and ATLAS (b). Shown is the bending of a muon track in a plane perpendicular to the direction of the LHC beam (so-called 'r–ϕ' plane, both on the right) and in a plane orthogonal to this containing the LHC beam (so-called 'r–z' plane, both on the left). The solid lines with arrows (left) show the orientation of the magnetic field. The dashed drawn rectangle indicates the region of the magnetic flux return yoke (adapted from [513]).

are given in fig. 15.11 on page 596. Figure 14.24 shows the average longitudinal profile of electromagnetic and hadronic showers in iron in comparison [489].

To distinguish hadrons and electrons in calorimeters, besides the shower starting point and the ratio of deposited energy in front and rear parts of the calorimeter, also the substantially larger width of hadronic showers can be used. For example, an electromagnetic shower in lead is 95% contained in a cylinder of 3.2 cm, whereas a hadronic shower has a 95%-extension of 17.6 cm (see sections 15.2.2 and 15.3.3).

Usually electrons and hadrons or jets are separated from each other by fitting measured shower profiles to energy dependent shower shape templates. Figure 14.25 shows for ATLAS that suppression factors around 1000 at an electron efficiency of 90% can be obtained already by using the calorimeter only. By additionally employing tracking detector information plus the TRT detector, suppression factors above 10^5 at an electron efficiency of 70% can be obtained [385], a value that is for example needed to select $W \to e\nu$ decays sufficiently background free despite the large jet-to-electron

abundance ratio [7].

At energies accessible to accelerator experiments muons in general do not initiate showers, neither electromagnetic nor hadronic, but lose energy predominantly by ionisation (dE/dx). Figure 14.26 exemplifies the much smaller energy deposition by muons compared to particles that interact hadronically like pions or electromagnetically showering electrons. For isolated particles this is an important characteristic to distinguish muons from electromagnetically or hadronically showering particles. More difficult is the identification of muons inside jets. A separation from hadrons requires a combination of measurement information from the calorimeters and from the tracking detectors, as explained in section 14.3.

14.5 Detection of neutrons

The detection of neutrons has gained much in importance in recent decades. Besides increased requirements in reactor security and radiation protection (see for instance [911]), scientific projects have also developed an increased interest in neutron detection. An example is the worldwide network of neutron monitors for observation of the cosmic neutron flux to determine the primary energy spectrum of cosmic rays and to aid

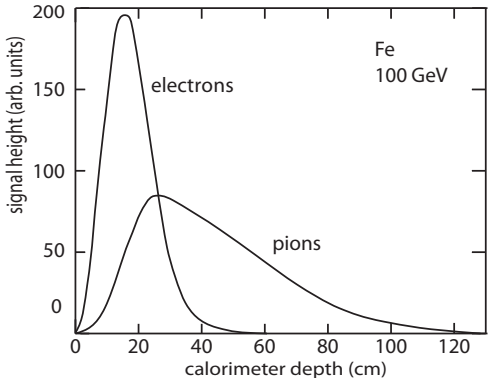


Fig. 14.24 Longitudinal shower development of electromagnetic showers initiated by electrons in comparison to pions (adapted from [489]).

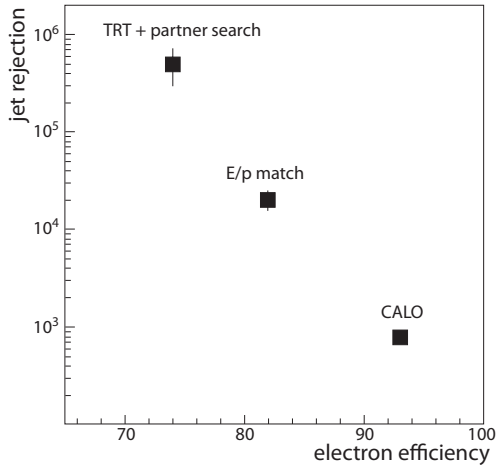


Fig. 14.25 Separation of electron and hadron jets in ATLAS (simulation): Jet suppression factors as a function of the detection efficiency for electrons employing (i) shower profiles in the calorimeter (CALO), (ii) the additional condition that a track with momentum p measured by the inner tracker aligns with the shower of energy E with $p \approx E$ (E/p match), and (iii) with additional usage of the transition radiation detector TRT to recognise electrons with positron partners as photon conversions (TRT + partner search). Adapted from [385], with kind permission of the American Physical Society.

weather forecast [739]. Detailed reviews on neutron detection can be found for example in [812] and [616].

Since they are neutral particles, neutrons do not ionise atoms in matter. Collisions with atomic nuclei, however, can generate charged reaction products which in turn are detectable by ionisation in detectors. Apart from elastic collisions, many neutron-induced reactions with nuclei proceed in a way in which the neutron is first bound in the nucleus, then causing the nucleus to decay or undergo fission. Often an excited nucleus is created which de-excites into the ground state by emission of a gamma quantum ((n,γ) reaction). The cross sections for reactions relevant for neutron detection strongly depend on the (kinetic) neutron energy T_n. For low energies the cross sections approximately follow $\sigma \propto 1/T_n$. For neutron detection the following reactions are important, classified according to the neutron energy (see also fig. 14.27):

– At low neutron energies of $T_n < 20\,\text{MeV}$ besides (n,γ) capture we mainly have fission reactions:
  (a) $\text{n} + {}^{6}\text{Li} \rightarrow \alpha + {}^{3}\text{H}$,
  (b) $\text{n} + {}^{10}\text{B} \rightarrow \alpha + {}^{7}\text{Li}$,
  (c) $\text{n} + {}^{3}\text{He} \rightarrow \text{p} + {}^{3}\text{H}$,
  (d) $\text{n} + {}^{235}\text{U} \rightarrow {}^{236}\text{U}^{*} \rightarrow$ fission products.

  This energy range ($T_n < 20\,\text{MeV}$) includes cold ($T_n < 25\,\text{meV}$) and ultra-cold neutrons ($T_n < 2\,\text{meV}$), thermal (25 meV – 1 eV) and slow (1–10 eV) neutron as well as medium (300 eV – 1 MeV) and fast (1–20 MeV) neutrons .
– For energies between $T_n > 20$ MeV and 1 GeV elastic (n,p) reactions are most abundant.
– High energy neutrons with energies $T_n > 1\,\text{GeV}$ cause hadronic showers.

Common to all detection methods is that the cross section of all suitable reactions strongly falls with increasing neutron energy, also for the elastic (n,p) reaction above 100 keV. At very low energies (<1 eV) the cross section reaches values of more than 1000 b. Moderation of neutrons thus enhances the detection efficiency. This can be obtained most effectively by elastic scattering with large energy transfers to collision partners with equal mass, that is, nucleons or light nuclei (compare eq. (15.71) on page 639).

Thermal and slow neutrons. For the detection of thermal or slow neutrons gas-filled as well as semiconductor detectors or scintillators can be employed. To efficiently detect neutrons ^{3}He (only for gases), lithium or boron should be contained in the

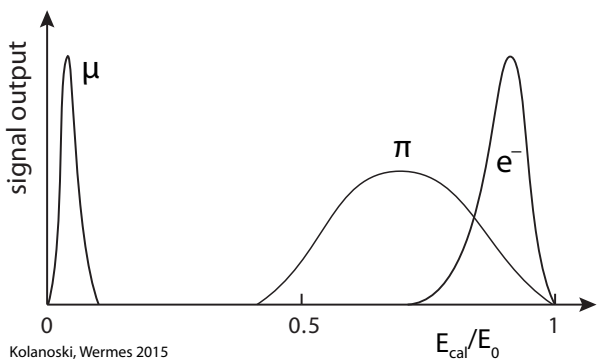


Fig. 14.26 Energy deposits from electrons, pions and muons in absorbing calorimeters (schematic). Due to different interactions being responsible the energy deposits are very different.

detector material to allow exploitation of reactions (a),(b) or (c) above. Chambers or tubes filled with BF_3 gas and operating as ionisation or proportional chambers (for gas detector operation modes see section 7.4.2 on page 184) can hence be used as neutron detectors. To increase the detection efficiency an outer shell enveloping the detector, composed for example of paraffine or water, is often added as moderator. This slows down the neutrons to lower energies with higher reaction cross sections.

Lithium-coated semiconductor detectors exploit the same principle: α particles or tritium generated in reaction (a) are detected. For scintillators the use of lithium iodide scintillator material (LiI(Eu)) is an option, where Eu acts as activator (see section 13.3) similar to Tl in NaI(Tl) [489]. Since thermal neutrons have a large fission cross sections in ^{235}U (>100 b), also fission reactions (d) are suited for neutron detection. Figure 14.28(a) shows a neutron counter composed of a proportional counter containing a ^{235}U-coated cathode. The fission products are detected by the counter. The device is surrounded by a paraffin moderator (C_nH_{2n+2}, $n = 18$–32) for neutron deceleration.

Often used as neutron counters are proportional or Geiger counters, embedded in a polyethylene sphere. To utilise reaction (c) special counter gas fillings, for example ^{3}He/CH_4 or pure ^{3}He, are used. A photograph of a neutron detector of this type (Bonner Ball Neutron Detector) is shown in fig. 14.28(b) [703]. The counter in the centre operates as a proportional tube under high pressure (about 6 bar of pure ^{3}He). It is surrounded by a massive, 9 cm thick polyethylene moderator shell preventing other radiation apart from neutrons (α, β, γ) from entering the tube. This type of neutron counter is a reliable device for neutron flux measurements in radiation areas.

Fast neutrons. So-called fast neutrons in the MeV energy region up to about 100 MeV can efficiently be detected by organic scintillators with a large hydrogen content, utilising the elastic (n,p) reaction (see the corresponding discussion in chapter 15 on calorimetry, especially section 15.6.3.1 and the cross section in ^{238}U in fig. 15.38 on page 642). The neutron counter should have dimensions which are large compared to the neutrons' mean free path λ_n such that they are detected with high probability. Even beyond 100 MeV up to the GeV region the elastic (n,p) reaction can be utilised. However, the detection efficiency decreases strongly with increasing neutron energy due to the energy dependence of the elastic cross section being roughly constant only

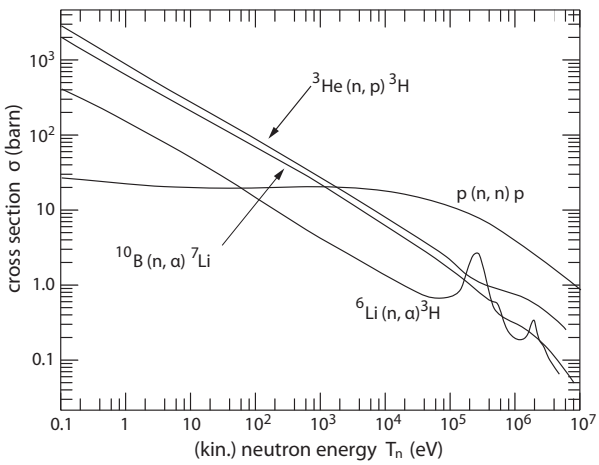
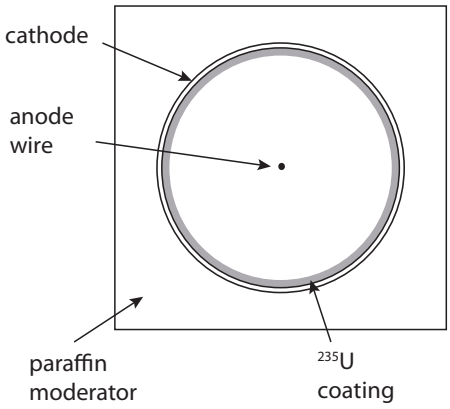
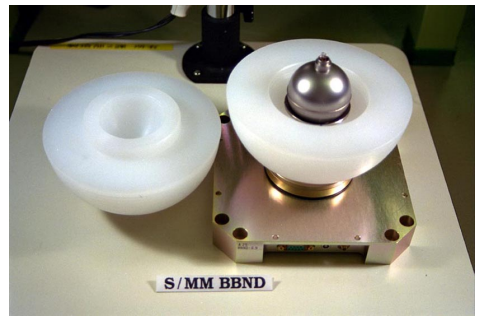


Fig. 14.27 Cross section for neutron-induced reactions as a function of kinetic neutron energy (adapted from [732]).



(a) Neutron counter (fission reactions).



(b) Neutron counter (^{3}He reactions).

Fig. 14.28 (a) Neutron counter based on a proportional counter tube. The tube's cathode is coated with ^{235}U in which fission reactions are produced. The outer paraffine shell moderates the neutrons (adapted from [855]); (b) photo of a spherical neutron counter (proportional counter) filled with ^{3}He at 6.1 bar exploiting the elastic (n,p) reaction; diameter 23 cm, moderator shell thickness 9 cm [1010] (with kind permission of Taylor & Francis).

up to 100 keV (fig. 14.27). For 10 MeV neutrons we have $\sigma(n,p) \approx 1$ b. In a 1 cm thick plastic scintillator ($\rho = 1.2\,\mathrm{g/cm^3}$) with a molar proton fraction of about 30% the interaction length is $\lambda_n = (n\,\sigma)^{-1} = 2.2$ mm, that is, a maximum detection efficiency of 22% results.

High energy neutrons. Like other strongly interacting particles, neutrons with GeV energies or more release hadronic showers which can be detected in calorimeters (see section 15.3 on page 592). In high energy experiments with hadron calorimeters, as for example at LHC, a typical signature for a neutron is a shower in the hadron calorimeter, to which neither a track of a charged particle nor an energy deposit in the electromagnetic calorimeter can be associated, as fig. 14.2 shows.

14.6 Detection of particles with measurable decay lengths

In typical high energy physics experiments particles can be measured and potentially identified if their lifetime is long enough that they traverse and interact in the detector media by leaving 'signals', for example by ionisation or by showering. This typically means flight distances of the order of metres and lifetimes longer than nanoseconds. For particles with lifetimes much shorter than this, identification can only achieved by reconstructing their decay by a measurement of the decay products and then, for example, by computing their invariant mass. For not too short lifetimes, especially in the 10^{-12} s to 10^{-13} s range which applies to many heavy hadrons, a measurement of the flight distance l from production to decay point (decay length) can be added as an observable for particle identification:

$$l = \beta\gamma c\tau\,, \tag{14.24}$$

where $\beta\gamma c = p/m$ (p = momentum) is the Lorentz $\beta\gamma$ product of the particle in the laboratory frame and τ its lifetime (in the particle rest frame).

Table 14.2 Lifetimes of some particles for which decay length measurements are used in particle identification [746]. For charged particles only the positive charge state is listed.

	m (GeV/c^2)	τ	$c\tau$	l ($p = 10$ GeV/c)	Particle type
τ^+	1.776	0.290×10^{-12} s	87 µm	0.5 mm	lepton
K_S^0	0.498	0.895×10^{-10} s	2.68 cm	54 cm	meson ($d\bar{s}$)
Λ	1.116	2.632×10^{-10} s	7.89 cm	71 cm	baryon (uds)
D^0	1.865	0.410×10^{-12} s	123 µm	0.7 mm	meson ($c\bar{u}$)
D^+	1.869	1.040×10^{-12} s	312 µm	1.7 mm	meson ($c\bar{d}$)
Λ_c^+	2.286	0.200×10^{-12} s	60 µm	0.3 mm	baryon (udc)
B^0	5.279	1.519×10^{-12} s	456 µm	0.9 mm	meson ($d\bar{b}$)
B^+	5.279	1.641×10^{-12} s	492 µm	0.9 mm	meson ($u\bar{b}$)
Λ_b	5.619	1.425×10^{-12} s	427 µm	0.8 mm	baryon (udb)

In table 14.2 lifetimes of some leptons, mesons and baryons are collected, for which a decay length measurement is important for identification.

14.6.1 Detection of K_S^0 mesons and Λ baryons

K_S^0 mesons and Λ baryons have lifetimes in the order of 10^{-10} s, depending on their momentum. Their flight distance to decay typically is in the region of several tens of centimetres to metres (see table 14.2). The dominant decay modes contain two particles in the final state:

$$K_S^0 \to \pi^+\pi^- \quad (70\%)\,, \tag{14.25}$$

$$\Lambda \to p\pi^- \quad (64\%)\,. \tag{14.26}$$

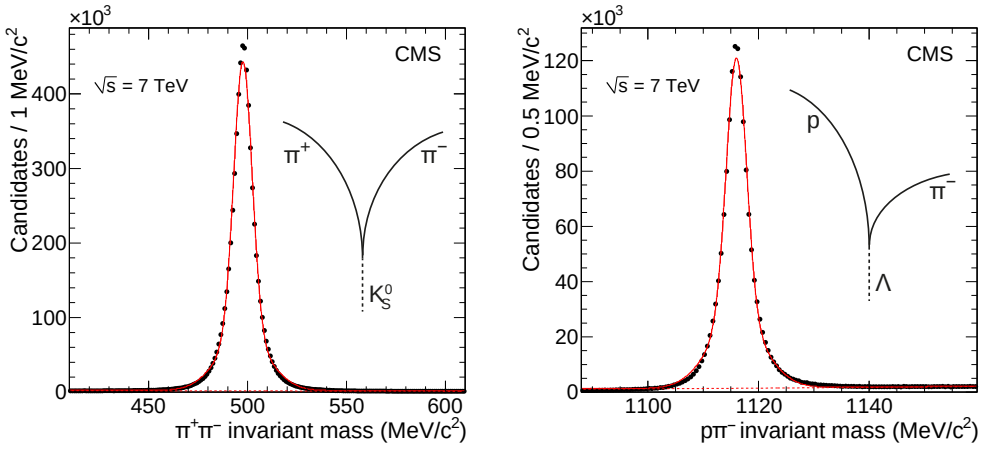
The signature of such decays are two particles visible as tracks in a track detector forming a 'V-shape' when leaving the detector. The momenta of the decay products can generally be well measured and hypothesising the masses of the decaying particles[2] in (14.25) and (14.26) the invariant mass can be determined. In most cases K_S^0 and Λ can be cleanly separated from background this way, as shown in fig. 14.29.

14.6.2 Detection of quarks and leptons with lifetimes in the picosecond range

For particle lifetimes in the order of picoseconds the decay length is in the range of 300 µm and above depending on momentum. Precise measurement of secondary vertices in this range is a primary goal of vertex detectors (see also section 8.1), because τ leptons and $q\bar{q}$ bound states containing a heavy quark (c, b) have picosecond lifetimes (table 14.2).

The corresponding length $c\tau$ lies in range of some hundred micrometres. Since the momentum spectra of heavy particles at high energy colliders can result in Lorentz factors of $\gamma > 10$, decay lengths in the range of millimetres are possible. These can be determined by precise measurement of track points by so-called vertex detectors placed as close as possible to the interaction point. The measured tracks are extrapolated back to their origin and a potential common intersection point (secondary vertex), being

[2]Stand-alone identification of π and p is usually not attempted. However, in the Λ decay, the pion in general is the decay particle with the smaller momentum.



(a) K^0_S peak. (b) Λ peak.

Fig. 14.29 (a) $\pi^+\pi^-$ invariant mass spectrum, (b) $p\pi^-$ invariant mass spectrum, measured with the CMS tracking detector after selection of outgoing 'V-shaped' tracks with decay lengths in the region of centimetres to metres [601]. The experimental resolutions (average of two Gaussians) are 8.2 MeV and 3.4 MeV, respectively, with little background (dashed lines). (source: CERN, CMS Collaboration, CC-BY 3.0). Sketched in (a) and (b) are typical decay topologies of K^0_S and Λ decays.

different from the primary vertex, is determined (see also section 8.1 on page 255) and section 9.4.6 on page 401.

Particularly suited as vertex detectors are semiconductor-based microstrip or pixel detectors, having spatial resolutions for a track point in the region of 5–10 µm. In the 1980s, before microstrip detectors became established, special gas-filled vertex detectors were also developed that obtained spatial resolutions of about 50 µm (see e.g. [273]).

Table 14.2 also shows that it is difficult to identify certain particles from decay length measurement only or to distinguish particles with bottom quark content from those with charm quark content. To achieve this either the reaction process should already tell that the examined vertices predominantly result from only c- or only b-quark decays or that a decay chain (b → c → s) occurred where the (secondary) b vertex occurs earlier in time than the (tertiary) c vertex (fig. 14.30), possibly followed by a fourth vertex due to the decay of the strange hadron, for example a K^0_S.

The decays of hadrons (mesons or baryons) with heavy quark content are largely dominated by the weak decay of the heavy quark (e.g. $b \to c + l\nu$). The decays are largely governed by the underlying interactions ruling the decay of the heavy quark of the system, while the other quarks of the hadron are less important for the system's lifetime (spectator model). Heavy quarks in hadrons often decay under emission of a lepton with comparatively large momentum transverse to the flight direction of the hadron. This results from the fact that states with heavy quarks are massive objects having decay products with high momenta emitted in all directions in the rest frame of the mother particle. This results in comparatively high momenta (up to $Mc/2$, where M is the mass of the heavy particle) transverse to the flight direction of the heavy quark. Leptons with high transverse momenta together with decay lengths significantly

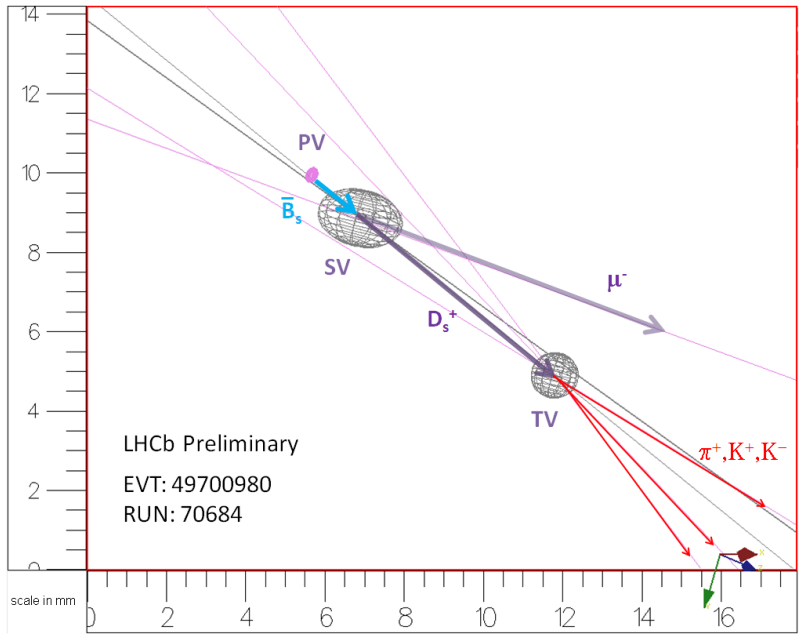


Fig. 14.30 Collision event recorded by the LHCb experiment in a proton–proton collision [656]: $pp \to B_s^0 + X \to D_s^+ \mu^- \bar{\nu}_\mu + X \to K^+K^-\pi^+ + \mu^- \bar{\nu}_\mu + X$. Starting from the primary vertex PV a B_s^0 meson ($s\bar{b}$ state) is created together with many other particles not shown here. After a flight distance of about 1.5 mm the B_s^0 decays at a secondary vertex (SV) into a μ^- and a D_s^+ meson ($c\bar{s}$ state) plus a neutrino which is not detected. The D_s^+ in turn decays after a flight distance of 6.5 mm into $K^+K^-\pi^+$ at a tertiary vertex (TV). The grey ellipsoids around the decay vertices indicate the uncertainty of the decay point measurement.

different from zero are thus signatures of heavy quark states.

The techniques developed over time to measure secondary vertices and to assign quark flavour contents of particles or jets (flavour tagging), use a multitude of algorithms, for example neural nets or similar multivariate techniques. A review can be found for example in [240, 863]. We here only mention the essential features and methods used for the design of tagging and identification observables.

Figure 14.30 illustrates the vertex identification problem using as example a pp-collision event $pp \to B_s + X$ where the B_s meson decays into a D_s meson containing a charm quark [656]. To identify particles that originate with high probability from the decay of a long-lived hadron with bottom- or charm-quark content, essentially two methods are employed: the impact parameter method and explicit reconstruction of decay lengths l.

The impact parameter d_0 defines the shortest perpendicular distance of a particle track to the primary vertex (see also section 9.4.6 and fig. 9.14 on page 402). Tracks with impact parameters significantly larger than the experimental resolution of this quantity are a signature for the existence of a secondary vertex. Figure 14.31(a) shows tracks coming from a secondary vertex (e.g. track 1). Track 2 for comparison is a badly measured track with an impact parameter different from zero though likely originating from the primary vertex. While the reconstruction of a secondary vertex requires at least two tracks, the advantage of the impact parameter method is the fact that d_0

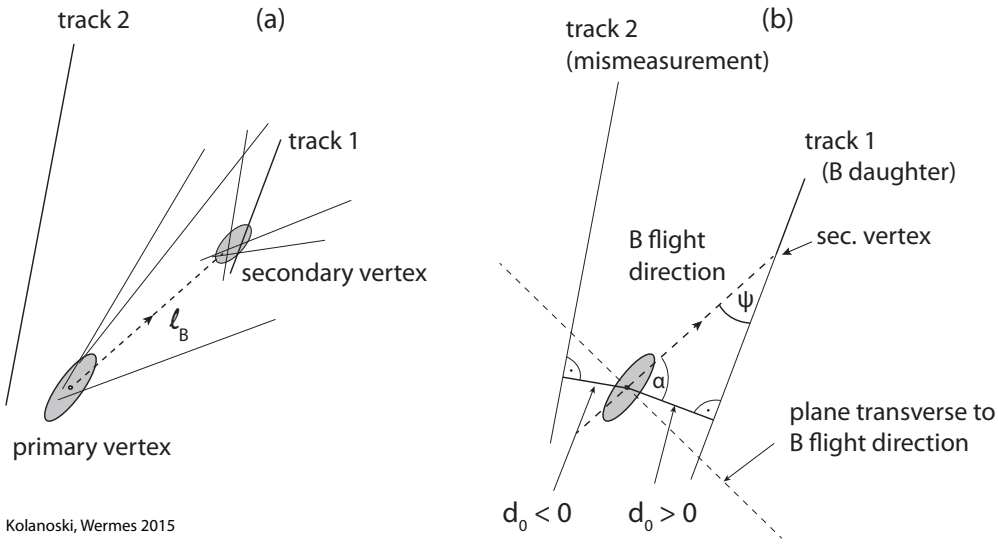


Fig. 14.31 (a) Tracks in a reaction with a secondary decay vertex (here, decay of a B meson). The ellipses indicate the measurement accuracies for the vertices. Track 1 originates from the secondary vertex. Track 2 is a badly measured track originating from primary vertex. (b) The impact parameters d_0 of the tracks are defined as positive or negative, respectively, depending on whether the angle α, between the direction perpendicular to the track towards the primary vertex and the assumed flight direction of the B meson (reconstructed jet direction), is smaller or larger than 90°.

can be determined for every track, even if only one charged track is emitted from the secondary vertex, as for example in the decay $\tau^- \to \mu^- + \bar{\nu}_\mu + \nu_\tau$.

Often d_0 is only measured in a plane perpendicular to the magnetic field (xy plane in a coordinate system, as shown for example in fig. 14.23), in which the measurement resolution is oftentimes better. The average decay length in this plane is $\langle L_{xy} \rangle = \beta\gamma c\tau \sin\theta_\mathrm{B}$ with τ being the lifetime of the decaying mother particle. The impact parameter of a given track $d_0 = L_{xy} \sin\psi$ thus on average depends on the lifetime as

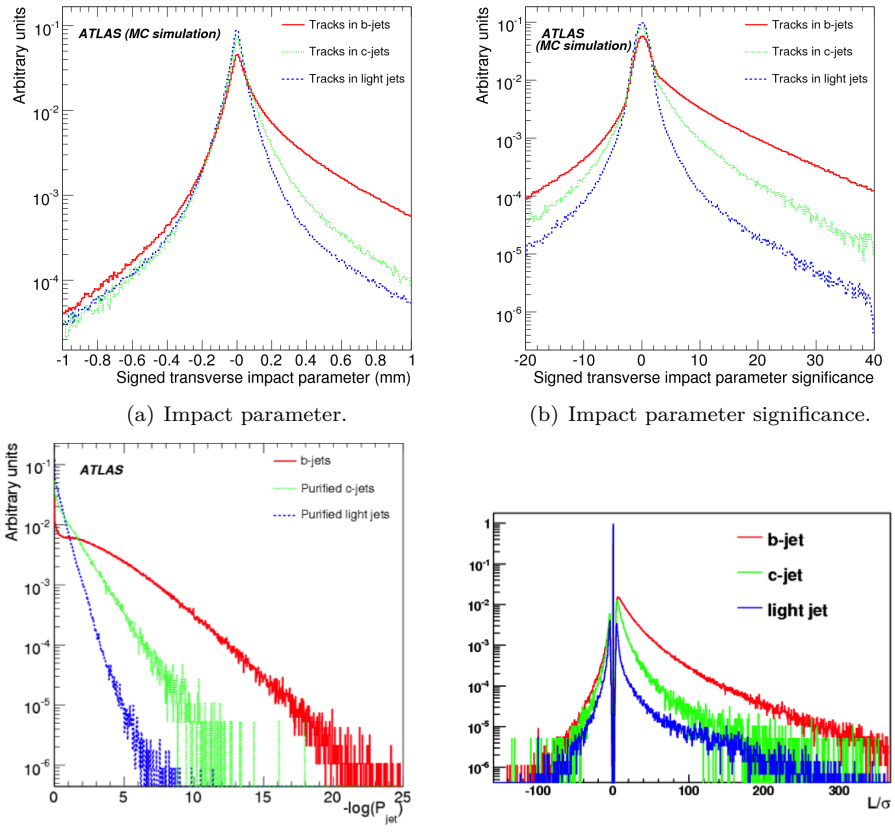
$$\langle d_0 \rangle = \langle L_{xy} \rangle \sin\psi = \beta\gamma \mathrm{c}\tau \sin\theta_\mathrm{B} \sin\psi \,, \tag{14.27}$$

where θ_B is the polar angle of the B meson (approximated by the direction of the decaying particle estimated from its decay products) relative to the beam axis and $\psi = \phi - \phi_B$ (ϕ, ϕ_B = azimuthal angles) the angle between the B flight direction (experimentally often equated with the 'jet' direction) and the track in the transverse plane.

With $\sin\psi \propto 1/\gamma$ and for an isotropic decay angle distribution of the mother particle the impact parameter distribution is independent of the γ-factor. To first order the transverse impact parameter $\langle d_0 \rangle$, averaged over many tracks and decays, is directly proportional to the lifetime:

$$\langle d_0 \rangle \approx \frac{\pi}{2}\, c\tau \tag{14.28}$$

and does not depend on the momentum of the decaying particle. In high energy physics



(a) Impact parameter.

(b) Impact parameter significance.

(c) Jet probability.

(d) Decay length significance.

Fig. 14.32 *Impact parameter tagging* [5]. (a) Monte Carlo simulation of the impact parameter d_0 of tracks in the xy-plane (see fig. 14.23) of the ATLAS detector for b quarks (red), c quarks (green) and 'light' uds quarks (blue); (b) impact parameter significance of the same tracks (same color assignment) and (c) resulting jet-flavour probabilities obtained from the positive significances S_{d_0}; (d) decay length significance distributions. Source: CERN/ATLAS Collaboration.

experiments a measured non-zero flight distance of heavy hadrons inside jets is interpreted as a signature for heavy quarks being the mother particles of the jet (b or c tagging). In order to identify heavy quarks or leptons this way, many algorithms use positive and negative impact parameters $\pm d_0$ of the tracks of a jet, where the sign of d_0 is defined to sort impact parameters likely originating from a heavy particle decay ($+d_0$) from those being likely due to a mis-measurement ($-d_0$). A positive d_0 is assigned to tracks which cross the assumed flight direction (often the reconstructed jet direction) of the decaying mother particle (heavy-quark hadron) in the same hemisphere as the mother; a negative d_0 is assigned if the track crosses the flight axis of the mother in the opposite hemisphere, where the two hemispheres are defined relative to the plane through the primary vertex and perpendicular to the flight direction of the mother hadron (see also the illustration in fig. 14.31(b)).

Figure 14.32(a) shows an impact parameter distribution of tracks from jets originating from different quark flavours [5]. The lifetime of b quarks typically is three times larger than that of c quarks (see table 14.2). With the help of the sign convention it is possible to separate the resolution and background contributions in a measured d_0 distribution from contributions coming from decay lengths of long-lived particles. The right side of the distribution contains positive impact parameter values $d_0 > 0$, predominantly tracks from secondary vertices of a long-lived b or c quark, whereas negative ($d_0 < 0$) values predominantly originate from background contributions and from mismeasured tracks reflecting the detector resolution, both contributing on both sides of the distribution in fig. 14.32(b). From the negative part of the distribution a calibration of resolution and background contributions is possible without the need of Monte Carlo simulations. From the positive side then the efficiency for heavy quark detection and the rejection power of light quarks can be determined.

The impact parameter resolution σ_{d_0} strongly varies with the track momenta and track angles relative to the beam direction, the number of hits available for track reconstruction, as well as contributions from multiple scattering. The impact parameter significance

$$S_{d_0} = \frac{d_0}{\sigma_{d_0}} \tag{14.29}$$

accounts for the resolution varying from jet to jet. For tracks originating from the primary vertex (PV), one can determine the probability that a significance S_{d_0}, more positive/negative or equal to the measured one, occurs [9]:

$$P(S_{d_0}) = \int_{-\infty}^{-|S_{d_0}|} R(s)ds\,, \tag{14.30}$$

where $R(s)$ is the resolution function which can be obtained from a fit to measured data from the left side of the distribution of fig. 14.32(b), to which only resolution effects contribute. This can be used to separate heavy-quark jets, containing secondary decays with tracks not originating from the PV, from light (u, d, s)-quark jets where all tracks come from the PV. A jet probability is calculated [9] as

$$P_{jet} = P_0 \times \left(\sum_{j=0}^{N_{trk}-1} \frac{(-\ln P_0)^j}{j!} \right), \qquad P_0 = \prod_{i=1}^{N_{trk}} P_i(S_{d_0})\,, \tag{14.31}$$

with $P_i(S_{d_0})$ as in (14.30). P_{jet} is given by the product of the individual track probabilities to originate from the primary vertex (P_0), multiplied by a weighting factor (the term in brackets which accounts for the number of tracks in a jet). The distribution of P_{jet} should be flat for light-quark jets, free of tracks from long-lived particles, while for heavy b- or c-quark jets it assumes values near zero. Note that for a flat P_{jet} distribution the variable $-\log P_{jet}$ as plotted in fig. 14.32(c) is $1/x$ distributed ($x = -\log P_{jet}$), as the (blue) histogram for light quarks shows.

This information, together with the chosen specifications for S_{d_0}, leads to a representation of the *b-tagging* properties of a detector as shown in fig. 14.33 [427], where on the x-axis the detection efficiency for a b jet and on the y-axis a measure of the suppression of light quark jets (= one minus *light-jet efficiency*) is plotted. The light-jet efficiency should be as small as possible while keeping a high b-jet efficiency. For a chosen b-jet efficiency of 60% for example, the suppression of light quark jets amounts to

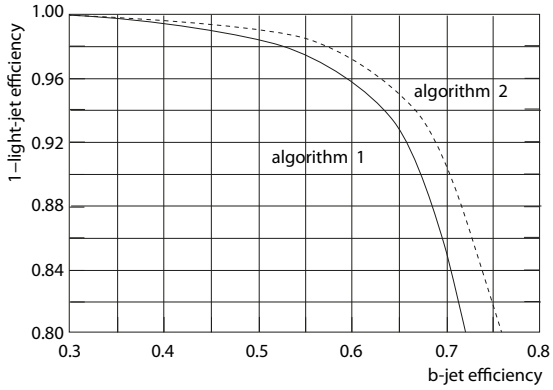


Fig. 14.33 Suppression (= one minus the *light-jet efficiency*) of jets from light (uds) quarks versus b-quark-jet efficiency (adapted from [427]). The two curves show the performance of two different algorithms: algorithm-2 performs better than algorithm-1.

about 96% (97% for algorithm-2). More advanced algorithms include likelihood ratios for tracks to originate from heavy or light quark decays, respectively (see e.g. [9]).

If more than one track is emitted from the decay vertex, instead of the impact parameter the decay length l can be used as observable to identify heavy quarks or τ-leptons [240,5]. Figure 14.32(d) shows the decay length significance, calculated very similar to the impact parameter significance in (14.29), $S_l = l/\sigma_l$, for different quark flavours. By statistical comparisons of template distributions to experimental distributions likelihood quantities or likelihood ratios are formed, from which probabilities for the respective b- or c-quark flavour or the (uds)-quark flavour can be calculated (see e.g. [5]). Multivariate analysis (MVA) methods are used to optimise the performance for b-tagging.

15
Calorimeters

15.1 Introduction

The determination of the energy of a high-energetic particle is called 'calorimetry' and the corresponding detectors are *calorimeters*. The detection of particles in calorimeters is destructive. In general a particle hitting a calorimeter initiates a chain of inelastic reactions by which the particle's energy is distributed amongst many secondary particles leading to the formation of a *particle shower*. At the end of the reaction cascade the charged particles in the shower are stopped, mainly by ionising the detector medium.

Despite the designation 'calorimeter' it is not the generated heat which is measured but it is the generated charge or the scintillation light or sometimes also Cherenkov light generated in the shower. For the calorimeters under discussion in this chapter the temperature increase caused by dissipation of the particle energy is not measurable (in contrast to the 'cryo detectors' in section 16.7.2). When the signals, for example due to charge or light, are collected from the whole volume, we are talking about *homogeneous calorimeters*. Otherwise, if dense 'passive' media ('absorbers') alternate with light 'active' media for readout of the signals, one talks about *sampling calorimeters* (section 15.4.1). An important design criterion for a calorimeter is the requirement that in the considered energy range the measurable signals are as proportional to the incident particle energy as possible.

According to the way the shower develops we distinguish between *electromagnetic calorimetry* for electrons, positrons and photons and hadronic calorimetry for hadrons (singly and in jets). The radiation length, introduced in section 3.3.2, is a characteristic length measure for electromagnetic showers while the hadronic absorption length, introduced in section 3.6, plays this role for hadronic showers. More detailed accounts of calorimeters are given for example in [982, 385].

For charged particles the possibility exists to measure the energy through momentum measurement in a magnetic field. However, the relative momentum resolution deteriorates proportionally with increasing momentum, typically like (see section 7.10.9)

$$\frac{\sigma_p}{p} \approx 0.1 \ldots 1\,\% \, p/\mathrm{GeV}\,, \tag{15.1}$$

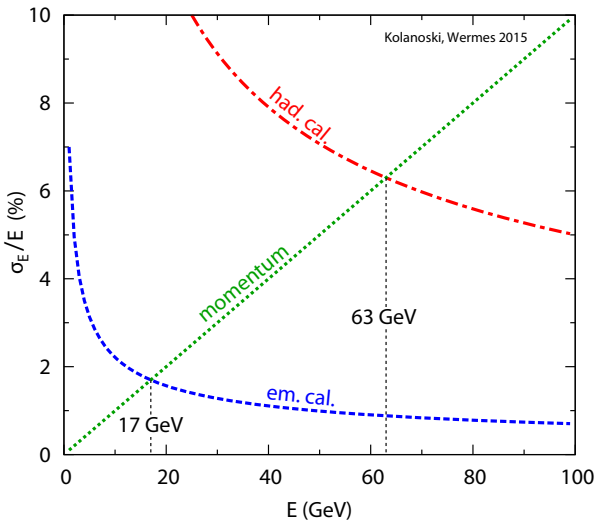


Fig. 15.1 Comparison of typical energy resolutions in a detector. In the example the energy resolutions of electromagnetic and of hadronic calorimeters are assumed to be $\sigma_E^{em}/E = 7\,\%/\sqrt{E}$ and $\sigma_E^{had}/E = 50\,\%/\sqrt{E}$, respectively. For the energy determination by momentum measurement a momentum resolution of $\sigma_p/p = 0.1\,\%\,p$ is assumed. With these values the energy determination by calorimetric measurements becomes better than by momentum measurements at 17 GeV and 63 GeV, respectively.

while the energy resolution of a calorimeter improves with increasing energy about like

$$\frac{\sigma_E}{E} \approx \begin{cases} \dfrac{2\text{–}15\,\%}{\sqrt{E/\mathrm{GeV}}} & \text{elm.} \\ \dfrac{35\text{–}120\,\%}{\sqrt{E/\mathrm{GeV}}} & \text{had.} \end{cases} \tag{15.2}$$

The ranges of the numerical values apply for most of the calorimeters.[1] The $1/\sqrt{E}$ dependence of the relative energy resolution, typical for totally absorbing calorimeters, can be traced back to the Poisson statistics of the shower particles. In contrast, momentum determination through measurement of the track curvature in a magnetic field deteriorates with increasing momentum (section 9.4.3). Therefore the resolution curves cross at a point above which calorimetry is better than momentum measurement (fig. 15.1). The choice of the crossing point for detectors which feature both tracking detectors and calorimeters is usually determined by cost considerations. According to (9.63) (page 397) the momentum resolution is proportional to $1/(B\,L^2)$. Therefore, for a given magnetic field B the measured track lengths L have to increase quadratically to compensate the resolution deterioration with increasing momentum. In contrast, the size of a shower increases only logarithmically, as will be explained in the following. Since the cost of a detector increases roughly proportionally to its volume, for cost reasons alone calorimetry is advantageous at higher energies.

An additional important advantage of calorimeters is their sensitivity to electromagnetically and strongly interacting particles, in particular also to neutral particles like photons, neutrons or neutral kaons. With these capabilities calorimeters are the appropriate detectors for measurement of highly energetic hadron jets which generally

[1] The momentum and energy dependences in (15.1) and (15.2) are valid in a range of medium energies and momenta, respectively. Additional terms which more adequately describe the behaviour at high or low energies, respectively, will be discussed later (see section 15.4.3).

consist of a mixture of charged and neutral, electromagnetically and strongly interacting particles. For the measurement of jets it is important that all particles, independent of their individual interaction with the detector, deliver the same signal for the same energy. This is a great challenge for hadron calorimetry (see section 15.6.3).

Overview of the chapter. This chapter has the following structure. After this introduction the next two sections describe the development of electromagnetic and hadronic showers, respectively. It follows an overview of the criteria which have to be accounted for when constructing a calorimeter and which determine its quality. In the two following sections we discuss different realisations of electromagnetic and hadronic calorimeters. Finally typical examples of complete calorimeter systems in experiments at accelerators are compared.

In this chapter we want to restrict ourselves to calorimetry with detectors at particle accelerators and to sufficiently high energies a (larger than about 10 MeV for electromagnetic showers and larger than about 1 GeV for hadronic showers). Applications of calorimetric methods in experiments which are not conducted at accelerators, for example for the detection of cosmic ray initiated air showers or of nuclear recoils in reactions of dark matter, will be described in chapter 16. In nuclear physics energies are also determined by range measurements, see section 3.2.4 and fig. 3.21.

15.2 Electromagnetic showers

Highly energetic electrons, positrons and photons develop so-called *electromagnetic showers* when passing through a dense medium. Showering can be described by repeated radiation of photons by electrons and positrons and the subsequent pair creation by the photons. In the following we will discuss models and characteristic parameters describing electromagnetic showers.

15.2.1 Models of shower development

For describing the development of electromagnetic cascades an often used reference is still Rossi's analytical calculation published 1952 in his meanwhile classical book 'High Energy Particles' [835]. The most essential properties of a shower are very well accounted for by Rossi's 'Approximation B' (chapter 5 in [835]). The most important assumptions are:

(i) Only bremsstrahlung and pair production are considered using the respective cross sections in the asymptotic high-energy approximation (sections 3.3 and 3.5.5).

(ii) The energy loss per path length by ionisation is energy independent and can be approximated by the critical energy per radiation length (section 3.3.3).

(iii) Multiple scattering is neglected and the shower development is treated one-dimensionally without lateral spread.

With these assumptions a very good assessment of the characteristic properties of longitudinal shower development can already be obtained. An important property is the so-called 'X_0 scaling', which is the approximate material independence of shower profiles if lengths are measured in units of the radiation length introduced in chapter 3 (eq. (3.89)).

For a long time analytical calculations based on parametrisations, as presented in [835], have been the only possibility to evaluate experimental data involving electromagnetic showers. Meanwhile very mature simulation programs exist which hardly

(a) Bremsstrahlung. (b) Pair production.

Fig. 15.2 Diagrams for bremsstrahlung and pair production, the most important processes for the development of electromagnetic showers.

require any restricting assumptions (if not for curtailing computing time) and which very well describe experimental data (see the corresponding description on page 588). On the other hand the analytical calculations have led to a deeper understanding of the shower parameters which are essential for detector reconstruction and to the development of simple formulae describing the showers. In the following we will therefore employ analytical calculations in order to convey a basic understanding and to derive some important rules of thumb.

The most important shower processes. According to Rossi's 'Approximation B' [835] the most important processes for the development of electromagnetic showers are bremsstrahlung for electrons and positrons as well as pair production for photons (fig. 15.2, see also the sections 3.3 and 3.5.5), which both proceed in the field of a nucleus with nuclear charge corresponding to the atomic number Z.

Since the cross sections of both processes increase proportionally to the square of the nuclear charge,

$$\sigma \propto Z^2 \qquad \text{(bremsstrahlung, pair production)}\,, \tag{15.3}$$

the detector medium should have high Z making lead, tungsten and uranium the preferred materials. In both processes the secondary particles are produced with increasing energy more and more into the forward direction:

$$\theta \propto \frac{1}{\gamma} = \frac{m_e}{E}\,. \tag{15.4}$$

The characteristic length for both processes is the radiation length $X_0 \propto 1/Z^2$ (see eq. (3.88) and table 3.4). At high energies the absorption length of photons becomes proportional to the radiation length,

$$\lambda \approx \tfrac{9}{7}\, X_0\,, \tag{15.5}$$

and for electrons X_0 determines the relative energy loss per path length:

$$\frac{dE}{E} = \frac{dx}{X_0}\,. \tag{15.6}$$

With the critical energy E_c, that is, the energy at which the energy loss through bremsstrahlung equals that through ionisation (see fig. 3.27 and eq. (3.92)), it follows from (15.6) that

$$\left.\frac{dE}{dx}\right|_{ion} (E_c) \approx \frac{E_c}{X_0}\,. \tag{15.7}$$

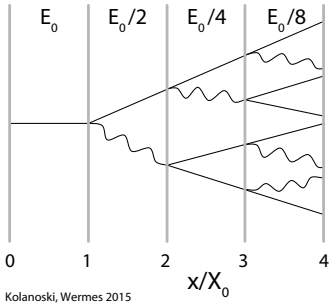


Fig. 15.3 Simplified model of shower evolution. At each step of length X_0 the number of particles doubles and the particle energy is halved.

A simple model for the shower evolution. For the visualisation of the evolution of electromagnetic showers a simplified model introduced by Heitler [517] (see also the description by Rossi [835]) is often reverted to. The model assumes that electrons and photons only interact with matter through bremsstrahlung and pair production until the critical energy E_c is reached. Subsequently the remaining electrons and positrons only lose their energy through ionisation. From this model it follows that the total deposited energy E_0 is proportional to the total number of produced electrons and positrons and proportional to their total path length in the medium:

$$\text{total number} \qquad N_{tot} \approx \frac{E_0}{E_c}\,, \tag{15.8}$$

$$\text{total path length} \qquad s_{tot} \approx \frac{E_0}{E_c}\,X_0\,. \tag{15.9}$$

Further we want to assume that after each radiation length one of the two processes occurs and that each of the two particles emerging from a reaction carries half of the energy of the incoming particle. This is sketched in (fig. 15.3):

$$\text{bremsstrahlung} \qquad E_e(n\,X_0) = E_\gamma(n\,X_0) = \frac{1}{2}E_e\left[(n-1)\,X_0\right], \tag{15.10}$$

$$\text{pair production} \qquad E_{e^+}(n\,X_0) = E_{e^-}(n\,X_0) = \frac{1}{2}E_\gamma\left[(n-1)\,X_0\right]. \tag{15.11}$$

When the electrons and positrons in this cascade reach the energy E_c they deposit their whole remaining energy without a further radiative process.

After a distance $s = t\,X_0$ (t is the path length in units of radiation length) the number and the energy of the particles are

$$N = 2^t, \qquad E = \frac{E_0}{2^t}\,. \tag{15.12}$$

The maximal number and maximal length of a shower is reached at the energy

$$E = E_c = \frac{E_0}{2^{t_{max}}}\,, \tag{15.13}$$

yielding

$$N_{max} = \frac{E_0}{E_c}\,, \tag{15.14}$$

$$t_{max} = \frac{\ln E_0/E_c}{\ln 2}\,. \tag{15.15}$$

Table 15.1 Parameters for the development of electromagnetic showers for four different materials (material parameters taken from [762]). The values for t_{max} and $t^{98\%}$ have been calculated using the Longo formula (15.17) with the relations (15.18), (15.19) and the parameter b as given in the table (partially extracted from the respective figure in [201]).

Material	Z	X_0	E_c	b	t_{max}		$t^{98\%}$		R_M	$\frac{R_M}{X_0}$
					10	100	10	100		
		(mm)	(MeV)		(GeV)	(GeV)	(GeV)	(GeV)	(mm)	
H_2O	1, 8	361	78.6	0.63	4.3	6.6	13.8	17.3	83	0.23
Al	13	89	42.7	0.58	5.0	7.3	15.4	18.8	45	0.51
Fe	26	17.6	21.7	0.53	5.6	7.9	17.1	20.6	18	1.02
Pb	82	5.6	7.4	0.50	6.7	9.0	19.3	22.7	16	2.86

The important results are that N_{max} depends linearly on the energy E_0, making it usable as a measure for the energy, and that t_{max} only logarithmically increases with the energy, which obviously makes calorimetry attractive, in particular at high energies:

$$N_{max} \propto E, \qquad t_{max} \propto \ln E + const\,. \tag{15.16}$$

This highly simplified model already reflects the essential features of the development of an electromagnetic shower. For a more precise assessment measurements and simulations are necessary.

Simulation of electromagnetic showers. As already mentioned, the development of electromagnetic showers is theoretically well known and therefore can be simulated very precisely. A popular simulation program for electromagnetic showers is the EGS code (EGS = electron–gamma shower) [592] which is specialised on electron–photon transport. In addition there are respective parts in more general detector simulation programs, for example in the Geant4 program package (see section 3.7).

15.2.2 Characteristic size of electromagnetic showers

The dimensions of showers essentially determine the construction of calorimeters. In the following we give some empirical formulae for the calculation of shower sizes with parameters which have to be taken from data or simulations.

15.2.2.1 Longitudinal shower profile

An empirical formula for the longitudinal energy distribution of a shower was derived by Longo and Sestili [672]:

$$\frac{dE}{dt} = E_0 \frac{b^a}{\Gamma(a)} t^{a-1} \mathrm{e}^{-bt}\,. \tag{15.17}$$

The formula corresponds to a gamma distribution function, which in this context is often called the *Longo formula*. The parameters a and b depend on E_0 and Z, and Γ is the gamma function [976]. The maximum of the function (15.17) is called the *shower maximum* and lies at

$$t_{max} = \frac{a-1}{b}\,. \tag{15.18}$$

A handy formula for t_{max} is given in [746]:

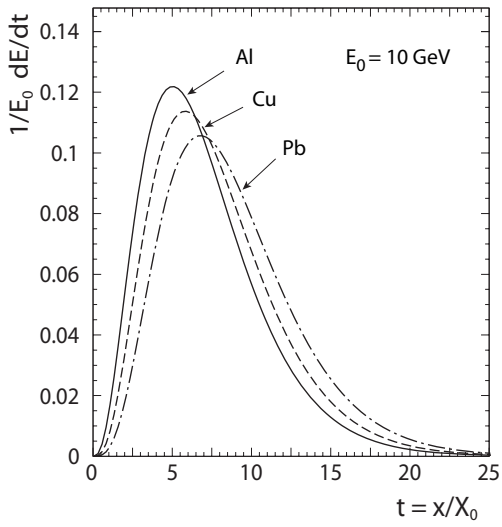


Fig. 15.4 Longitudinal profiles of 10-GeV electron showers calculated according to the Longo formula (15.17) for three different materials. The parameters are: $b = 0.5$, the same for all materials, and a calculated from (15.18) and (15.19). Due to the normalisation factor $1/E_0$ the area under each curve is the same. The differences in the shower development are due to the different critical energies of the three materials (see table 3.4).

$$t_{max} = \ln \frac{E_0}{E_c} + \begin{cases} -0.5 & \text{(electrons)} \\ +0.5 & \text{(photons)}\,. \end{cases} \tag{15.19}$$

Note that t_{max} (in units of X_0) is material dependent through the critical energy E_c, which is also a feature of the corresponding model formula (15.15). For decreasing critical energy the shower becomes more extended, as demonstrated in fig. 15.4. The parameter b in (15.17), which determines the decline after the shower maximum, amounts to about 0.5 for heavy elements with relatively minor energy dependence. In [746] the b values of some detector media, as determined by simulation, are graphically displayed as a function of energy (fig. 32.21 in [746], see also table 15.1). If b is assumed to be known, (15.18) together with (15.19) also yields a, and with this input the Longo formula (15.17) can be evaluated. In fig. 15.5 shower profiles of simulated electron showers for different energies are compared to fits to the Longo formula. The plot demonstrates that the Longo formula is for most purposes a very good approximation for the longitudinal extension of a shower.

Since the longitudinal distribution of the showers has very long tails, leakage losses have to be accepted in order to stay within a reasonable detector length. For 98% detection of the deposited energy a rough estimate yields [383]:

$$t^{98\%} \approx t_{max} + 13.6 \pm 2.0\,. \tag{15.20}$$

Examples for shower parameters are compiled for four materials and two energies in table 15.1.

15.2.2.2 Lateral shower profile

At high energies the two shower processes scatter the particles under a very small angle, proportional to $1/\gamma$ (see eq. (15.4)). Therefore the lateral shower dimension is determined by other mechanisms, namely by multiple scattering of the low-energy charged particles and Compton scattering of the photons. The lateral shower dimension is characterised by the 'Molière radius'

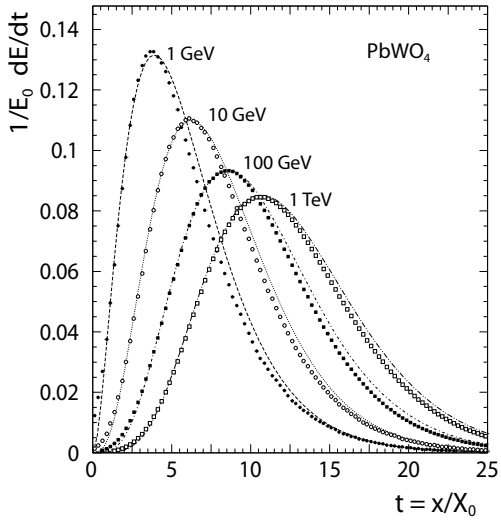


Fig. 15.5 Longitudinal profiles of simulated electron showers in $PbWO_4$ crystals (table 3.4) with various primary energies (simulated data where taken from slides of a tutorial given by [684]). The simulations are fitted with the Longo formula (15.17) which reproduces the essential features even if not perfectly. The shower maxima which can be calculated from the fitted parameters a and b using (15.18) agree within less than 7% with the estimates from (15.19).

R/R_M	1	2	3.5
$\Delta E/E_0$ [%]	90	95	99

Table 15.2 Fractions of a shower in a cylinder with a radius given in units of the Molière radius.

$$R_M = \frac{E_s}{E_c} X_0 \,. \tag{15.21}$$

Because of the connection to multiple scattering (section 3.4) the energy $E_s = 21.2\,\text{MeV}$ in (15.21) is the same as the one appearing in the calculation of the scattering angle parameter in (3.100). Inside a cylinder with the radius R_M around the shower axis about 90% of the energy is deposited (table 15.2). For the measurement of the shower position the granularity of the shower detector has therefore to be of the order of a Molière radius or finer.

Since the longitudinal dimension scales with the radiation length X_0 the shape of a shower is given by the ratio

$$\frac{R_M}{X_0} \propto \frac{1}{E_c} \propto Z \,. \tag{15.22}$$

The increase of this ratio with Z means that, for example a shower in aluminium is slimmer than in lead (last column in table 15.1). Absolutely, however, the aluminium shower is three times broader than the lead shower (second but last column in table 15.1).

15.2.2.3 LPM effect

At this point we also want to mention the Landau–Pomeranchuk–Migdal effect (LPM effect) which becomes essential for the longitudinal shower extent at very high energies (above about 1 PeV in water and 10 TeV in lead). A derivation of the effect is given in appendix G, while a more detailed treatment of the LPM effect and similar effects can be found in [610] (see also [746]).

The effect implicates the suppression of the two most important processes of the shower development, bremsstrahlung and pair production, with increasing energy due

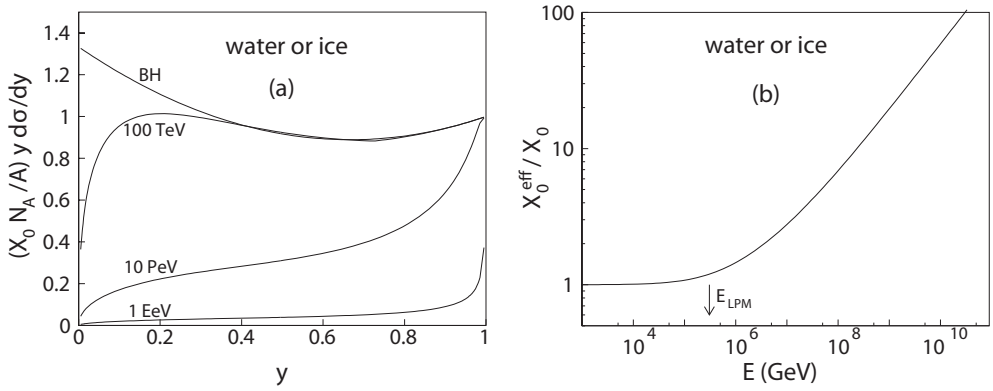


Fig. 15.6 (a) Spectra of bremsstrahlung from electrons in water or ice as a function of the radiated relative energy $y = k/E$ for different primary energies E. For reference the Bethe–Heitler cross section ('BH') for low energy bremsstrahlung is also plotted. The vertical axis denotes the number of photons, multiplied by $y = k/E$, emitted by an electron per radiation length in a y interval (the factors can be derived from (3.3) and (3.2) with $l = X_0/\rho$). Here the radiation length X_0 has the units mass per area. The LPM curves have been calculated following the quantum mechanical formulation by Migdal [707] with the numerical approximations of [912]. (b) The suppression factor as a function of energy given as the ratio of an *effective radiation length*, as defined in the text, to the normal radiation length. The onset of suppression is roughly given by E_{LPM}, defined in (15.23) and (G.6).

to quantum mechanical interference. This suppression leads to an extension of the longitudinal shower dimension accompanied by a concurrent increase of fluctuations. In appendix G the characteristic energy E_{LPM}, at which the effect is becoming important, is derived as

$$E_{\mathrm{LPM}} = 7.7\,\mathrm{TeV}\,\frac{X_0}{\mathrm{cm}} = \begin{cases} 4.3\,\mathrm{TeV} & \text{lead} \\ 305\,\mathrm{TeV} & \text{water or ice} \end{cases} \tag{15.23}$$

where the radiation length is given in absolute length units (cm). Even at the highest energies of accelerators the implications for particle detection are small but can play a role for the detection of cosmic radiation, as demonstrated by the following example.

Example: LPM effect in water or ice. As an example of the influence of the LPM effect we take the detection of electromagnetic showers with the IceCube detector (section 16.6.5 and fig. 16.34) in the Antarctic ice [230] (the relevant shower parameters are included in table 15.1 on page 588). With the IceCube detector highly energetic electromagnetic showers which are triggered by electron or τ neutrinos (so-called *cascade events*) can be detected through the Cherenkov radiation generated by the showering particles.

Figure 15.6(a) shows for highly energetic electrons the normalised differential cross section for bremsstrahlung in ice as a function of the normalised radiated energy $y = k/E$ for different primary energies E. With the chosen normalisation the vertical axis displays the number of photons radiated per radiation length in a y interval multiplied by y, hence the fraction of the primary energy radiated in the y interval. The plot demonstrates that low k/E values are increasingly suppressed with rising energy and that on the whole the probability for radiation becomes smaller.

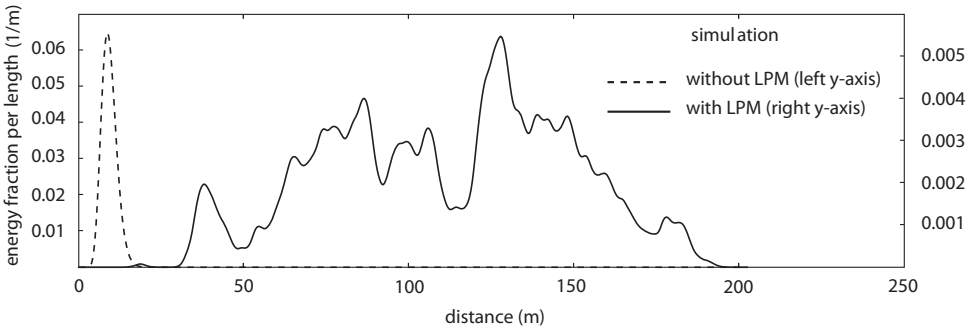


Fig. 15.7 Simulation of the shower of a 10-EeV electron in ice: longitudinal shower profile without and with taking into account the LPM effects. Author: B. Voigt (see also [961]).

Figure 15.6(b) shows the ratio of the integral over the Bethe–Heitler curve in pane (a) to the integral over the corresponding suppressed curve as a function of the primary energy E. This ratio can be interpreted as an *effective radiation length* in units of the normal radiation length X_0. The growth of the effective radiation length starts roughly around the energy E_{LPM}, thus for ice at about 305 TeV. At larger energies this results in an increased stretching of the shower accompanied by a growth of the statistical fluctuations in the longitudinal shower direction (fig. 15.7).

Due to the suppression of low-energy radiation and symmetric pair production, highly energetic shower particles arise more frequently but at larger distances, generating local showers with high particle density. In fig. 15.7 the shower with an energy of 10 EeV (10^{19} eV) is about 200 m long with the energy distributed on average relatively uniformly over the whole distance. However, this random example cannot be generalised because the variations from shower to shower are large. Without the LPM effect the shower would have a profile as shown in fig. 15.5 with the shower maximum at about 9 m and a 98% energy deposition at about 14 m (sketched in fig. 15.7 by the dashed curve).

With the enormous longitudinal stretching of the shower the direction resolution for high-energy showers can be considerably improved. This has significant advantages, for example, in the search for astronomical point sources emitting electron neutrinos which generate electron showers in ice or water detectors [230].

15.3 Hadron showers

Highly energetic hadrons also form particle showers when passing through dense media. These hadronic showers can be exploited for calorimetric energy measurements. In the case of hadrons, however, the reactions contributing to the shower formation are much more abundant than in the electromagnetic case and therefore the theoretical modelling and simulations of hadronic showers are much more difficult. The fluctuations of the contributions of the various reactions with very different signal efficiencies pose a large problem for the energy resolution.

In this section we present the essential properties of hadronic showers, in particular the fluctuations in the shower development. The impact on the design of hadron calorimeters with optimal energy resolution will be discussed in section 15.6.

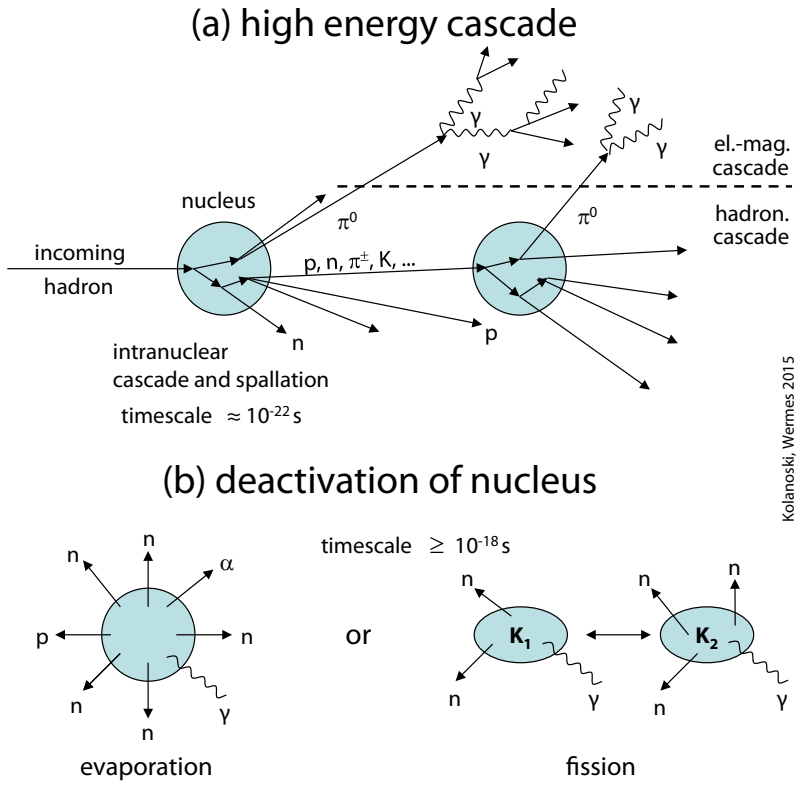


Fig. 15.8 (a)Schematic development of a hadronic cascade in a dense medium. The incoming hadron interacts inelastically with a nucleus transferring most of its energy onto a single nucleon whereby highly energetic particles are emitted into the forward direction. A minor part of the energy is distributed among the other nucleons forming a so-called *intranuclear cascade*. The high-energy particles, as long as they do not spontaneously decay into photons, like the neutral pions, again interact with nuclei thus forming a hadronic cascade. The photons from π^0 and η decays start electromagnetic cascades which develop independently from the hadronic shower. The remaining nucleus (left) is highly excited and initially releases its energy in a spallation process through spontaneous emission of nucleons and nuclear fragments with energies around 100 MeV. (b) Deactivation of the nucleus. When the remaining excitation energy reaches the level of the binding energies of the nucleons the residual energy is dissipated by evaporation of nucleons, primarily neutrons, and light nuclear fragments. For heavy nuclei, like uranium and lead, nuclear fission instead of evaporation followed by the emission of neutrons and gammas can take place as well. More details are discussed in the text.

15.3.1 Shower development

In a dense medium a hadronic cascade schematically develops as depicted in fig. 15.8(a).

High-energy cascade. An incoming highly energetic hadron interacts inelastically with a single nucleon of a nucleus generating highly energetic particles, preferentially into the forward direction. If these secondary particles leave the nucleus with sufficient energy they can again inelastically generate particles so that eventually a cascade

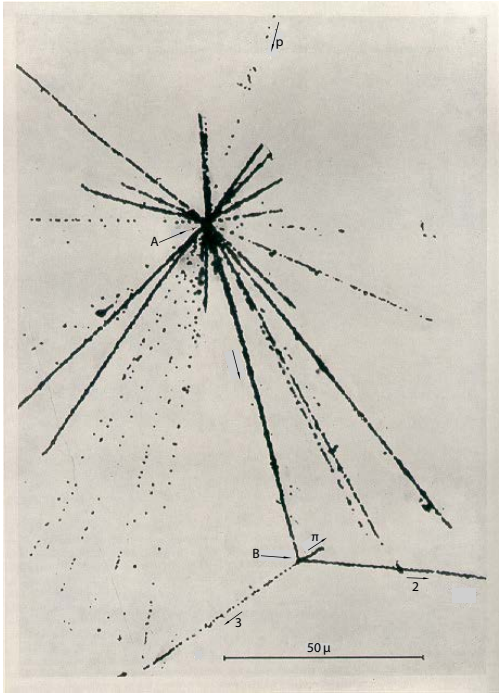


Fig. 15.9 Nuclear star generated by the interaction of high-energy cosmic radiation (coming from above, labelled 'p') with a silver atom in a photoemulsion. The thinner tracks come out of the high-energy cascade and of the spallation. The thicker tracks designate strongly ionising slow nucleons and nuclear fragments from the evaporation process. Reprinted from [340], with friendly permission of Taylor & Francis Ltd.

develops.

Spallation and intranuclear cascade. The particles of the high-energy cascade are preferentially produced on a single nucleon of a nucleus while the other nucleons take on the role of 'spectators'. The nucleus becomes highly excited by the collision and by the interaction of the created hadrons with the other nucleons of the same nucleus. In a first step, the so-called *spallation*, the excited nucleus releases its energy through emission of nucleons and light fragments (fig. 15.9). This process is often described by the formation of a 'intranuclear cascade' in which the produced particles scatter in the nucleus and generate more particles as long as they are either above the threshold for inelastic reactions or they have left the nucleus. Spallation nucleons and fragments have kinetic energies in the order of 100 MeV with tails up to the GeV region and are emitted on a time-scale of 10^{-22} s. Neutrons are about 1.5 times more frequent than protons, corresponding roughly to the relative abundances in the target nuclei [982]. The charged spallation products, mainly protons, substantially contribute to the total energy deposit in the medium by ionisation. The spallation neutrons trigger further nuclear reactions in the calorimeter medium which lead, amongst others, to the production of more neutrons.

In the end the 'fast' neutrons also lead to evaporation processes (fig. 15.8(b)) with energies in the MeV range (see the description in the next paragraph). In calorimeters containing heavy nuclei, like uranium and lead, spallation neutrons can also induce nuclear fission in which short-range nuclear fragments as well as neutrons and gammas are generated.

Evaporation. When the spallation stops a highly excited nucleus remains, which after about 10^{-18} s releases its energy by evaporation of nucleons and nuclear fragments,

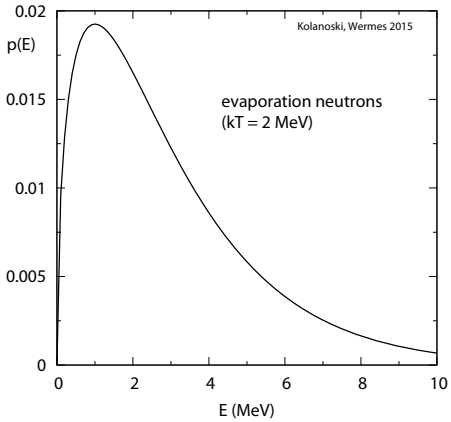


Fig. 15.10 Kinetic energy distribution of neutrons from nuclear evaporation, here approximated by a Maxwell distribution with $kT = 2\,\mathrm{MeV}$.

sometimes also by fission (fig. 15.8(b)). The energies of the evaporation products lie in the range of a few MeV, hence of fractions of the binding energy per nucleon in the nucleus. The kinetic energies roughly follow a Maxwell distribution with mean energies around 2 MeV (fig. 15.10). In this state heavy nuclei, like lead and uranium, which have high Coulomb barriers for protons and charged fragments, predominantly emit neutrons. By contrast, the neutron/proton ratio for lighter nuclei, for example for iron, is relatively balanced. The neutrons can cool down by collisions before they are captured by a nucleus. The corresponding energy deposition can be considerably delayed, as is also the gamma radiation from radioactive decays of the remaining nucleus and the fragments. The energy spent in breaking up the nuclear bindings is lost for the continuation of the process and is referred to as 'invisible energy'.

The charged nuclear fragments are strongly ionising and have a short range. Therefore, if dense 'passive' media alternate with light 'active' media, as is typical for hadronic calorimeters (*sampling calorimeters*, section 15.6.1), usually only the faster protons from spallation have a chance to cross from the passive to the active medium where they can be recorded. Neutrons can thermalise by transferring energy to charged particles in elastic collisions, which becomes more efficient at smaller target masses. Thermalised neutrons can be captured by a nucleus, thus releasing binding energy which is emitted in the form of gamma radiation. Reactions which are triggered by thermalised neutrons are delayed in the order of microseconds after the primary reaction, such that the contribution to the signal in a calorimeter depends on the signal integration time. The number of generated neutrons is strongly correlated with the binding energy loss and therefore plays an important role for the concept of 'compensation' of the undetectable binding energy (see section 15.6.3).

Electromagnetic and weak processes in a shower. The high-energy hadrons generated in the cascade interact again inelastically or decay. In dense calorimeters, as employed at accelerators, only decays of short-lived particles play an essential role, that means, particles with decay lengths well below the longitudinal dimension of the calorimeter. These are besides strong (hadronic) decays, electromagnetic decays and, less relevant for calorimeters, weak decays of heavy flavour particles. For shower evolution electromagnetic decays like those of neutral pions or η mesons are particularly relevant.

In a high-energy cascade pions are produced most frequently, of which one-third are

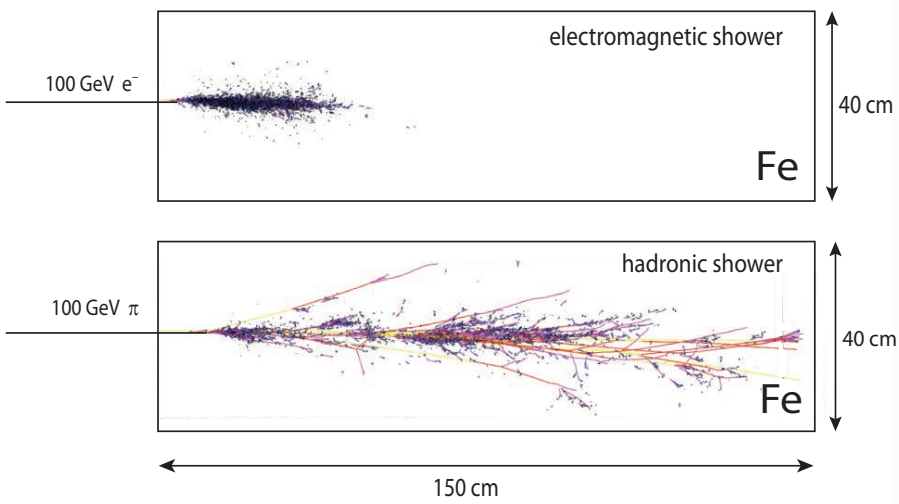


Fig. 15.11 Simulation of showers in an iron block. The picture compares an electron and a pion shower at the same energy at 100 GeV. Only ionising particles are plotted. Author: S. Menke.

neutral pions. The neutral pions decay with a short lifetime exclusively into photons and electrons, with 98.8 % into two photons:

$$\pi^0 \to \gamma\gamma \qquad (\tau \approx 10^{-16}\,\mathrm{s})\,. \tag{15.24}$$

The photons create electromagnetic sub-showers within the hadronic shower. Charged pions have a large probability of generating in turn pions via hadronic scattering, of which again one-third ends up as electromagnetic showers. These electromagnetic sub-showers essentially no longer contribute to the hadronic shower, therefore both shower types can be considered separately. The higher the initial energy the more frequently a split-off of an electromagnetic portion from the hadronic cascade can occur. In this way a substantial fraction of the energy can be deposited through electromagnetic interactions (see the following discussion of shower fluctuations).

In not very dense media the decay probabilities for unstable particles, like charged pions and kaons, could compete with the interaction probability in the medium. This plays an important role for air shower generated by cosmic radiation (chapter 16). In weak decays like $\pi^{\pm} \to \mu^{\pm}\overset{(-)}{\nu}_{\mu}$ and $K^{\pm} \to \mu^{\pm}\overset{(-)}{\nu}_{\mu}$ the neutrinos carry energy out of the detector volume, which cannot be detected. In dense calorimeters neutrinos produced by weak decays of pions, kaons and others contribute to the non-observable energy with only about 1% of the hadronically deposited energy [982].

These combinations of different processes—hadronic, electromagnetic and weak interactions—in the shower development lead to stronger shower fluctuations and in general worse energy resolutions for hadronic showers as compared to electromagnetic showers. Figure 15.11 shows the simulation of cascades initiated by an electron and a hadron in an iron block both at the same energy of 100 GeV. Obviously the electron shower is much shorter and much smoother than the hadron shower. The spots with increased density in the hadron shower indicate electromagnetic sub-showers, mostly initiated by neutral pions.

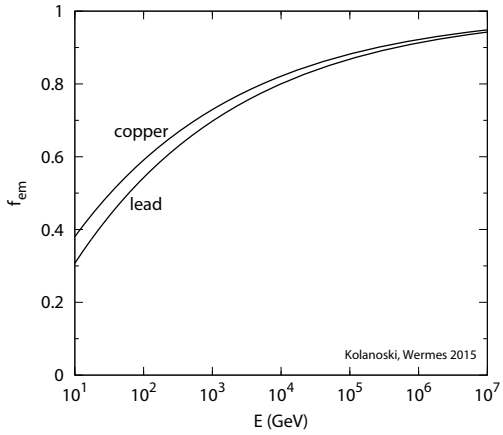


Fig. 15.12 The fraction f_{em} of the electromagnetically deposited energy in copper and lead as a function of the energy of the primary hadron. The curves have been calculated using (15.26) with the parameters $E_0 = 0.7\,\text{GeV}$ for copper, $E_0 = 1.3\,\text{GeV}$ for lead and $k = 0.82$ for both.

15.3.2 Shower components and shower fluctuations

We have already mentioned that shower fluctuations are much stronger in hadronic than in electromagnetic showers. This is caused by the splitting of the shower into different components for which the fractions f_i of the total deposited energy strongly fluctuate. The energy deposited in a block of matter consists of the following components:

$$E_{dep} = \left(f_{em} + \underbrace{f_{ion} + f_n + f_\gamma + f_B}_{f_h} \right) E\,, \tag{15.25}$$

where f_{em} is the electromagnetic portion (predominantly from π^0 decays) and $f_h = 1 - f_{em}$ is the hadronic portion; the other notations are explained further down in the text. Because in a hadronic cascade π^0 mesons are repeatedly generated which then decay into photons, the average electromagnetic energy fraction f_{em} increases with the particle multiplicity in the cascade and hence with the energy:

$$f_{em} \approx 1 - \left(\frac{E}{E_0} \right)^{k-1} . \tag{15.26}$$

The form of this equation is derived in [982] with E_0 and k to be determined as free parameters by fitting the formula to data or simulations. The parameter k is related to the particle multiplicity and is estimated to be $k \approx 0.82$; for the parameter E_0, which has the meaning of the mean energy necessary for the production of a π^0, simulations yield $E_0 = 0.7\,\text{GeV}$ for copper and $E_0 = 1.3\,\text{GeV}$ for lead [982]. Figure 15.12 shows the resulting energy dependence of f_{em} for copper and lead. At 10 GeV the electromagnetic fraction corresponds approximately to the π^0 fraction in an inelastic reaction of about 1/3, while it increases to about 70% at 1 TeV and to about 90% at 1 PeV. The highest energies in the plot are particularly relevant for the detection of cosmic radiation through the shower development in the atmosphere (see chapter 16).

The hadronic fraction in (15.25) is divided into contributions which differ in the way they are detected:

f_{ion}*:* Charged particles transfer energy to the medium through ionisation. Here one has to distinguish *relativistic* particles which are minimum-ionising and *non-rela-*

tivistic particles (slow protons and nuclear fragments) which are strongly ionising and thus have a small chance to reach the active medium in sampling calorimeters. Analysing particle reactions in photoemulsions these particles are are seen as 'thin' or as 'thick' tracks, respectively (fig. 15.9).

f_n: Neutrons transfer the energy through elastic collisions or nuclear reactions. In elastic collisions the contribution to the signal depends on the energy transfer and thus on the target mass. In hydrogen-rich active media this contribution can be relatively large (see 'compensation' in section 15.6.3). Additional neutrons can be generated in nuclear fission reactions induced by spallation neutrons. After thermalisation by elastic scattering neutrons can be captured by nuclei, which mostly release their excitation energy by gamma emission. Theses gammas are delayed in the order of microseconds because of the thermalisation process (see above). For such a delayed contribution to the signal the detection efficiency depends on the integration time of the signal.

f_γ: The photons from nuclear reactions are allocated to the 'hadronic' part of the energy because they arise from nuclear reactions and are strongly correlated with the other hadronic energies. Compared to the nuclear fragments these photons have a somewhat increased chance to reach the active medium and to deposit their energy there, predominantly through Compton scattering and the photoeffect. This component can be strongly delayed in the order of microseconds (like the gammas from neutron capture, see above).

f_B: The energy which is needed to break up a nucleus, corresponding to its binding energy, cannot be detected so that the last term in (15.25) does not contribute to the calorimeter signal. The escaping neutrinos are often included in this 'invisible energy' term.

The fraction of electromagnetically deposited energy in a hadron shower, f_{em}, fluctuates very strongly between the extremes of 0% and 100%, corresponding to reaction chains where either none or only neutral pions, respectively, are produced. These fluctuations are the largest challenge in hadron calorimetry.

On the other hand the composition of the hadronic component by itself is largely independent of the energy and particle species, as can be seen from the compilation in fig. 15.13. Between 10 and 500 GeV the fractions of energy going into ionisation ('total ionisation' is the energy deposited by charged particles), low-energy neutrons, gamma radiation and binding energy losses are surprisingly constant. However, the hadronically deposited energy is not a good measure of the total energy because the division into the hadronic and the electromagnetic parts fluctuates strongly and because in addition their mean values are energy dependent according to (15.26).

At this point two possible ways for reaching good energy resolutions with hadron calorimeters become apparent. Either one measures the division between hadronically and electromagnetically deposited energy or one attempts to make the signal response of a calorimeter the same for both components. The corresponding concepts and practical realisations will be discussed in more detail in section 15.6.

15.3.3 Characteristic size of hadronic showers

The characteristic length scale for the development of hadron showers is the *nuclear absorption length*, as introduced in (3.145):

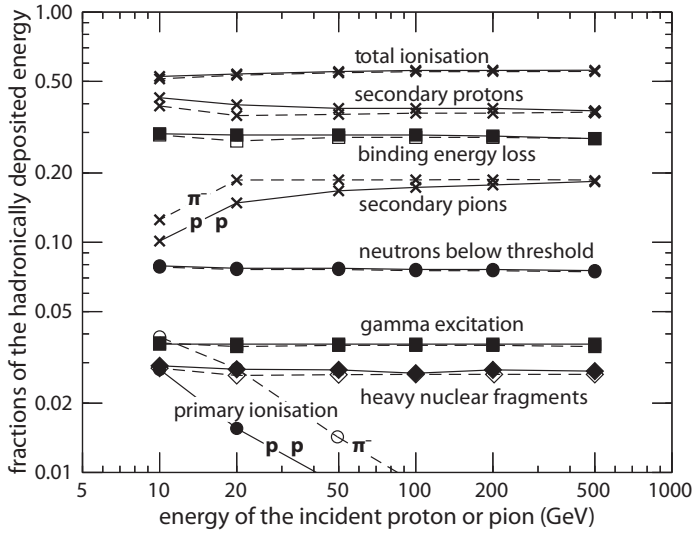


Fig. 15.13 The different fractional contributions to the hadronically deposited energy, $f_h E$, as a function of the primary energy for protons (solid line) and negative pions (dashed line) in an iron block. The simulation was carried out using the program CALOR (from [436], with kind permission of Elsevier).

Table 15.3 Atomic and nuclear properties of different materials: atomic number, mass number, density, radiation length, hadronic absorption length, their ratios and the ionisation energy loss for minimum-ionising particles (from [746, 762]).

Material	Z	A (g/mol)	ρ (g/cm^3)	X_0 (cm)	λ_a (cm)	λ_a/X_0	$dE/dx\|_{min}$ (MeV/cm)
H_2O	1, 8	18.0	1.00	36.1	83.3	2.3	1.99
air	7, 8	14.3	1.205×10^{-3}	3.0×10^4	7.5×10^4	2.5	2.19×10^{-3}
Be	4	9.0	1.85	35.3	42.1	1.2	2.95
C	6	12.0	2.21	19.3	38.8	2.0	3.85
Al	13	27.0	2.70	8.9	39.7	4.5	4.36
Fe	26	55.8	7.87	1.76	16.8	9.5	11.42
Cu	29	63.5	8.96	1.43	15.3	10.7	12.57
W	74	183.8	19.30	0.35	9.9	28.3	22.10
Pb	82	207.2	11.35	0.56	17.6	31.4	12.73
U	92	238.0	18.95	0.32	11.0	34.4	20.48

$$\lambda_a = \frac{A}{N_A \, \rho \, \sigma_{inel}} \approx 35 \, \frac{\mathrm{g}}{\mathrm{cm}^2} \, \frac{\mathcal{A}^{\frac{1}{3}}}{\rho} \, , \tag{15.27}$$

where for the inelastic cross section a $\mathcal{A}^{2/3}$ dependence was assumed (see also section 3.1). For some materials absorption lengths[2] are listed in table 15.3. The $\mathcal{A}^{1/3}$ dependence in (15.27) is not necessarily compensated by an increase of the density

[2]The data have been extracted from the *Review of Particle Properties* [746]. There the nuclear absorption length is defined as that for neutrons with a momentum of 200 GeV/c [480]. Since the energy dependences and the dependences from the hadron species are not negligible and the definitions

so that, for example, iron has a slightly shorter absorption length than lead. The size of a shower grows both in the electromagnetic and hadronic case with $\ln E$ and scales with X_0 or λ_a, respectively. The ratio of λ_a to X_0 (table 15.3) increases about proportionally to Z:

$$\frac{\lambda_a}{X_0} \approx 0.37\, Z\,. \tag{15.28}$$

Since for denser media λ_a is much larger than X_0 (table 15.3), in general hadronic calorimeters are much larger than electromagnetic ones (see the example in fig. 15.11).

For the detection of hadronic showers with good resolution it is important that the electromagnetic and the hadronic components deliver roughly the same signals for the same deposited energy (see the corresponding discussion in section 15.6). This can approximately be achieved by using passive media with small A and Z values in hadron calorimeters. With smaller Z values the ratio of the absorption length to the radiation length becomes smaller, as can be seen in table 15.3. From this point of view beryllium would be most favourable according to the table. However, the light elements have a very large absolute absorption length, which is highly unfavourable for practical applications. Therefore and for many practical reasons iron is the preferred passive medium in hadron calorimeters. However, besides the relative size of absorption and radiation length there are others aspects which can influence the approximate equality of the signal due to the electromagnetic and hadronic components. This again brings heavy elements into play, in particular uranium and lead, as will be discussed in the section 15.6.3 on 'compensating calorimeters'.

Also, for hadronic showers, as in the electromagnetic case, general properties of the shower profiles can be specified. In the following we quote empirical estimates which are valid in the energy range of some GeV up to several 100 GeV [383]. Here again the position of the shower maximum has a linear dependence on the energy (t = length in units of λ_a):

$$t_{max} \approx 0.2\, \ln(E/\mathrm{GeV}) + 0.7\,. \tag{15.29}$$

On average about 95% of the energy is contained within a longitudinal depth of

$$t_{95\%} \approx t_{max} + 2.5 \left(\frac{E}{\mathrm{GeV}}\right)^{0.13}. \tag{15.30}$$

Note however that the fluctuations about the mean values are much stronger than for electromagnetic showers.

The lateral distribution of the deposited energy is characterised by a dense kernel which is dominated by electromagnetic sub-showers and a relatively long tail which is mainly caused by neutrons. The lateral inclusion of 95% of the deposited energy can be roughly described by a cylinder with radius

$$R_{95\%} \approx \lambda_a\,. \tag{15.31}$$

More precise statements are material dependent [982].

The approximate scaling in λ_a for the longitudinal shower profile is demonstrated in fig. 15.14 [383]. For the lateral shower distribution this scaling is also obtained for the shower core within a FWHM. However, further out in the tails, as for 90% containment shown in the figure, the distribution becomes increasingly material dependent.

are not unique, somewhat different values for the absorption and interaction lengths can be found in the literature.

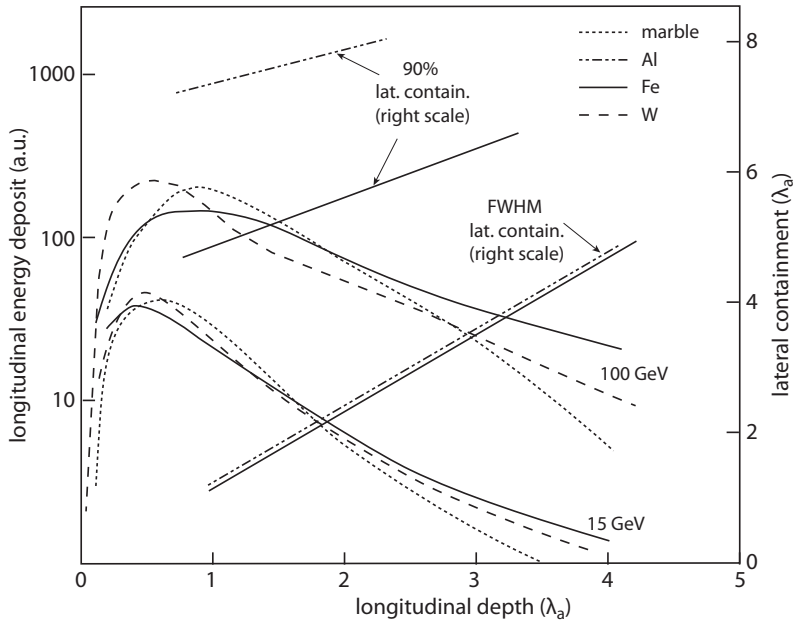


Fig. 15.14 Average longitudinal profiles of hadronic showers in different absorber materials for the primary energies 15 and 100 GeV (left scale) and lateral containment in terms of λ_a as a function of the longitudinal depth in units of λ_a for 10-GeV pions (right scale). The longitudinal profiles are plotted starting at the point of the first interaction (rather than at the entrance into the absorber) which makes the shower averages more compact with a more pronounced shower maximum. For the same primary energy the profiles are similar for different materials demonstrating the approximate scaling in λ_a. This scaling is also found for the lateral shower core defined by the FWHM of the lateral distribution at a fixed longitudinal depth. However, requiring 90% containment exhibits a strong material dependence for aluminium and iron. The plot was adapted from [383] where references to the input data can also be found.

15.3.4 Simulation of hadronic showers

The simulation of hadronic showers is more involved than for electromagnetic showers (see page 588). While electromagnetic reactions are at least in principle theoretically accessible, many hadronic reactions can only be described by employing models with parameters that have been experimentally determined. Therefore a larger variety of programs exists, some of them being optimised for specific aspects, like high or low energies, particle or nuclear physics. The program packages Geant4 [452] and FLUKA [416, 401] offer different options for adjusting the simulation to different objectives.

For the simulation of air showers, initiated by cosmic radiation, special programs have been developed, which we describe in chapter 16.

15.4 General principles for construction and operation of calorimeters

The basic requirements that a calorimeter or calorimeter system has to fulfil are optimal energy and position resolutions and usually also a complete, hermetic coverage of

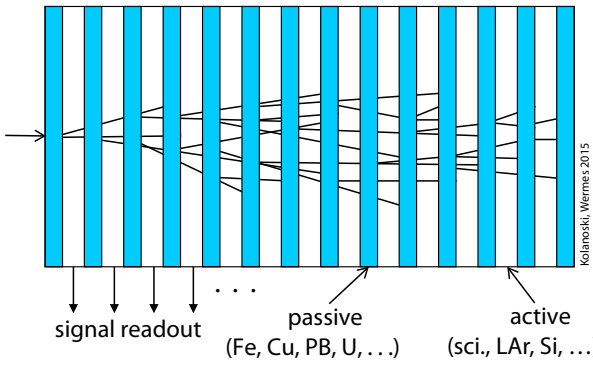


Fig. 15.15 Schematic of a sampling calorimeter, here in a sandwich arrangement.

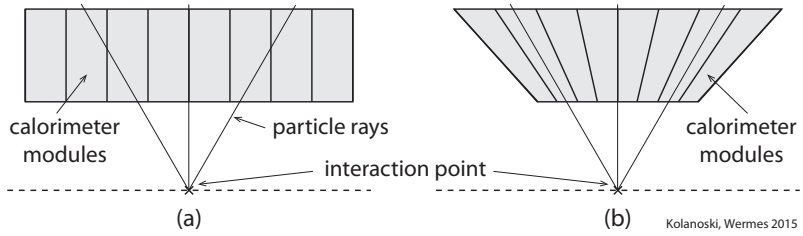


Fig. 15.16 Comparison of a non-projective (a) and a projective (b) alignment of calorimeter modules towards the interaction point.

the relevant solid angle sector. For practical realisation many other requirements and constraints supervene. Usually the requirements can be fulfilled in different ways, but in most cases only by compromising between different requirements. In the following we discuss some of the essential parameters for calorimeter construction.

15.4.1 Construction method

In a calorimeter one can in principle distinguish between the *passive medium* in which the shower develops and the *active medium* in which electronically[3] recordable signals of the shower particles are generated (e.g. as ionisation charge, scintillation or Cherenkov light). In *homogeneous calorimeters* (e.g. made of scintillating crystals or lead glass blocks) both functions are fulfilled by a single medium. In inhomogeneous, so-called *sampling calorimeters*, both functions are separated, which is mostly achieved by alternately stacking a passive and an active layer, see fig. 15.15. For reasons which we will later discuss in detail, hadron calorimeters are nearly always built as sampling calorimeters, while for electromagnetic calorimeters both types can be found.

In larger experiments calorimeters are subdivided into modules which, depending on the experiment, can be arranged as a sphere (e.g. the Crystal Ball, fig. 13.24 on page 537), as a cylinder (in the central region of most collider detectors; see examples in fig. 15.22 and 15.24) or in a plane (usually in fixed-target experiments, e.g. in fig. 15.28). Besides the mechanical partitioning owing to construction requirements there is also a segmentation of the readout. A module of a homogeneous calorimeter is usually at

[3]Here we focus on calorimeters with electronic readout as used in modern high-energy experiments. However, the optical representation of a shower by a bubble or cloud chamber (chapter 6) with converter plates (fig. 15.26) is still useful for the visualisation and the understanding of shower development.

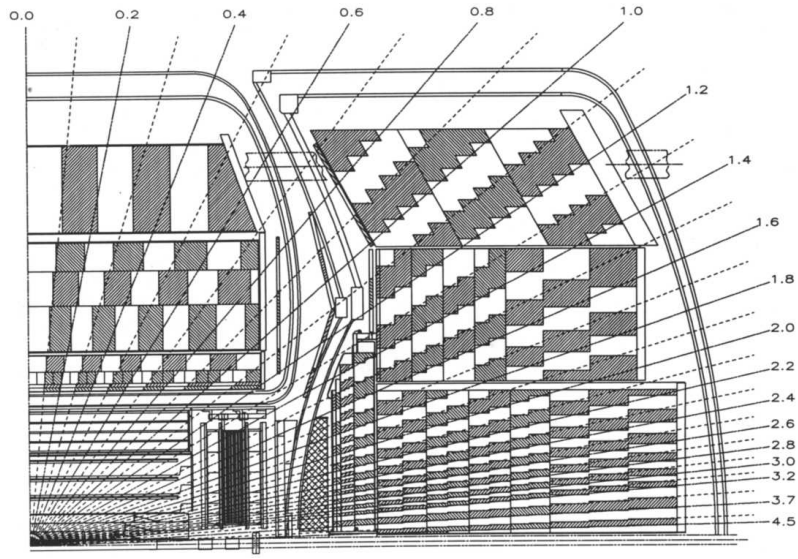


Fig. 15.17 The *trigger tower* structure of the D0 detector (from [13], with kind permission of Elsevier). The segments, drawn alternately white and grey, are electronically tied together forming an AND logic.

the same time also a readout unit. Sampling calorimeters offer more freedom to adapt the readout structures to physics requirements like resolution and particle separation.

Furthermore, one has to decide whether the arrangement should be *projective*, which means whether the modules are pointing towards the nominal interaction point (*pointing geometry*). Figure 15.16 shows drawings of non-projective and projective arrangements. The advantage of a projective arrangement lies mainly in an optimal and over the whole detector uniform position resolution for the reconstructed showers. An important practical disadvantage is the fact that many different module shapes have to be manufactured. If there is material between the modules, for example support structures or just an air gap, then a pointing geometry allows photons to pass through the calorimeter without producing a (sufficient) signal. Therefore, if hermetic detector coverage is important, often a non-projective geometry is chosen.

The signals from calorimeters are often employed for generating fast triggers. For this purpose several modules of a calorimeter are often combined to form *trigger towers* which have about the size of the showers to be recognised. As an example the trigger tower structure of the D0 detector [13] is shown in fig. 15.17.

15.4.2 Size and granularity of a calorimeter

The size of a calorimeter should be chosen such that as large a fraction of the shower energy as possible will be detected. Leakage losses degrade the resolution, as demonstrated by the examples in fig. 15.18.

For the signal readout a calorimeter is usually divided into segments, making details of the lateral and often also the longitudinal shower development visible.

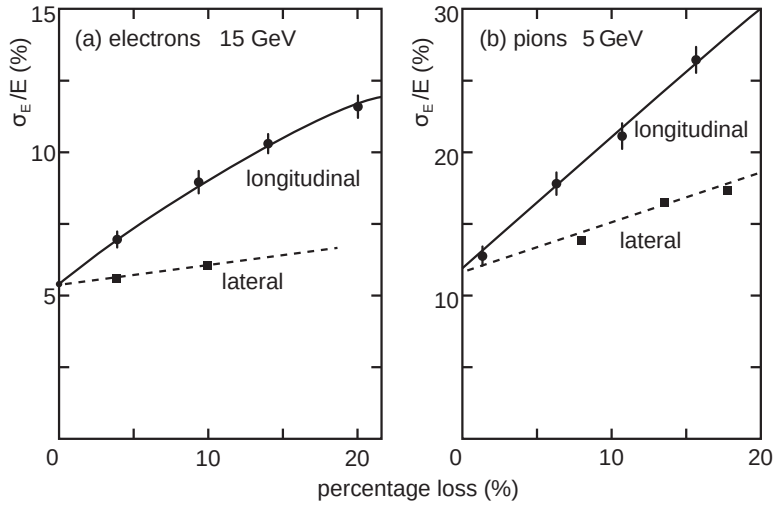


Fig. 15.18 Simulations of energy resolutions in dependence of the percentage shower loss (longitudinal and lateral) for (a) 15-GeV electrons and (b) 5-GeV pions (from [88], © The Royal Swedish Academy of Sciences, with kind permission of IOP Publishing).

15.4.2.1 Longitudinal structure

The choice of the longitudinal dimensions of a calorimeter depends on the energy region it should be optimised for. If for example the 98% containment of a shower is required the necessary detector depth $t^{98\%}$ can be estimated by means of formula (15.20). Some values for the 98% coverage are listed in table 15.1 on page 588. For example, in lead the 98% coverage is reached for 10-GeV electrons with $18.8\,X_0$ and for 100-GeV electrons with $22.2\,X_0$. For hadronic showers (15.30) is an approximate formula for the detection of 95% of the shower energy.

Longitudinal segmentation. In electromagnetic calorimeters a longitudinal segmentation of the signal readout can substantially improve the hadron–electron separation (see chapter 14) and in hadronic calorimeters a longitudinal segmentation can be exploited to improve the energy resolution (see 'software correction' in section 15.6.3).

Tail catcher. If shower losses cannot be avoided even a quite rough measurement of the loss by a so-called *tail catcher* can cause a distinct improvement of the energy. The scheme is sketched in fig. 15.19: behind the calorimeter an additional detector measures with coarse resolution the shower fraction which was not absorbed in the calorimeter. For hadron calorimeters such a measurement can often be achieved by subsequent muon detectors or the instrumented iron yoke of a magnet. Even with such a coarse measurement of the leakage losses the energy measurement by the main calorimeter can be corrected so that the energy resolution significantly improves. The effects of leakage losses and of potential corrections on the resolution are discussed for electromagnetic calorimeters in section 15.5.5 (page 633) and for hadronic calorimeters in section 15.6.5 (page 650).

Presampler. In front of the calorimeters detectors for tracking and possibly particle identification are often installed. Although these detectors are in principle conceived to be 'non-destructive', made of as little material as possible, the actual material thick-

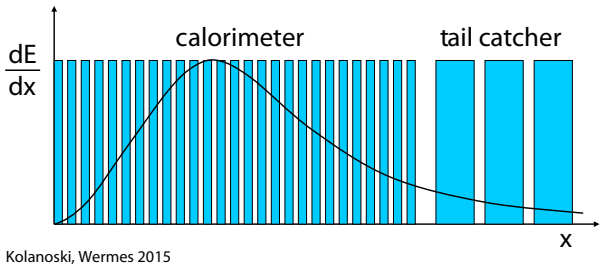


Fig. 15.19 Measurement of the shower loss by means of a subsequent coarser detector (*tail catcher*).

ness in front of the calorimeters can add up to some radiation lengths (e.g. in the LHC detectors). A thickness of the order of radiation lengths leads to a high probability that a shower already starts before the calorimeter, causing a degradation of the energy resolution, in particular for electromagnetic calorimeters. This can partially be remedied by installing a so-called *presampler*, which may just determine if the shower has already started (important for photons) or may already provide a rough energy measurement of an early shower. A presampler can simply be a first calorimeter layer with separate readout (e.g. the ATLAS EM calorimeter, described in section 15.5.3.2 on page 621) or a dedicated detector, for example a layer of scintillators, silicon detectors or wire chambers.

15.4.2.2 Lateral structure

For electromagnetic showers the lateral energy losses can be estimated by the lateral extent of the calorimeter in units of the Molière radius (see eq. (15.21) and table 15.2). According to fig. 15.18(a) 1–2 Molière radii are already sufficient in order not to downgrade the resolution by much. For hadronic showers the 95% containment radius is estimated to be about one absorption length according to (15.31).

A laterally segmented signal readout is important for the separability of individual showers and the position resolution of the shower axis (section 15.4.5). In general the algorithms for the separation of two showers require that a minimum of the deposited energy is visible in at least one readout segment between the showers.

15.4.3 Energy resolution

The energy resolution is the most important quality criterion of a calorimeter. In most calorimeters the energy resolution is determined by the stochastic fluctuations of the number of charged particles contributing to the signal. If their number N_S is proportional to the primary energy, $N_S \propto E$, with a standard deviation $\sqrt{N_S}$ according to Poisson statistics then the resolution is

$$\frac{\sigma_E}{E} \propto \frac{\sqrt{N_S}}{N_S} = \frac{1}{\sqrt{N_S}} \propto \frac{1}{\sqrt{E}}\,. \tag{15.32}$$

However, in a realistic calorimeter further terms arise in addition to the stochastic term, in general with different energy dependences. Employing a relatively general ansatz three contributions are quadratically summed:

$$\frac{\sigma_E}{E} = \sqrt{\frac{a^2}{E} + \frac{b^2}{E^2} + c^2} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\,. \tag{15.33}$$

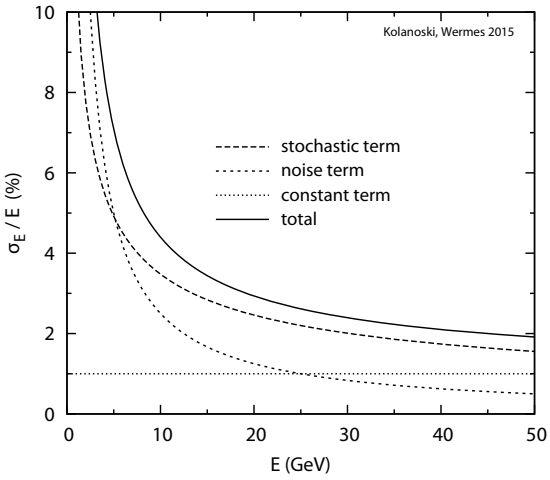


Fig. 15.20 Example for the energy dependence of the energy resolution of a calorimeter with the individual contributions of the resolution terms as given in (15.33), with $a = 0.11/\sqrt{\text{GeV}}$, $b = 250\,\text{MeV}$, $c = 0.01$.

The right hand side presents a common notation for the quadratic addition. The dominant sources for the different contributions, parametrised by a, b, c, are (fig. 15.20):

a stochastic fluctuations of the shower development being particularly strong for sampling calorimeters due to incomplete sampling;

b electronic noise which contributes a $1/E$ term to the relative resolution because the absolute fluctuations are energy independent;

c mechanical and electronic imperfections, fluctuations in the leakage losses, intercalibration errors and others with absolute contributions which rise with energy.

The first two terms become smaller with increasing energy until finally the constant term limits the resolution. At low energies the limitation is given by the noise term, which also determines the threshold for the smallest measurable energy. The coefficient b in (15.33) is called the 'equivalent noise energy', corresponding to the energy of a particle which yields the same electronic signal as the standard deviation of the noise background.

Because of the characteristic differences in the behaviour of electromagnetic and hadronic calorimeters we will discuss their energy resolutions separately in sections 15.5.5 and 15.6.5.

15.4.4 Linearity

Besides the energy resolution the *linearity* of the calorimeter, that is, the degree to which the measured signals depend linearly on the primary energy, is an important characteristics of a calorimeter. The linearity can be curtailed by the detector (e.g. by the leakage losses, typically increasing with $\ln E$, see (15.20) and (15.30)) or by the electronics (e.g. through saturation of large signals). Often the linearity of the electronics can be determined by means of an appropriate test pulse system (see also section 17.7.1 on the linearity of the signal digitisation step). However, in order to determine the linearity of the full system, the calorimeter response should be measured in a test beam of known particle species and known energy.

15.4.5 Position and direction resolution

The reconstruction of the kinematics of a showering particle (or jet) also requires the determination of the particle's direction in addition to the energy. This can be achieved by measuring at least two positions, for example the point where the particle enters the calorimeter and the interaction point.

If the shower is fully absorbed in one readout segment, the position resolution is determined by the lateral dimensions of the segment. For example, if the segment's cross section is quadratic with a side length a then the variance of a reconstructed coordinate (parallel to a side) is $\sigma^2 = a^2/12$ (corresponding to the variance of a uniform distribution over the length a, see eq. (E.4) in appendix E). The position resolution can be substantially improved if the deposited shower energy is distributed over several neighbouring segments. Optimal resolutions are obtained if the lateral dimension of a single segment is similar to or smaller than a Molière radius for electromagnetic calorimeters or an absorption length for hadronic calorimeters. The shower position can be determined by fitting a shower template, that is, a theoretical lateral shower profile (taking the shower direction into account, as well), to the energies measured in neighbouring segments (see also appendix E). In this way position resolutions of some millimetres for electromagnetic calorimeters and some centimetres for hadronic calorimeters can be reached.

Connecting the reconstructed entrance point of a showering particle or jet with the point at which the particle was produced the direction can be determined and thus, together with the measured energy, the particle kinematics can be reconstructed. As an example we consider a cylindrical calorimeter with the axis aligned parallel to a beam line (z-axis), as typical for collider experiments, and with a radial distance d of the calorimeter's entry surface from the beam. The shower reconstruction delivers the particle's entrance point at the fixed radius d with the z coordinate z_{cal} and an azimuthal coordinate (here the latter is not used). Let the resolution of z_{cal} be σ_z^{cal} and the resolution of the corresponding interaction point coordinate z_{ip} be σ_z^{ip}. Since d is fixed, the radial uncertainty of the shower position is included in the uncertainty of z_{cal} (through the shower reconstruction). The polar angle of the particle with respect to the z-axis is θ. Without magnetic field, the particle slope, $\tan\theta$, and its resolution are given by

$$\tan\theta = \frac{d}{z_{cal} - z_{ip}} \qquad \text{and} \qquad \sigma_{\tan\theta} = \tan\theta \, \frac{\sqrt{\left(\sigma_z^{cal}\right)^2 + \left(\sigma_z^{ip}\right)^2}}{z_{cal} - z_{ip}} \,. \tag{15.34}$$

15.4.6 Signal collection and time resolution

Depending on the chosen readout method calorimeters can have very different signal collection times and time resolutions. The signal collection times range from some nanoseconds for organic scintillators, which are often employed in sampling calorimeters, to microseconds for inorganic crystals (such as NaI or CsI) and liquid argon. A good energy resolution requires sufficiently long collection times in order to compensate fluctuations between early and late signal contributions and/or to suppress noise. In the so-called 'compensating hadron calorimeters' (section 15.6.3) the collection time of the electronics also determines the signal contributions from nuclear neutrons.

At high event rates a good time resolution is needed for the assignment of signals to specific events, to a beam crossing in colliders or to a beam bunch at fixed targets.

In many cases a good time resolution has to be already available at the trigger level (section 18.3). For fast triggering the signals are often electronically processed on a separate path in order to achieve a good time resolution by pulse shaping (section 17.3), however, usually not without loss in energy resolution on such a fast trigger path. In this way time resolutions can be reached which are orders of magnitude better than the collection times. For example, for liquid argon calorimeters with collection times of about 1 µs, time resolutions of about 10 ns are reachable.

15.4.7 Calibration

The design of a calorimeter should always take into account how the calorimeter can be calibrated. This is the more important as in practice an absolute energy scale cannot be calculated or determined by simulations, but one has to rely on experimental calibration procedures.

According to the application field and the technical implementation several different procedures exist which at best could be combined to investigate systematic uncertainties. Basically, one can distinguish between procedures using test beams and those exploiting kinematical constraints in the reconstruction of particle processes. Test beams of high-energy electrons, muons or hadrons are provided by accelerators. For calorimeters with a low noise level, especially crystal calorimeters, gamma lines of radioactive sources (see appendix A.2) are also used. More details, in particular also on the use of kinematical constraints, are given in the sections 15.5.4 and 15.6.4 for electromagnetic and hadronic calorimeters, respectively.

15.4.8 Radiation hardness

In experiments that are exposed to very high energies and intensities, as at the LHC (see table 2.2 on page 13), a high radiation load can prevail even in the regions of the calorimeters. Therefore in the developments for the LHC experiments special care was taken to assure *radiation hardness* of the electronics, the used materials and the construction methods. In particular worth mentioning in this context is the development of radiation hard scintillating crystals for the electromagnetic calorimeter of the CMS experiment [298] (see the description in section 15.5.2.2).

15.5 Electromagnetic calorimeters

15.5.1 Overview

In this section we discuss examples of electromagnetic calorimeters of various types following different construction principles according to the requirements of the respective experiment. Usually a certain requirement has its optimal solution (see table 15.4); however, the combination of different, maybe contradictory, requirements may demand compromises.

Characteristic properties of electromagnetic calorimeters of different experiments are compared in table 15.5. Although the best energy resolutions are reached with homogeneous calorimeters, often the decision is taken in favour of sampling calorimeters because of overriding requirements, like high granularity for electron–hadron separation, jet resolution, cluster separation and position resolution together with the cost factor.

Table 15.4 Typical requirements for electromagnetic calorimeters and the respective technical solutions which best meet the requirements.

Requirement	Optimal solution
energy resolution	homogeneous calorimeter, crystals
position and direction resolution	high lateral segmentation
electron–hadron separation	longitudinal segmentation
hermetic coverage	dense module assembly, little support material, few supply lines to the detectors in front of the calorimeter
energy resolution at high energies	small constant term, minimisation of mechanical tolerances, small intercalibration error, stable calibration (through continuous monitoring)
equal energy scale for electrons and hadrons ($e/h \approx 1$)	compensating calorimeter (section 15.6), uniform technology for both the electromagnetic and hadronic calorimeters
jet resolution	granularity, adjusted combination of the electromagnetic and hadronic calorimeters
linearity	sufficient depth of the calorimeter, adequate dynamic range of the electronic signal
absolute calibration	test beam calibration combined with in-situ calibration
low costs	small number of electronic channels, low-priced materials, small volume (in general demanding to limit the dimensions of the inner tracking detectors)

15.5.2 Homogeneous calorimeters

15.5.2.1 Detector materials and signal readout

The best energy resolutions for electromagnetic showers are reached with homogeneous calorimeters (fig. 15.21) which use either scintillation or Cherenkov light (seldom ionisation charge) as signal:

– Scintillators: suitable are inorganic crystals like NaI(Tl), CsI(Tl), BGO or $PbWO_4$ (table 13.3 on page 520), whose properties were already discussed in section 13.3. Typical resolutions are (see also section 13.5.2.2):

$$\frac{\sigma_E}{E} \approx \frac{(2\text{–}5)\%}{\sqrt[4]{E/\mathrm{GeV}}}\,.$$

Table 15.5 Parameters of some selected electromagnetic calorimeters. The numerical values are taken from the respective references of the detectors and from the reviews in [247], [982] and [746]. The specifications for 'distance' and 'front surface' refer to the region of the shortest (perpendicular) distance from the interaction point. The values for X_0 and R_M are effective values, averaged over both active and passive layers. 'LAr' and 'LKr' mean liquid argon and liquid krypton, respectively, and 'PWC' means proportional wire-chamber. The parameters a, b, c of the energy resolutions are defined by (15.33).

Type	X_0	R_M	Distance	Cell size at front surface	Thickness/X_0		Resolutions (E in GeV)				Experiment
					passive layer	total	$a/\sqrt{E}$ (%),	b (MeV)	c (%)	σ_θ (mrad)	
	(cm)	(cm)	(cm)	(cm^2)							
homogeneous calorimeters											
NaI(Tl)	2.59	4.8	25.4	12.9		15.7	$2.8/\sqrt[4]{E}$	≈ 0.05		26–35	C. Ball [749]
CsI(Tl)	1.85	3.5	92	4.7×4.7		16	$2.3/\sqrt[4]{E}$	≈ 0.15	1.4	$4.2/\sqrt{E}$	BaBar [128,840]
BGO	1.12	2.3	50	2×2		22	$\approx 2/\sqrt{E}$		0.7	≈ 10	L3 [50,253]
Pb glass	2.54	3.5	245	10×10		25	$6.3/\sqrt{E}$	11	0.2	4.5	OPAL [63]
$PbWO_4$	0.89	2.0	130	2.2×2.2		25.8	$2.8/\sqrt{E}$	120	0.3	≈ 0.7	CMS [298]
LKr	4.7	5.9	≈ 100 m	2.0×2.0		27	$3.2/\sqrt{E}$	90	0.42	0.001	NA 48 [954]
sampling calorimeters											
Pb/sci (sandwich)	3.2	5.0	230	10×10	0.18	12.5	$6.5/\sqrt{E}$	<10	7.2	$6.5/\sqrt{E}$	ARGUS [73]
Pb/LAr	1.1	2.66	90	10–100	0.42	20–30	$11/\sqrt{E}$	150	0.6	$\approx 15/\sqrt{E}$ *	H1 [98,99]
Pb/sci (shashlik)	1.7	4.15	1350	5.59×5.59	0.54	20	$11.8/\sqrt{E}$		1.4	$1.0/\sqrt{E} \oplus 0.2$	HERA-B [135]
Pb/sci (spaghetti)	0.9	2.55	150	4.05×4.05		28	$7.1/\sqrt{E}$		1.0	≈ 1 at 30 GeV	H1 [111]
Pb/LAr	≈ 2	≈ 4.1	150	14.7×0.47	≈ 0.4	22–24	$10/\sqrt{E}$	190	0.5–0.7	$\approx 1/\sqrt{E}$	ATLAS [4]
U/sci	0.56	1.66	120	115–200	1.0	25	$18/\sqrt{E}$			$\approx 40/\sqrt{E}$ *	ZEUS [331]
Pb/gas (PWC)	≈ 1.85	4.65	185	3×3	0.36	22	$18/\sqrt{E}$		0.9	$3.7/\sqrt{E}$	ALEPH [346]

* Estimated using data on position resolutions and (15.34).

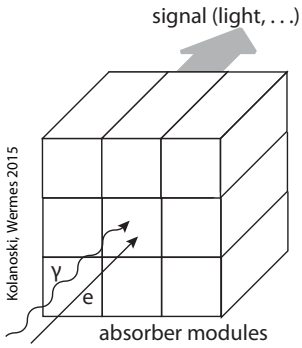


Fig. 15.21 Schematic representation of a homogeneous calorimeter. Incoming radiation generates light which is detected by suitable devices (photomultipliers, photodiodes, APDs; see section 13.4).

– Liquid noble gases with high Z (krypton, xenon): The small radiation lengths and Molière radii of liquid krypton and xenon make these noble gases interesting for calorimetry, see table 15.6. While liquid argon is quite common as active medium but less suitable as passive medium because of its long radiation length, liquid krypton and xenon can serve both as converters and as active media. In both gases the deposited energy is converted to scintillation light as well as to free charges. The best resolutions are obtained if both components are measured (see also the applications for the detection of dark matter in section 16.7). For practical reasons, however, often only one component is detected. Charges can be collected on relatively thin electrodes such that little dead material is introduced in the otherwise homogeneous detector. In this way a shower sampling with high granularity and hence high position resolution is possible. Therefore, such a 'quasi-homogeneous' calorimeter combines the advantages of a homogeneous calorimeter with those of a sampling calorimeter. Typical resolutions are

$$\frac{\sigma_E}{E} \approx \frac{(3\text{–}5)\%}{\sqrt{E/\mathrm{GeV}}} \,.$$

– Cherenkov detectors: Absorbers which emit Cherenkov radiation (chapter 11) are employed in accelerator experiments, for example lead glass or PbF_2, or in detectors for cosmic radiation, mostly water or ice. Typical resolutions for lead glass are

$$\frac{\sigma_E}{E} \approx \frac{(5\text{–}10)\%}{\sqrt{E/\mathrm{GeV}}} \,.$$

The characteristic properties of homogeneous calorimeters of some selected experiments are listed in table 15.5.

Crystal and lead glass modules are densely stacked, as depicted schematically in fig. 15.21. The light is collected at the backside of a module and converted to electronic signals by means of photomultipliers or photodiodes (section 13.4). The better energy resolution obtained with scintillating crystals as compared to lead glass arises from the higher light yield of crystals: for example 4×10^4 photons per MeV in the CsI of the BaBar experiment [128] as compared to 60 photons per MeV in lead glass calorimeter of the OPAL experiment [63].

Table 15.6 Properties of liquid noble gases which are used for calorimetry [762].

	Z	ρ (g/cm^3)	X_0 (cm)	R_M (cm)	E_c (MeV)
Ar	18	1.4	14.0	9.2	32.8
Kr	36	2.4	4.7	6.1	17.0
Xe	54	3.0	2.9	5.5	11.3

15.5.2.2 Examples for homogeneous calorimeters

In the following we present a small selection of homogeneous calorimeters which are exemplary for a detector material or a construction type. Characteristic properties of such calorimeters are summarised in table 15.5.

Crystal Ball. The NaI(Tl) calorimeter Crystal Ball (fig. 13.24), which was operated at the electron–positron collider SPEAR in Stanford starting in 1978, was a milestone for the deployment of inorganic crystals in particle physics experiments. As such the Crystal Ball is also described in the chapter on scintillators in section 13.5.2 on page 531. It was the first crystal calorimeter at a storage ring with nearly full solid angle coverage, being built for the detection and precise measurement of the photon transitions in the charmonium system with till then unprecedented energy resolution. Operated without magnetic field the detector was specialised on the observation of photon lines in inclusively measured photon spectra and the reconstruction of final states containing mainly electromagnetically showering particles (photons, electrons and positrons), see fig. 13.24(b).

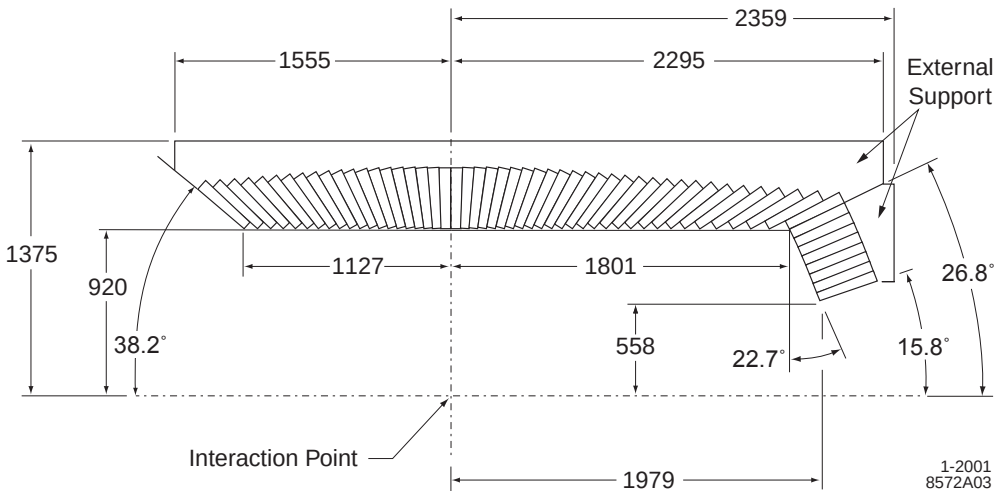
With the development of CsI and other scintillating crystals, which are less hygroscopic than NaI while offering similar energy resolutions at smaller radiation lengths and Molière radii, NaI has lost its importance in particle physics.

The CsI(Tl) calorimeter of the BaBar detector. The BaBar detector was operated at the electron–positron collider PEP-II starting in 1998 [128]. The experiment was built for the discovery and investigation of the violation of matter–antimatter symmetry (CP violation) in bottom meson systems. This goal required a calorimeter optimised for the measurement of photons and electrons in the energy range between 20 MeV and 9 GeV with high efficiency as well as high energy and angle resolution [840]. To cope with these requirements a CsI(Tl) calorimeter was chosen, consisting of 6580 crystals, each read out by two photodiodes.

Figure 15.22 shows (a) a photograph of the installation of the CsI(Tl) calorimeter together with (b) a half cross section in a plane containing the beam axis. Since the PEP-II accelerator operates with asymmetric beam energies, the detector is also not symmetric in the polar angular coverage, featuring a better coverage in the direction of the high-energy beam. The crystal arrangement is 'projective' in the azimuth angle, meaning that in the projection onto the plane perpendicular to the beam all crystals are aligned towards the nominal interaction point (*pointing geometry*). In contrast, for the polar angle the geometry is not exactly projective, with the crystal alignment deviating from the direction to the interaction point by 14–45 mrad. The advantages and disadvantages of projective and non-projective arrangements have been discussed in section 15.4.1.

CsI(Tl) calorimeters are preferentially used for meson spectroscopy at not too high energies, in particular because of their very good energy resolution (e.g. for detectors at so-called 'factories' for hadrons with charm or bottom flavours and tau leptons).

(a)

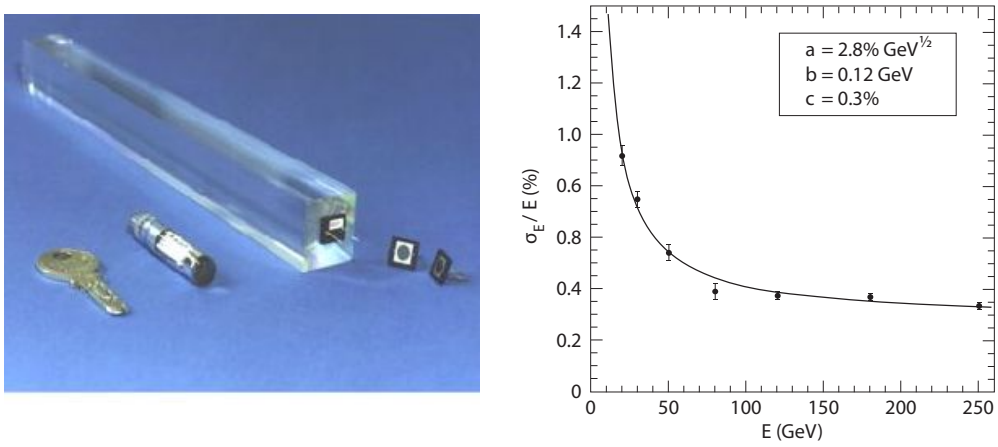


(b)

Fig. 15.22 The electromagnetic CsI(Tl) calorimeter of the BaBar detector. (a) Installation of the detector. Source: SLAC, BaBar experiment, retrieved from [574]. (b) Stacking scheme of the crystals; dimensions in mm (from [128], with kind permission of Elsevier).

The $PbWO_4$ calorimeter of the CMS detector. The CMS detector at the LHC [298] (fig. 2.11) features an electromagnetic calorimeter which consists of pulled monocrystals of the very radiation resistant material $PbWO_4$ [102]. For an experiment which is operated at TeV energies the choice of a crystal calorimeter is somewhat unexpected. The decision for a homogeneous calorimeter was taken in order to achieve the best mass resolution for the two-photon decay of a light Higgs boson. The scintillating crystals available at the time of planning the detector did not offer the radiation hardness necessary for operation at the LHC, therefore a lot of effort was invested in the development of radiation hard crystals.

Since $PbWO_4$ crystals deliver by far less scintillation light per deposited energy as compared, for example, to CsI crystals (table 13.3), a readout sensor with integrated signal amplification is preferred to a non-amplifying photodiode (fig. 15.23(a)). In CMS



(a) (b)

Fig. 15.23 (a) $PbWO_4$ crystal of the CMS detector with avalanche photodiodes (*barrel*) and to the left of the crystal a vacuum phototriode (*endcaps*). (b) Energy resolution of the electromagnetic calorimeter of the CMS detector [298]. The measurement was performed employing a matrix of 3×3 crystals with the beam aligned onto the central crystal. The function (15.33), $\sigma_E/E = a/\sqrt{E} \oplus b/E \oplus c$, was fitted to the measured resolution values. The fitted value for the stochastic term a, the noise term b and the constant term c are displayed in the figure. Source: CERN/CMS Collaboration.

the central region crystals are equipped with avalanche photodiodes (APDs) and in the endcap region with vacuum phototriodes (VPT) [173] (see section 13.4, pages 420 and 424).

A measurement of the energy resolution is shown in fig. 15.23(b). For the (3×3) matrix used in this study the equivalent noise energy (see section 15.4.3) was determined to be 120 MeV, while about 40 MeV is obtained for a single crystal [298]. For comparison, the CsI(Tl) crystals of the BaBar experiment with PIN diode readout feature a corresponding noise energy of about 250 keV [128]. However, comparing these calorimeters one has to take into account that the two experiments operate in completely different energy regimes.

The lead glass calorimeter of the OPAL detector. Figure 15.24 shows the electromagnetic calorimeter of the OPAL detector [63] which took data at the LEP storage ring until 2001. The calorimeter is constructed from 9400 blocks of lead glass ($PbO + SiO_2$). In lead glass (table 11.1) electrons and positrons generate Cherenkov light (threshold $\beta_{th} < 1/n = 1/1.46 = 0.685$, corresponding to an electron energy of 0.7 MeV). The light yield per deposited energy is smaller than for typical scintillating crystals by about two orders of magnitude (table 13.3) which, however, is sufficient for the detection of electromagnetically showering particles in the energy range above some 10 MeV. In OPAL a resolution of about 6% was reached at 1 GeV, corresponding to about 300 Cherenkov photons per GeV. Further properties of the calorimeter can be found in table 15.5.

The liquid argon calorimeter of the NA48 experiment. As an example for a homogeneous liquid calorimeter we present the liquid krypton calorimeter of the NA48 experiment (adapted from [954]). This experiment measured the direct CP violation

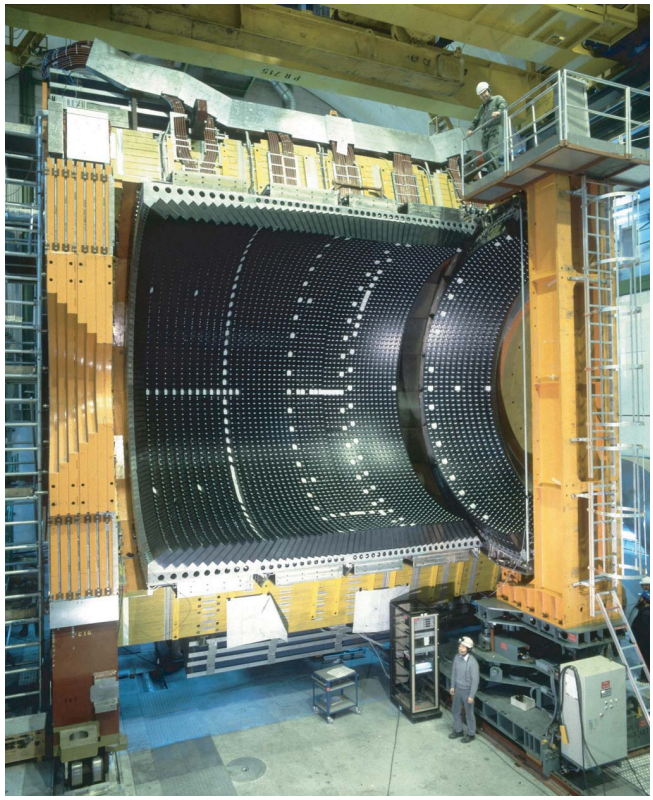


Fig. 15.24 The lead glass calorimeter of the LEP detector OPAL. The detector has been opened so that one half out of the total of 9400 crystal blocks, each one of size 10×10 cm$^2 \times 24\, X_0$, becomes visible. Source: CERN/OPAL.

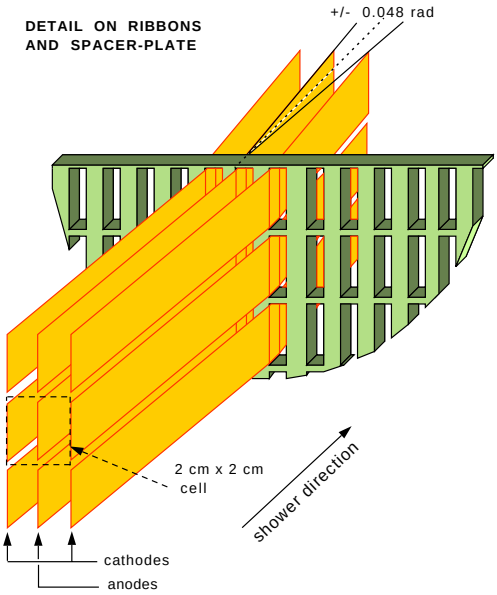


Fig. 15.25 Detail of the liquid krypton calorimeter of the NA48 experiment [954]: ribbon structure of the electrodes. The ribbons extend into the shower direction fixed to a zigzag line by spacer structures (in the drawing only one of these structures is shown). The charges from two adjacent drift cells are collected on one anode ribbon.

in decays of neutral kaons by comparing the K_L and K_S decay rates into $\pi^0\pi^0$ with those into $\pi^+\pi^-$ [393]. The suppression of the background of the $\pi^0\pi^0$ decay with four photons in the final state impose very high requirements on the energy and position resolutions of the electromagnetic calorimeter.

The calorimeter consists of about $10\,\mathrm{m}^3$ liquid krypton with depth in the shower direction of 127 cm, corresponding to about 27 radiation lengths. The ionisation charge generated by the shower particles is collected on thin electrode ribbons.[4] Figure 15.25 shows the arrangement of the electrodes: 40 µm thin, 1.8 cm wide copper–beryllium ribbons are stretched in the longitudinal direction (direction of shower development) in vertical planes with about 1 cm horizontal distance and a 2-cm vertical grid. Every 21 cm (see figure) the ribbons are fixed by spacers which at the same time keep them on a zigzag line with ±48-mrad kinks in order to avoid a shower development along an electrode. The vertical planes are alternately connected to high voltage (about 3000 V at the anode) and to ground (cathode). In the direction transverse to the shower development the ribbons form 2 cm × 2 cm readout cells which extend over the whole calorimeter depth (no longitudinal sampling). The calorimeter has a total of 13 212 readout channels. The maximal drift time is about 3.2 µs but the signals have a width of only 70 ns after pulse shaping.

The properties of this calorimeter are listed in table 15.5. In particular, the following resolutions have been reached:

- position resolution better than 1 mm above 25 GeV;
- time resolution per photon 500 ps and 250 ps for a $\pi^0\pi^0 \to 4\,\gamma$ event;
- energy resolution better than 1% above 20 GeV with a constant term of less than 0.5%.

15.5.3 Sampling calorimeters

In *sampling calorimeters* the functions of shower development and of signal generation are separated into a medium with high Z (passive medium) and a medium with rather low Z (active medium), respectively. While the shower evolves it is continuously sampled by the low-Z medium. In general, advantages of sampling calorimeters over homogeneous calorimeters are lower costs and the principally better possibility to measure the longitudinal shower shape (e.g. for electron–hadron separation, section 14.4). In contrast, better energy resolutions are reached with homogeneous calorimeters.

15.5.3.1 Technologies

Typical passive materials are lead, tungsten and uranium. The active layers are mostly designed as scintillation detectors, ionisation chambers (ionisation medium: mostly liquids or solids, for example liquid argon or silicon) or proportional chambers. Bubble and cloud chambers are nearly exclusively used for educational demonstrations of shower formation (fig. 15.26).

In the following we present some typical structural forms of sampling calorimeters as depicted in fig. 15.27:

- Sandwich calorimeter: This construction type features converter plates with high Z alternating with active layers stacked in the shower direction. The cloud chamber

[4] Because of the electrodes in the passive medium this calorimeter type is also called quasi-homogeneous.

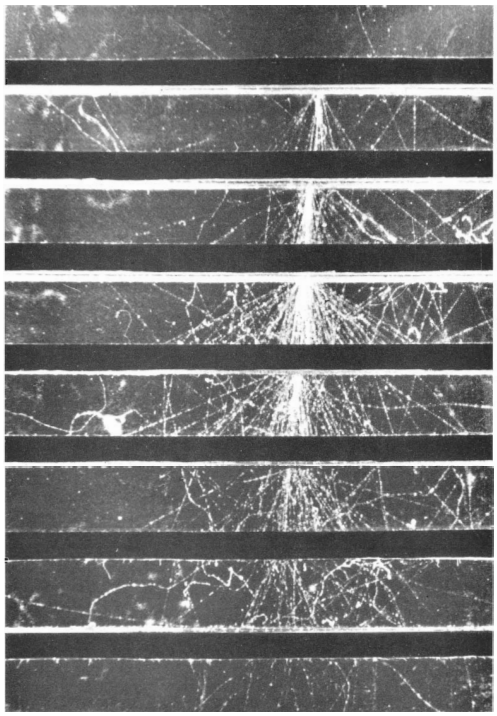


Fig. 15.26 Development of an electromagnetic shower, initiated by a photon with an energy of about 4 GeV in a cloud chamber which is equipped with converter plates consisting of 1.3 cm thick lead (from [827], with kind permission of Elsevier). The picture shows not only the high ionisation density in the shower core where single tracks can no longer be resolved due to the strong forward bundling, but also the relatively small number of tracks from low-energy electrons and positrons which diffuse away from the shower axis due to the strong Coulomb scattering.

photograph in fig. 15.26 shows an electromagnetic shower. The shower develops in lead plates which are arranged according to the sandwich principle. All electrons and positrons which reach the active material contribute to the signal. Prevalent readout methods for sandwich calorimeters are:

- Scintillators with readout via light guides (fig. 15.27(a)). A disadvantage is that the light guides, sticking out laterally, prevent dense packaging of several modules.
- Scintillators with readout via wavelength shifters. This design allows for dense packing. Typical combinations are 2–5 mm Pb + 3–5 mm scintillator. The wavelength shifters are formed as plates (fig. 15.27(b)) or as fibres (fig. 15.27(f)).
- The ionisation charge can also be measured employing the ionisation chamber principle (i.e. without gas amplification, fig. 15.27(c)). In order to obtain a sufficient amount of charge liquids are used instead of gases, like liquid argon (LAr). The argon is kept liquid at about −185 °C in a cryostat. The advantage of liquid argon calorimeters is primarily the possibility of reaching a high position resolution by means of segmentation of the active planes (pad structure). Disadvantages are the additional material due to the cryostat walls and the considerable effort necessary to keep the argon free of electronegative contaminants, in particular free of oxygen. In order to circumvent the drawback of thick cryostat walls liquids are also employed at room temperature [533]. An example is the hadron calorimeter of the air shower experiment KASCADE using tetramethylsilane (TMS) as readout medium [374,790]. Because of the even more demanding purification requirements and lower signal yield as compared to liquid argon, 'warm liquids' have not gained much currency as a readout medium.
- Wire chambers with gas amplification like MWPC or streamer tubes

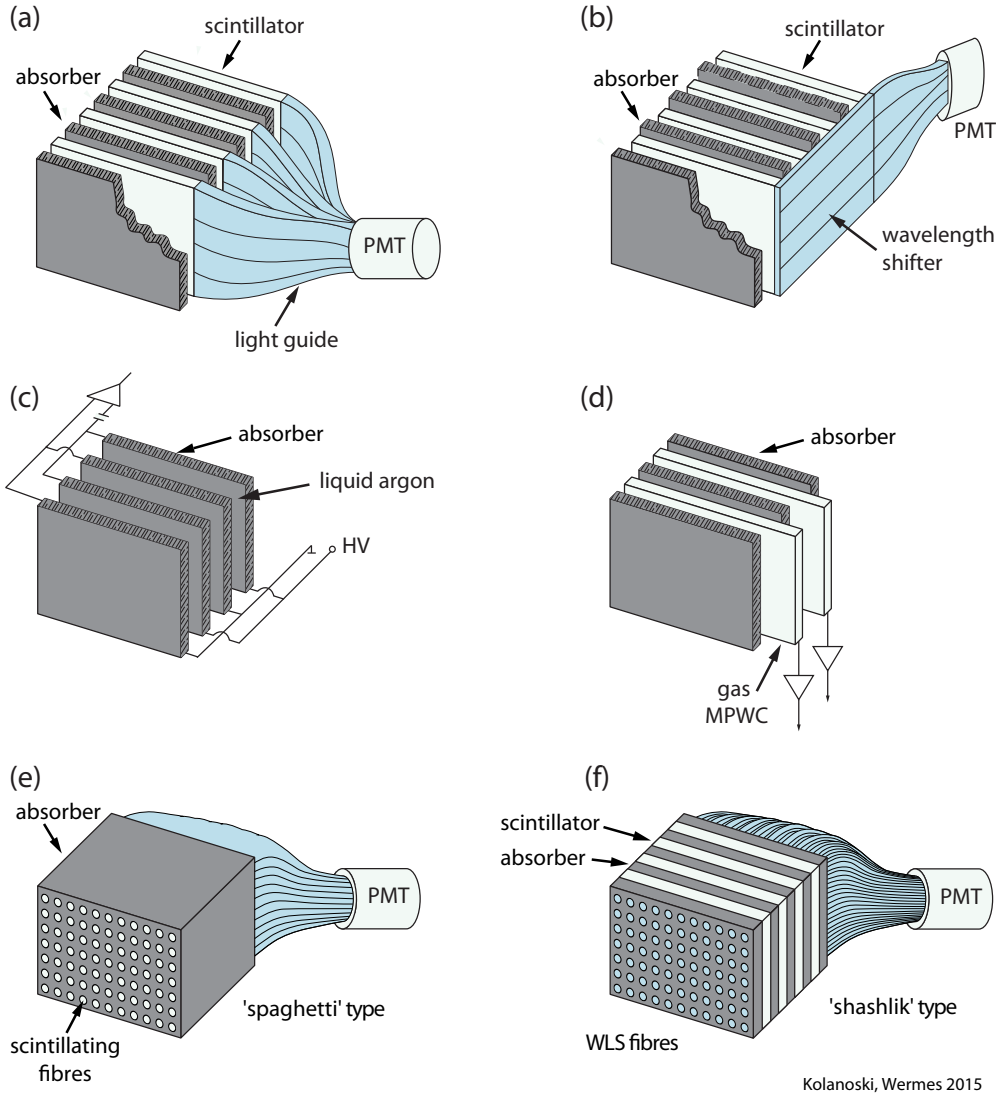


Fig. 15.27 Examples for readout methods of sampling calorimeters (adapted from [386]).

(a) Sandwich configuration with scintillator plates as active layers and readout by light guides perpendicular to the shower axis.
(b) As (a) but with wavelength shifter readout redirecting the light into the shower direction.
(c) Sandwich configuration with ionisation chambers as active layers.
(d) As (c) but with gaseous multiwire proportional chambers as active layers.
(e) Scintillating fibres as active medium strung in the shower direction through the absorber ('spaghetti' type). The light of several fibres can be collected by one light guide for the light transport to the PMT.
(f) Sandwich configuration as in (b) but read out by wavelength shifter fibres which go through all layers ('shashlik' type).

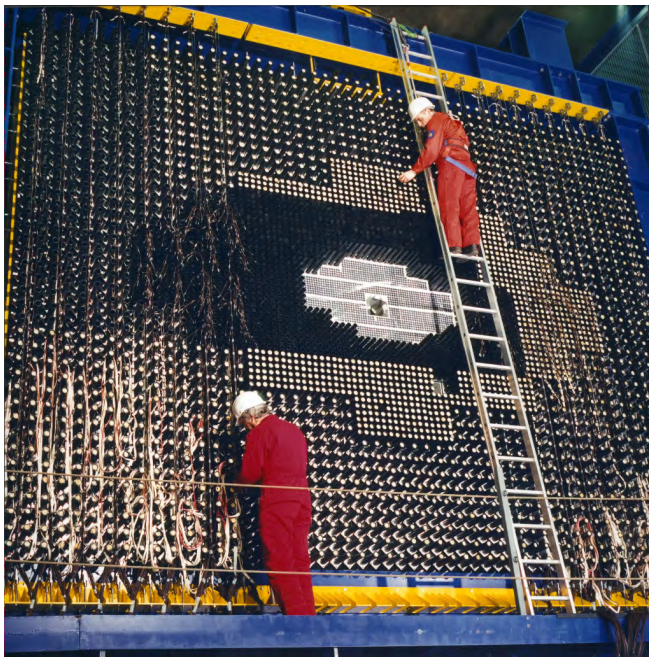


Fig. 15.28 The electromagnetic calorimeter of the HERA-B experiment during the installation of the photomultiplier tubes. Though not yet completed the three different regions inner, middle and outer ECAL are clearly visible (Source: DESY).

(fig. 15.27(d)). In this case either the amplified ionisation charge is measured or just the counting rate because it is sufficiently correlated with the deposited energy.

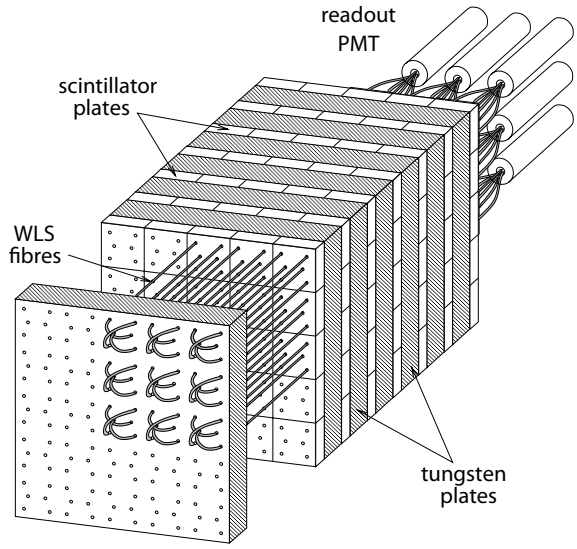
- Measurement of ionisation in silicon detectors. For this application the detectors are segmented into pad structures. So far semiconductor sampling detectors have primarily been used to construct particularly dense calorimeters to be deployed in areas which are difficult to cover, for example near the accelerator beam. An often-used combination is tungsten–silicon offering a particularly high calorimeter density (e.g. the luminosity monitor of the OPAL experiment [18] or the calorimeter of the balloon experiment CREAM on page 664).

– Spaghetti calorimeter (fig. 15.27 e): In this design scintillating fibres (about 0.5–1 mm thick) are strung through the passive medium. The spaghetti fibres are combined into bundles and read out by PMTs. An example is the spaghetti calorimeter of the H1 detector described in the next sub-section. In this context see also the application for hadron calorimeters discussed in section 15.6.3.

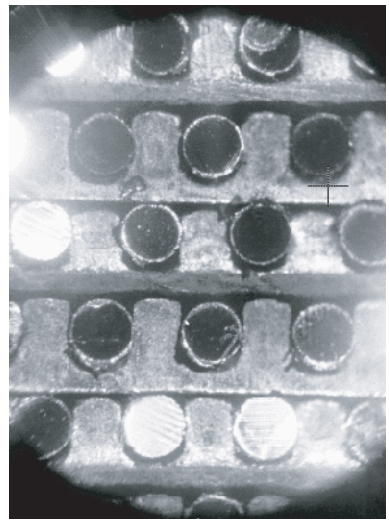
15.5.3.2 Examples for sampling calorimeters

As for the homogeneous calorimeters in section 15.5.2.2 we present a small selection of sampling calorimeters which are exemplary for passive and active materials or the design features.

The shashlik calorimeter of the HERA-B detector. The experiment HERA-B has used the 920-GeV proton beam of HERA (DESY, Hamburg) to study B-meson production and other phenomena in interactions of protons with the nuclei of wire targets. The electromagnetic calorimeter (fig. 15.28) was designed for measuring energies up to some 100 GeV [135].



(a) HERA-B shashlik calorimeter.



(b) H1 spaghetti calorimeter.

Fig. 15.29 Shashlik and spaghetti type calorimeters. (a) Schematic layout of a module of the inner calorimeter of the HERA-B detector (Source:DESY/HERA-B). Wavelength shifter fibres for the light transport to the photomultipliers are fed through the lead-scintillator sandwiches ('shashlik'). The module is subdivided in 5×5 readout units each of which is connected through a fibre bundle with a photomultiplier. (b) Magnified detail of a spaghetti calorimeter (H1 experiment) showing scintillating fibres surrounded by lead (from [111], with kind permission of Elsevier).

The calorimeter is constructed from lead-scintillator layers with wavelength shifter readout. The wavelength shifters are fibres which are fed through the plates, hence the term 'shashlik' in fig. 15.29(a). Guiding the light in the shower direction to the back allows for a dense packing of the modules and for a nearly arbitrary lateral segmentation of the readout by bundling fibre groups. The fibres transport the light to photomultipliers. Since in the case of HERA-B there is no magnetic field in the region of the calorimeter, photomultipliers can be used without the complications due to adverse magnetic field effects on the amplification process.

The calorimeter is subdivided into an inner (forward direction near the proton beam), middle and outer ECAL, with increasingly coarser segmentation of the cells. The more compact inner ECAL has a W–Ni–Fe alloy instead of lead as passive material, yielding a Molière radius of 1.24 cm in the inner region as compared to 4.15 cm in the outer regions. In table 15.5 on page 610 properties of the middle ECAL are listed as an example.

The spaghetti calorimeter of the H1 experiment. A *spaghetti calorimeter*, called SPACAL[5] [111], was deployed by the H1 experiment at the electron-proton storage ring HERA for measuring electrons scattered at small angles from the beam. It has an electromagnetic and a hadronic section, each with separate readout. The

[5]The concept of a spaghetti calorimeter (fig. 15.27(e)) was first applied to a prototype device [37] with the aim to provide optimal conditions for both electron and hadron calorimetry (see section 15.6.3).

Source	Lines (MeV)
^{137}Cs	0.662
^{60}Co	1.173, 1.333
^{228}Th	0.239, 0.583, 2.614

Table 15.7 Radioactive sources with γ lines which are employed for calibration of calorimeters with low noise level (see also table A.1).

passive medium of the SPACAL is lead, in which scintillating fibres as the active medium are strung along the shower direction. In contrast to the shashlik layout in this case light generation and light extraction are combined. It turned out to be technically too difficult to drill holes for the fibres in a long lead block. Therefore the lead was longitudinally cut into slices into which grooves for the insertion of the fibres were milled (see fig. 15.29(b)). The electromagnetic section (see table 15.5 on page 610) consists of 0.5 mm thick scintillating plastic fibres, embedded in a lead matrix with a lead-scintillator volume ratio of 2.3 : 1. In the hadronic section the fibres are 1.0 mm thick and the lead-scintillator ratio[6] is 3.4 : 1.

The ATLAS accordion calorimeter. The largest part of the accepted solid angle of the ATLAS detector [4] is covered by a lead–liquid-argon calorimeter for the detection of electromagnetic showers (see table 15.5 on page 610). As depicted in fig. 15.30 the passive layers are folded like an accordion. In contrast to sandwich calorimeters each layer is oriented parallel to the preferred shower direction with the folds running perpendicular to the shower axis. An electrode layer divides the space between two lead layers into two drift volumes, each 2.1 mm wide. Both the signal and HV connections go to the backside of the calorimeter (except for the first sampling where the signals go to the front). This allows for a straightforward transverse segmentation and an approximately complete coverage. The accordion folding also offers the advantage that the ionisation paths of the shower particles are longer than the corresponding paths for charge collection along the electric field perpendicular to the active layers. This is favourable for the signal-to-noise ratio and for the signal timing.

As shown in fig. 15.30 the calorimeter is divided into longitudinal sections with different granularities, finer at the beginning of the shower and coarser at the end. These longitudinal segments provide the functions of a presampler (page 604) and of a tail catcher (page 604) together with the possibilities of electron–hadron separation (section 14.4) and of logical cell connections to form trigger towers (see also section 15.4.1).

15.5.4 Calibration of electromagnetic calorimeters

Electromagnetic calorimeters are calibrated by employing test beams or by exploiting kinematical constraints of particle reactions. Examples for such reactions, which are suitable for calibration purposes and which can be analysed in parallel to the data acquisition, are Bhabha scattering at electron–positron colliders, the π^0 reconstruction from photon pairs and the reconstruction of resonances like J/ψ or Z^0 from electron–positron pairs. The signal of minimum-ionising particles (mips) can be used for calibration or at least for monitoring the stability of the signals.

[6] See the explanations in section 15.6.3 concerning the optimal lead-scintillator ratio for hadron calorimeters. For the H1 SPACAL the combination of the electromagnetic and hadronic calorimeters is not accordingly optimised because the requirements had other priorities (proximity to the beam, optimal resolution for electromagnetic showers).

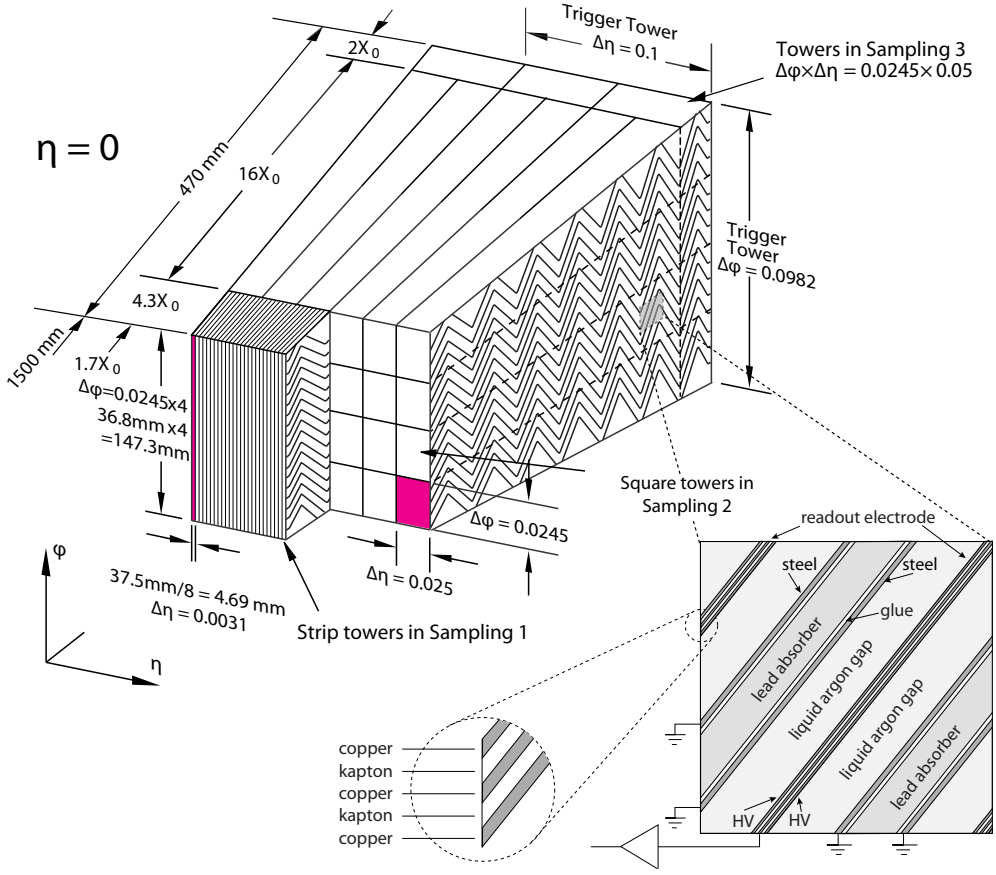


Fig. 15.30 Top: The lead–liquid-argon accordion calorimeter of the ATLAS experiment (barrel region) [4, 129]. The particles come from the interaction point at the lower left. The calorimeter covers a certain range in azimuth angle φ (around the beam direction) and in rapidity η, corresponding to the polar angle range relative to the beam ($\eta = 0$ corresponds to the direction perpendicular to the beam). In the η and φ directions the readout is segmented into $(\Delta\eta,\ \Delta\varphi)$ cells ('towers') and along the shower direction into three sections ('samplings') with different granularities. Bottom right: Layer structure of the middle section (adapted from [6]). The space between the lead absorbers is divided into two ionisation gaps by a readout electrode which receives the signals on the inner copper layers via capacitive coupling. The readout electrode is sandwiched between high voltage electrodes (HV) providing the drift field in the ionisation gaps. Source: CERN/ATLAS Collaboration.

For the calibration of calorimeters with noise levels in the keV range, which mostly applies to electromagnetic crystal calorimeters, radioactive sources with known gamma lines can be used. Some often-used sources are listed in table 15.7. By installing the sources in the calorimeter, optionally mechanically movable, one can calibrate in situ. However, gamma energies are too low to allow the extrapolation of the calibration to, for example, GeV energies. The necessary further calibration points at higher energies can be obtained by employing the above mentioned kinematical reconstructions.

15.5.5 Energy resolution of electromagnetic calorimeters

15.5.5.1 Energy dependences of resolution contributions

In general the energy resolution as a function of energy (fig. 15.20 on page 606) can be described by the formula (15.33):

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\,. \tag{15.35}$$

The predominant sources for the individual contributions have already been given in section 15.4.3: (a) stochastic fluctuations, (b) electronic noise and (c) various imperfections in the realisation of calorimeters. The constant term limits the resolution at high energy and the noise term at low energy.

Example. In order to compare a crystal calorimeter with a sampling calorimeter we take the values for the $PbWO_4$ calorimeter of the CMS detector and the liquid-argon calorimeter of the H1 detector from table 15.5 on page 610:

$$\begin{array}{lll} \text{CMS} & a = 2.8\,\%\ \sqrt{\text{GeV}}, & b = 120\,\text{MeV}, \quad c = 0.3\,\%\,, \\ \text{H1} & a = 11\,\%\ \sqrt{\text{GeV}}, & b = 150\,\text{MeV}, \quad c = 0.6\,\%\,. \end{array} \tag{15.36}$$

For a shower that was initiated by a 100-GeV electron one gets:

$$\begin{array}{llll} \text{CMS} & a/\sqrt{E} = 0.3\,\%, & b/E \approx 0, & c = 0.3\,\% \quad \Rightarrow \sigma/E \approx 0.4\,\%\,, \\ \text{H1} & a/\sqrt{E} = 1.1\,\%, & b/E \approx 0 & c = 0.6\,\% \quad \Rightarrow \sigma/E \approx 1.3\,\%\,. \end{array} \tag{15.37}$$

At this energy the constant term begins to dominate for the CMS crystal calorimeter while for the H1 sampling calorimeter this would only be the case at about 340 GeV.

In the following we discuss in detail the individual contributions to the energy resolution of calorimeters (see e.g. [385]). Because of its importance and because a lot of information from detailed research is available the discussion will become quite detailed. For a first reading we recommend to skip certain parts that we will indicate.

15.5.5.2 Stochastic term of homogeneous calorimeters

In homogeneous calorimeters the fluctuations in the deposited energy are usually small if the total energy can be deposited in the active volume. If in addition the conversion of the energy to a signal is very efficient, as for example the generation and processing of light in scintillating crystals (section 13.3, see the photon yield in table 13.3 on page 520), the resolution can get better by the Fano factor (see section 17.10.2). This factor quantifies the improvement over what would be expected from Poisson statistics of the signal contributions, here the light quanta in the crystal.

In such cases the intrinsic resolution can be limited by other factors, like position dependences of energy depositions and light generation and transport, leading to a slower improvement of the energy resolution than with $1/\sqrt{E}$. Often this can be parametrised by a $1/\sqrt[4]{E}$ dependence (e.g. for NaI and CsI calorimeters, see the description of homogeneous calorimeters in sections 15.5.2 and 13.5.2.2).

In contrast the light yield of lead-glass detectors (Cherenkov light) is much smaller (about 1000 photons per GeV, hence more than three orders of magnitude less than of inorganic scintillators, see also table 13.3 on page 520), such that here the $1/\sqrt{E}$ dependence prevails. In addition to the photon statistics one has also to account for the

fluctuations in the number of radiating electrons in the shower that lie above the Cherenkov threshold and thus contribute to the signal. An estimate of both contributions is given for a typical device in [383]:

$$\sigma_{tot} = \sqrt{\sigma_{ph}^2 + \sigma_{cut}^2} = \sqrt{0.02^2 + 0.032^2} = 3.8\,\% \text{ at } 1\,\text{GeV}\,, \tag{15.38}$$

where σ_{ph} and σ_{cut} denote the fluctuation in the number of photons and of radiating electrons, respectively. The actually measured resolutions (see examples in table 15.5 on page 610) tend to be somewhat worse, but in general not by much.

15.5.5.3 Stochastic term of sampling calorimeters

In sampling calorimeters, with alternating passive and active media, the energy that is deposited in the active medium and thus generates the signal can statistically fluctuate. The 'sampling fluctuations' mostly determine the energy resolution in sampling calorimeters. The fraction of the visible energy of the total deposited energy, the *sampling fraction*,

$$f_s = \frac{E_{vis}}{E_{dep}}\,, \tag{15.39}$$

typically amounts to only a few per cent. The number N_{vis} of the shower particles depositing the energy E_{vis} in the active medium is in a 'linear calorimeter' proportional to the primary energy E. In first approximation the resulting energy resolution can be estimated based on the Poisson statistics of the N_{vis} observable particles:

$$\frac{\sigma_E}{E} \approx \frac{\sqrt{N_{vis}}}{N_{vis}} = \frac{1}{\sqrt{N_{vis}}}\,. \tag{15.40}$$

In the following we assume that the fluctuations in the number of shower particles dominate the $1/\sqrt{E}$ term.

Rossi's 'Approximation B'. In the framework of the 'Approximation B' as presented in section 15.2.1 the total number N_{tot} of shower particles is estimated by (15.8) and their path length s_{tot} in the passive medium by (15.9):

$$N_{tot} \approx \frac{E_0}{E_c}, \qquad s_{tot} \approx \frac{E_0}{E_c} X_0\,. \tag{15.41}$$

If the thickness of a passive layer, measured in units of radiation length, is $t = d/X_0$ (d is the geometrical thickness) and the thickness of an active layer is comparatively negligible, then the number of transitions of shower particles from the passive to the active medium is:

$$N_{vis} \approx \frac{s_{tot}}{t\,X_0} = \frac{E_0}{t\,E_c}\,. \tag{15.42}$$

The term $1/t$ is the *sampling frequency* which specifies how frequently the shower is sampled over a radiation length:

$$\nu_s = \frac{1}{t} = \frac{X_0}{d}\,. \tag{15.43}$$

With these ingredients the sampling contribution to the resolution can be estimated:

$$\left(\frac{\sigma_E}{E}\right)_{samp} = \frac{1}{\sqrt{N_{vis}}} = \sqrt{\frac{t\,E_c}{E_0}} = 3.2\,\% \sqrt{\frac{t\,E_c/\mathrm{MeV}}{E/\mathrm{GeV}}} = a_t \sqrt{\frac{t}{E/\mathrm{GeV}}}\,, \qquad (15.44)$$

where the percentage value $3.2\% \approx 1/\sqrt{1000}$ comes from fixing the units to MeV and GeV. The coefficient a_t carries the index t to distinguish it from the coefficient a in (15.35): $a_t = a/\sqrt{t}$. The energy resolution improves with decreasing thickness $d = t\,X_0$ of the passive layer proportional to $\sqrt{d}$ because thinner passive layers imply more frequent sampling of the shower in the active layers. For this estimation the properties of the active layer, like thickness and material, do not play a role since the energy deposits in the active layers are neglected, which means that the particles crossing a layer are merely 'counted'. Therefore it does not matter if the active layer is for instance a scintillator or a gaseous medium as long as the sampling frequency is the same and both count shower particles with the same efficiency.

For example, the estimation (15.44), using radiation lengths and critical energies from table 15.1, yields for lead $8.6\% \sqrt{t/(E/\mathrm{GeV})}$ and for iron $15\% \sqrt{t/(E/\mathrm{GeV})}$ corresponding to $N_{vis} = 136/t$ per GeV for lead and $N_{vis} = 45/t$ per GeV for iron.

Necessary corrections to 'Approximation B'. The estimation of the energy resolution in (15.44) has to be taken as a lower limit. The experimentally observed resolutions are always worse, even if the condition of an efficient signal readout is fulfilled or the respective inefficiencies are taken into account. Most of the corrections to Rossi's Approximation B become necessary because the assumptions (see section 15.2.1) are too coarse:

- The assumption that the energy loss by ionisation is energy independent and equal to the critical energy per radiation length is not correct. Low-energy particles tend to have a much larger energy loss per path length. The portion of particles far below the critical energy is essential with contributions, for example, of electrons originating from photoeffect and Compton processes which are neglected in the approximation.
- The assumption of a constant energy loss does not account for Landau fluctuations (section 3.2.3).
- The assumption of an energy-independent absorption of photons, exclusively proceeding through pair production, is not correct. In fact, the energy spectra of electrons and photons extend to much lower energies than assumed in 'Approximation B'. In active media with low Z values this leads to photoeffect and Compton processes (neglected in 'Approximation B') with the corresponding creation of low-energy electrons.
- In pair production processes electrons and positrons are jointly produced leading to a high correlation and in the extreme case to a reduction of the Poissonian standard deviation by a factor $\sqrt{2}$. A reduction of this parameter, hence a worsening of the resolution, is also expected if the single layers are so thin that the same particle passes through several layers.
- The shower grows not only longitudinally but also laterally because of large scattering angles, in particular for low-energy particles. Therefore the particles have different path lengths in the active medium.
- The assumption that the resolution in (15.44) depends only on the thickness and the properties of the passive layers is too coarse because, amongst other reasons, the critical energy and hence the particle spectra change at the transition between materials with high and low Z (*transition effect*).

Using 'Approximation B' the signal response ϵ of a calorimeter per deposited energy would be the same for a mip and an electron or photon: $\epsilon_e/\epsilon_{mip} \approx 1$. Actually, however, the observations always show:

$$\frac{e}{mip} < 1 \qquad (\text{typically } \approx 0.5\text{–}0.7)\,. \tag{15.45}$$

Here we have used the convention $\epsilon_e/\epsilon_{mip} = e/mip$, pronounced *'e over mip'*. This behaviour will be discussed in more detail in the context of hadron calorimeters in section 15.6.2.

The transition to a realistic shower behaviour implies not only that the resolution becomes worse but also, in contrast to (15.44), that it becomes dependent on the properties of the active medium. This dependence has been confirmed by many studies with experiments and simulations. For example, in [916] a resolution of 12.9% (at 1 GeV) was measured with 4.2 mm thick lead absorber plates and 6.3 mm thick scintillator readout while with the same set-up but with a scintillator thickness of 12.6 mm the resolution improved to 10.8%. In the following we show that these differences can be explained by the physical behaviour of electromagnetic showers.

Prior to that, however, we still have to discuss the contribution due to fluctuations in the readout process. Since these fluctuations do not directly depend on the shower development in the calorimeter, the resolutions are often corrected for these effects if it comes to compare different detectors. Therefore we next discuss the readout fluctuations. For the further discussion we assume that the experimentally measured resolutions are corrected by quadratic subtraction of this contribution to obtain the 'intrinsic' resolution of a specific calorimeter.

Those readers that are less interested in the details can go directly to the summary on page 631.

Fluctuations of the readout quanta. The number N_Q of quanta detected by a readout device depends on the primary energy E, the sampling fraction f_S, the quantum yield n_Q per energy deposited in the active medium and the quantum efficiency η_Q:

$$N_Q = n_Q\, f_s\, \eta_Q\, E\,. \tag{15.46}$$

Both the quantum yield and efficiency introduce statistical fluctuations in the observed signal. While these are fluctuations in the observed quanta for a given deposited energy, the sampling fluctuations describe the fluctuations in the deposited visible energy. Although they are statistically not fully independent we add them in quadrature in order to make a comparison of both contributions to the resolution:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{a_S}{\sqrt{E}}\right)^2 + \left(\frac{a_Q}{\sqrt{E}}\right)^2}\,. \tag{15.47}$$

Here a_S and a_Q describe the sampling and signal quanta fluctuations, respectively. Because $a_Q/\sqrt{E} = 1/\sqrt{N_Q}$ and with (15.46) it follows that

$$a_Q = \sqrt{\frac{1}{n_Q\, f_s\, \eta_Q}}\,. \tag{15.48}$$

Example. Let the sampling term for a lead-scintillator calorimeter be $a_S = 7\%$. In order to have negligible readout fluctuations, the number of signal quanta, in this case

the number of photoelectrons, should be distinctly larger than $1/0.07^2 \approx 200$. At an assumed sampling fraction $f_s \approx 15\%$ the energy deposited in the active medium per GeV primary energy is 150 MeV. With about 10 000 photons per MeV deposited energy in a scintillator (Table 13.1 on page 506), quantum efficiencies of the photocathode of 20%, and a light collection efficiency of 10% the resulting number of photoelectrons per GeV is 30 000. The parameter of the corresponding readout fluctuations is $a_Q = 0.6\%$ and thus in this case completely negligible. However, dense packing of calorimeter modules usually requires photon readout via wavelength shifters (fig. 15.27(b, f)) leading to substantial losses in the light collection efficiency with typically 1000 photoelectrons per GeV or less (see section 13.4). With $1/\sqrt{1000} \approx 3\%$ this contribution to the resolution is often not negligible. For example, the uranium calorimeter of the ZEUS detector (section 15.7.1) has a light yield of 'at least 100 photoelectrons' per GeV for electromagnetic showers [331]. This yields an appreciable contribution from photon statistics of about 10% to the total resolution of about 18% (at 1 GeV).

Energy threshold for shower particles. In Rossi's 'Approximation B' (section 15.2.1) the assumption is made that the energy deposition is proportional to the total path length s_{tot} of all charged shower particles whose energy loss per path length is constant and equal to E_c/X_0. In reality the shower particles have an energy spectrum which favours low-energy particles with high ionisation density. Often low-energy particles contribute to the signal only above an energy threshold above which they are effectively 'counted'. Thresholds explicitly arise, for example as Cherenkov thresholds, if the detector signal is produced by Cherenkov light or by the saturation of scintillation light for low-energetic, strongly ionising particles (Birks' law, section 13.2.3), which also introduces an effective threshold. Such a threshold reduces the number N_{vis} of detectable particles. According to [835] a correction can be given by

$$N'_{vis} = F(\xi)\, N_{vis}\,. \tag{15.49}$$

The function $F(\xi)$ depends on the variable

$$\xi = 4.58\,\frac{\mathrm{g}}{\mathrm{mol}}\,\frac{Z}{A}\,\frac{E_{cut}}{E_c} \tag{15.50}$$

and thus on the minimal kinetic energy of the detectable particles E_{cut} normalised to the critical energy. An approximation for the function $F(\xi)$ for $\xi \leq 0.3$ is given in [88]:

$$F(\xi) \approx \mathrm{e}^{\xi}\left[1 + \xi \ln\left(\frac{\xi}{1.529}\right)\right]. \tag{15.51}$$

Quantitative evaluations of the function are listed in table 15.8.

Scattering angle distribution. 'Approximation B' assumes a one-dimensional, longitudinal shower development. In reality low-energetic electrons and positrons experience an appreciable scattering, a sizeable fraction even into the backward direction. Then the thickness of the sampling layer effectively increases, on average according to

$$t' = \frac{t}{\langle\cos\theta\rangle}\,, \tag{15.52}$$

where $\cos\theta$ is averaged over the distribution of the scattering angle θ. In [88] the average of $\cos\theta$ is estimated for not too large cut-off energies ($E_{cut} \lesssim 1\,\mathrm{MeV}$):

Table 15.8 Comparison of measured and calculated resolutions for electromagnetic calorimeters with with scintillator readout (from [88] with the additional column a_t). The columns specify the thickness x_a of the active layer (the numbers correspond to about 1 cm because scintillators have a density of about 1 g/cm^3), the experimental and theoretical values $a'_{t,exp}$ and $a'_{t,theo}$ for the modified constant of the sampling resolution defined in (15.54), where for $\langle\cos\theta\rangle$ the approximation was used. For comparison the last column displays the uncorrected coefficient, as introduced in (15.44).

Pass. med.	t	x_a $\left(\frac{\text{g}}{\text{cm}^2}\right)$	E (GeV)	$a'_{t,exp}$ (%)	E_{cut} (MeV)	ξ	$\frac{1}{\sqrt{F(\xi)}}$	$\frac{1}{\sqrt{\langle\cos\theta\rangle}}$	$a'_{t,theo}$ (%)	a_t (%)
Al	1.0	3.0	10–50	20	3.0	0.168	1.16	1.00	23.0	19.8
Fe	0.3–1.5	0.65	0.2–2.5	16.9	0.65	0.068	1.09	1.03	16.1	14.3
Pb	0.3–1.5	1.3	0.2–2.5	12.6	1.3	0.328	1.21	1.29	13.2	8.4

$$\langle\cos\theta\rangle = \cos\left(\frac{E_s}{\pi\, E_c}\right) \tag{15.53}$$

with the energy parameter $E_s = 21\,\text{MeV}$ that appears in Molière's theory according to (3.101) on page 67. The term $\langle\cos\theta\rangle$ becomes smaller with increasing Z since the critical energy E_c is about proportional to $1/Z$. This means that the effective path lengths in the passive layers increase due to scattering in the passive layers more strongly than in the active layers with typically lower average Z values.

With both corrections (15.49) and (15.52) an improved estimate of the energy resolution (15.44) is obtained (which, however, is not applicable for gaseous active media, see the corresponding explanations somewhat below):

$$\left(\frac{\sigma_E}{E}\right)_{samp} = 3.2\,\%\sqrt{\frac{t\,E_c/\text{MeV}}{F(\xi)\,\langle\cos\theta\rangle\,E/\text{GeV}}} = a'_t\sqrt{\frac{t}{E/\text{GeV}}}\,. \tag{15.54}$$

The factors $F(\xi)\,\langle\cos\theta\rangle$ effectively account for the dependences on the properties of the active layers, as discussed above, and also other effects as, for example, the mentioned correlation between electrons and positrons. If one wants to describe the resolution with an effective number of detected particles one can make the ansatz:

$$N'_{vis} = \frac{1}{t\,a'^2_t} = N_{vis}\,F(\xi)\,\langle\cos\theta\rangle\,. \tag{15.55}$$

In table 15.8 measured resolutions for Al, Fe and Pb calorimeters with scintillator readout are compared with those calculated using (15.54). The table shows the size of the various effects. While the resolution correction according to formula (15.54) does not become visible for aluminium, the lightest element in the table, the effect is quite pronounced for iron and lead. For lead (highest Z) the resolution deteriorates by about 60%.

Path length fluctuations. In the previous paragraph it was shown that the thickness of a layer effectively increases for scattered particles according to (15.52). While this effect is independent of the properties of the active layer because only the radiation length of a total sampling layer (passive plus active) has been included, we discuss in the following the impact of different path lengths in the active medium, caused by

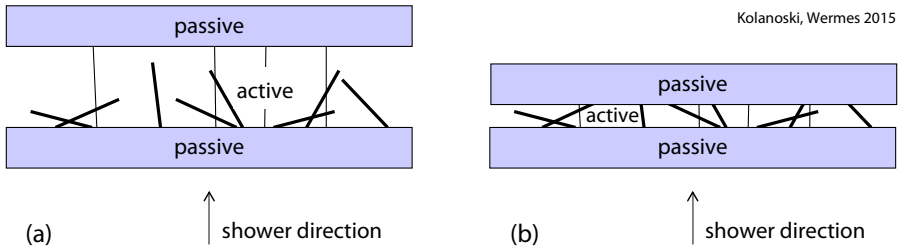


Fig. 15.31 Schematic representation of shower particles which pass a thick (a) and for comparison a thin (b) active layer. In order to simplify the picture it is assumed that there are only two particle classes: low-energy particles (thick lines) with an isotropic scattering angle distribution and high ionisation density which are absorbed in the thicker active layer, and high-energy particles (thin lines) which go into the forward direction and deposit an energy proportional to the layer thickness.

scattering, which lead to a dependence of the fluctuations from the thickness of the active medium.

There is a strong correlation between the particle's energy and average scattering angle: at lower energies the angular distribution becomes more isotropic, at higher energies the forward direction increasingly dominates. With reference to the schematic sketch in fig. 15.31 we demonstrate how the thickness of the active layer influences the fluctuations in energy deposition from low- and high-energy shower particles. In the thicker readout layer of the sketch (fig. 15.31) the low-energy particles are absorbed and therefore always deposit their full kinetic energy. The more energetic particles always deposit an energy proportional to the layer thickness. If the thickness of the thinner layer is smaller than the range of the low-energy particles then their energy deposition strongly depends on the crossing angle and thus on the path length in the active layer. With decreasing active layer thickness the signal of the perpendicularly crossing high-energy particles becomes smaller and at the same time the fluctuations of the signals from low-energy particles become larger because the signal becomes dependent on the angle.

So far we have assumed that the thickness of the active layer can be neglected or that it enters only into the sum with the passive layer. The path length fluctuations imply an explicit dependence of the fluctuation on the thickness of the active medium since the transitions from the passive to the active medium are no longer only counted but get weighted by the ionisation density. Using the terms 'thin' and 'thick' as a property of an active layer we refer to the mass density weighted geometrical thickness or column density ρx, as defined in (3.28).

The degradation of the resolution through path length fluctuations is demonstrated in fig. 15.32(a) by a simulation study of a lead–liquid argon calorimeter [411]. The energy resolution deteriorates with decreasing thickness of the active layer, with a particularly strong decline below a thickness of about 2 mm liquid argon.

While for liquid argon a thickness below about 2 mm is not possible (technically difficult), for gases as active media thin layers ('thin' in terms of small column density) cannot be avoided because gas densities are only about a thousandth of the densities of solids or liquids. Correspondingly the energy loss per minimum-ionising particle is in gaseous detectors only some keV and due to the low ionisation density Landau fluctuations additionally degrade the resolution.

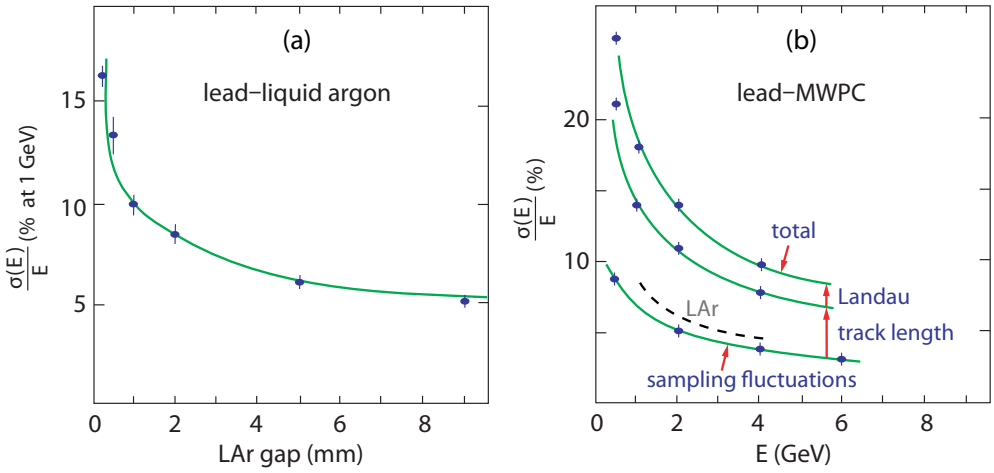


Fig. 15.32 Simulations of effects caused by the readout layer thickness (from [411], with kind permission of Elsevier). (a) Simulation of the energy resolution of a lead–liquid argon calorimeter. The resolution for 1 GeV primary energy is displayed as a function of the thickness of the active liquid argon layer with the passive lead layer thickness kept fixed at $X_0/3$. (b) The contributions of the sampling, Landau and path length fluctuations to the total energy resolution of a lead-MWPC calorimeter. The curves have been computed by simulations of 48 layers of lead with MWPC readout, each 1/3 radiation length thick. The resolution contributions at 1 GeV are about 7% (sampling), 12% (path length), 12% (Landau). The resulting total resolution of 18% (which was also achieved by the ALEPH calorimeter, see table 15.5) can be compared to a resolution of 8.8% for a liquid argon readout.

Semiconductors as active layers are also usually thin in the sense discussed here, that is, they have a low column density (3.28). For example, tungsten–silicon calorimeters can be built very compactly so that they are often deployed in spatially tight detector regions, like near the beam pipe. In a 300 μm thick silicon detector a minimum-ionising particle loses about 120 keV leading to strong path length fluctuations. In this case, however, the statistical fluctuations of charge generation are lower than in gases because of the lower ionisation threshold leading to more charge carriers for a given energy loss.

Landau fluctuations. The distribution of the ionisation energy loss dE/dx is given by the Landau distribution (section 3.2.3), which is asymmetric with a tail to large energy losses due to δ-electrons. The energy loss is subject to strong fluctuations which, due to the ionisation statistics, become stronger the thinner the layer, in particular if the δ-electrons are not fully contained in the active medium. For example, in 5 mm argon gas, which is a possible layer thickness for a readout with a proportional chamber, a mip produces only 15 primary ionisations each followed by on average three secondary ionisations with strong fluctuations in number and deposited energy of the secondary electrons (see table 7.1).

An estimate of the contribution of Landau fluctuations to the energy resolution is given for example in [383] on the basis of the model developed in [590]:

$$\left(\frac{\sigma_E}{E}\right)_{\text{Landau}} \approx \frac{1}{\sqrt{N_{vis}}} \frac{3}{\ln(1.3 \times 10^4\, \Delta E/\text{MeV})}\,. \tag{15.56}$$

Here ΔE is the mean energy loss of a track in a readout plane. Typical values are

$$(\sigma_E/E)_{\text{Landau}} \approx 30\% \frac{1}{\sqrt{N_{vis}}} \quad \text{for } \Delta E = 1\,\text{MeV (solid, liquid)},$$
$$(\sigma_E/E)_{\text{Landau}} \approx 100\% \frac{1}{\sqrt{N_{vis}}} \quad \text{for } \Delta E \simeq 10^{-3}\,\text{MeV (gaseous)}.$$

Figure 15.32(b) displays the effects of both the Landau and the path length fluctuations as obtained by a simulation of a lead-MWPC calorimeter [411]. The comparison with a corresponding arrangement with liquid-argon readout supports the general observations that the energy resolutions obtained with gaseous media are much worse than those obtained with solid or liquid media. Although MWPCs and drift tubes are otherwise attractive for calorimeter readouts (e.g. they have been employed for the calorimeters of the LEP detectors ALEPH [346], see also table 15.5, and DELPHI [11]), the poor energy resolution limits their importance for electromagnetic calorimeters in accelerator experiments.

Summary on 'stochastic term of sampling calorimeters'. For the design of sampling calorimeters it is one of the most important goals to minimise the contributions of sampling fluctuations to the energy resolution. Therefore, these contributions have been discussed in some detail in the preceding text. Here we summarise the most important points.

We started with the simple formula (15.44), based on Rossi's 'Approximation B',

$$\left(\frac{\sigma_E}{E}\right)_{samp} = 3.2\,\% \sqrt{\frac{t\,E_c/\text{MeV}}{E/\text{GeV}}}\,, \tag{15.57}$$

which depends only on the sampling frequency $\nu_s = 1/t$ and on the critical energy E_c of the passive medium, in particular not on the sampling fraction, that is, the energy fraction deposited in the active layer. Experimentally the assumption that the resolution is independent of the properties of the active layer requires corrections which we discussed in the following sequence:

Fluctuations of the readout quanta: This contribution is independent of the shower development and can be subtracted from the total resolution if one wants to study the intrinsic calorimeter resolution.

Energy threshold for shower particles: This reduces the number N_{vis} of detectable shower particles according to (15.49). In addition, defining an effective energy threshold, a dependence on the thickness of the active layer results.

Scattering angle distribution: Scattering effectively increases the thickness of a sampling layer according to (15.52); the Z dependence of scattering makes this effect different for the passive and the active layers.

Path length fluctuations: Due to the energy–angle correlations, sketched in fig. 15.31, this contribution is important in thin (small column density) active layers.

Landau fluctuations: According to (15.56) this contribution becomes important in thin active layers with small ionisation statistics.

The resolution contributions can be summarised by ordering them according to whether they depend on the thickness of the passive layer $x_p = t\,X_{0,p}$ or on the thickness of the active layer $x_a = s\,X_{0,a}$ ($X_{0,p}$ and $X_{0,a}$ are the respective radiation

Table 15.9 Parameters of the formula (15.58) for sampling calorimeters with various passive and active media [777].

Passive	Active	σ_0 [%]	α	β
C	scint.	16.48±2.50	0.72±0.03	0.16±0.02
Al	scint.	11.02±1.21	0.70±0.03	0.15±0.02
Fe	scint.	6.33±0.52	0.62±0.03	0.21±0.02
Sn	scint.	4.53±0.32	0.65±0.03	0.25±0.03
W	scint.	3.61±0.17	0.70±0.03	0.29±0.03
Pb	scint.	3.46±0.19	0.67±0.03	0.29±0.03
U	scint.	3.28±0.15	0.67±0.03	0.30±0.03
Pb	Si	5.04±0.20	0.66±0.03	0.24±0.03
Pb	LAr	6.49±0.31	0.62±0.03	0.19±0.03

lengths[7]). For these dependences a compact formula is given in [777]:

$$\frac{\sigma_E}{E} = \frac{\sigma_0}{\sqrt{E}} \frac{t^\alpha}{s^\beta} \tag{15.58}$$

with empirical parameters σ_0, α and β which, however, always apply only to definite combinations of passive and active media. In table 15.9 some examples are listed for parameters of media typical for sampling calorimeters. Obviously with $\alpha = 0.6$–0.7 the $\sqrt{t}$ dependence of the Rossi formula (15.44) is approximately reproduced. Also, the dependence on the critical energy is similar. For a certain active medium, for example the scintillator in table 15.9, the dependence on the absorber layer is a good approximation:

$$\sigma_0 \propto \sqrt{E_c} \propto \frac{1}{\sqrt{Z}} \,. \tag{15.59}$$

Gaseous active layers were not considered in [777] and are thus not included in table 15.9. However, the resolution value given in fig. 15.32 for the combination Pb–MWPC with a gas thickness $s \approx 10^{-4}$ would be described by similar parameters as for PB–scintillator, that is, $\beta \approx 2/3$ and $\alpha \approx 1/4$.

15.5.5.4 Noise term

The electronic noise of the readout electronics has a constant, energy-independent variance so that the contribution to the relative resolution decreases with energy proportionally to $1/E$. The coefficient b in (15.33) is called 'equivalent noise energy' (see the explanations for eq. (15.33), page 606) defining the smallest energy measurable with the particular calorimeter. For example, table 15.5 yields for the liquid argon calorimeter of ATLAS an equivalent noise energy of 190 MeV, but for the CsI calorimeter of BaBar only 150 keV. Therefore, concerning the energy resolution for the detection of photons in the MeV range CsI is better suited, while for GeV photons other criteria are more important. A general discussion of the influence of noise on detector resolutions is presented in section 17.10.

The influence of noise depends on the strength of the primary signal, thus on the amount of light or charge generated per deposited energy, and on the further signal

[7] For the active layer the appropriate reference length would be rather the column depth or some other measure for the ionisation loss than the radiation length.

reprocessing. For example, for light readout by a photomultiplier with high amplification, noise is often negligible compared to other effects. This is usually not the case for a readout with photodiodes. However, photodiodes compared to photomultipliers can be operated in more compact geometries and in magnetic fields. Therefore intensive research is ongoing for the development of photosensors that combine the positive properties of both types (see the corresponding section 13.4 and table 10.2 on light readout).

If the primary signals are relatively small, which is in particular the case for readout based on the principle of ionisation chambers (e.g. liquid argon calorimeters), a low-noise characteristic of the readout system up to the first signal amplification is crucial. Since the noise level increases with the capacitance at the input of the amplifier (see section 17.10.3) the capacitances of a single detector channel, of the cable connections and further parasitic capacitances should be minimised. In addition the noise increases with the bandwidth of the amplifier, causing higher noise levels for faster signals. Fast signals are often required to cope with high particle rates. Therefore an optimum between signal speed and noise suppression has to be found by pulse shaping and filtering.

15.5.5.5 Constant term

The contribution of a constant term to the relative energy resolution limits the resolution at high energies because all other terms fall off with energy. Correspondingly, the minimisation of the constant term demands most attention in experiments at the highest energies. Well-designed calorimeters reach values of the constant term around 1%.

Mechanical tolerances and intercalibration errors. Contributions to the constant term arise primarily from irregularities of the calorimeter in different detector regions, which are caused by mechanical or electronic imperfections. For example, such imperfections can be missing detector coverage and 'dead material' at transitions between detector modules as well as support structures necessary for the detector installation. In particular when integrating detector modules into a large system such imperfections cannot be avoided. Many measurement deviations can be reduced by a correct calibration. However, in particular for large systems, a so-called *intercalibration error* will remain which enters into the constant term of the resolution. Since such errors naturally show up only for large systems it is problematic to infer from test measurements on prototypes what the resolution for the final detector will be.

Leakage losses. Leakage losses arise when the longitudinal or transverse shower profile extends beyond the detector dimensions. Related to this subject is the discussion of construction requirements for calorimeters in section 15.4.2 where the reduction of fluctuations by employing *tail catchers* and *presamplers* is also described.

The longitudinal losses contribute more to the energy resolution because the shower development fluctuates more strongly in the longitudinal than in the transverse directions (see also fig. 15.18). The losses can be estimated using the Longo formula (15.17) where the target mark of 98% coverage is approximately given by formula (15.20). The influence of the longitudinal loss fluctuations on the energy resolution can be estimated as follows [383]:

$$\left(\frac{\sigma_E}{E}\right) = \left(\frac{\sigma_E}{E}\right)_{f_{\text{leak}}=0} \left(1 + 2\sqrt{E/\text{GeV}}\, f_{\text{leak}}\right) , \tag{15.60}$$

where f_{leak} is the fractional energy leakage.

For a given calorimeter the leakage losses f_{leak} increase about proportionally to the logarithm of the energy. Thus the expression (15.60) does not approach a constant for high energies and is therefore not really compatible with the parametrisation (15.35). This parametrisation also assumes a Gaussian behaviour of the resolution which is no longer the case for large leakage losses because the losses have an asymmetric distribution due to the boundary $f_{\text{leak}} \geq 0$. In practice, however, for most of the electromagnetic calorimeters, formula (15.35) can still be used for the parametrisation of the energy resolution.

15.6 Hadron calorimeters

In a *hadron calorimeter*, hadronic showers (section 15.3) are detected by measurement of the ionisation of an active medium, sometimes also by observing the Cherenkov light generated by charged particles. As compared to electromagnetic showers the development of hadronic showers, however, is much more complex and the multitude of contributing reactions with often very different signal efficiencies leads to strong intrinsic signal fluctuations. They are mainly due to the fluctuations in the splitting of the deposited energy into an electromagnetic and hadronic part together with the losses of binding energy by breaking up the nuclei. Since the signal responses of the various shower contributions are in general different, the fluctuations affect the resolutions of hadronic calorimeters which are therefore in general worse than those of electromagnetic calorimeters (see the respective typical values in (15.2) on page 584). In the following we investigate the different contributions of inelastic reactions and their influence on the measured signals and the resulting fluctuations.

In recent decades much development work has been achieved, leading to a systematic improvement of the properties of hadronic calorimeters by a steadily growing understanding of the basic processes of shower development and signal recording [982, 385]. The idea of 'compensation' was essential for the improvement of the energy resolution. The term compensation comprises methods for minimising the influence of the energy loss fluctuations on the signal (section 15.6.3).

15.6.1 Typical structure of a hadron calorimeter

Hadron calorimeters are practically always designed as sampling calorimeters. Homogeneous calorimeters would not yield the advantages seen for electromagnetic showers, primarily because any potential gain in energy resolution would be spoiled by the strong intrinsic fluctuations in a hadronic shower. In addition, homogeneous hadron calorimeters would be technically difficult to realise because of the large geometrical dimensions required for the containment of a hadron shower.

The choice of the passive and active media is influenced by several considerations. Table 15.3 on page 599 shows that media with lower atomic number Z have a smaller ratio of nuclear absorption length to radiation length than those with higher Z. Therefore media with lower Z offer a more similar development of the hadronic and the electromagnetic components of a shower than with higher Z. From this perspective alone and combined with the wish to have at the same time a density as high as possible to keep the calorimeter compact, iron becomes a preferable choice for the passive medium. In addition, it can be readily machined and may be part of an already existing detector piece which could be, for example, the return yoke of a magnet.

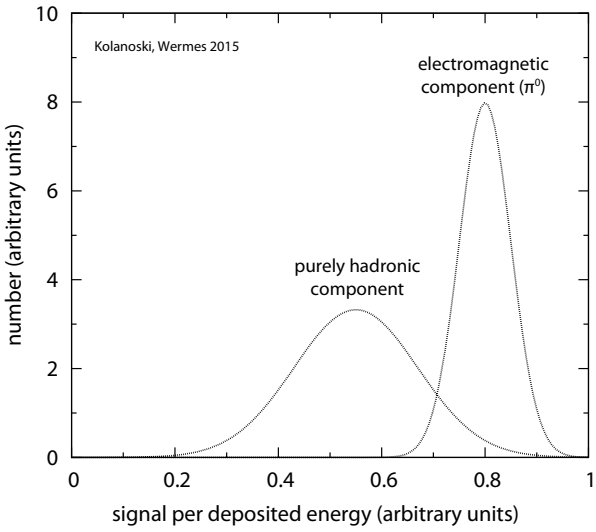


Fig. 15.33 Typical distribution of signal efficiencies for the hadronic and electromagnetic components of a hadron shower (adapted from [979]). The shown relations are typical for a (non–compensating) iron calorimeter with scintillator or liquid argon readout.

The 'classics' amongst the hadron calorimeters have 1–2 cm thick iron plates as passive layers and some millimetres scintillator or liquid argon as readout (examples are the calorimeters of H1, CDF and ATLAS, see table 15.11 on page 652).

'Compensating' calorimeters follow a quite different philosophy. In this case the requirement that the signal yield for the electromagnetic component is suppressed relative to that of the hadronic component leads to the choice of passive layers with particularly high Z values, like uranium, lead and tungsten. The choice and the adjustment of passive and active layers with the appropriate sampling fractions will be discussed in detail in section 15.6.3.

15.6.2 Calorimeter signals of electrons and hadrons

In section 15.3.2 we have discussed the different fractions f_i of the energy of a hadronic shower deposited in a matter block. If these fractions are detected with different efficiencies ϵ_i then the fluctuations of the f_i result in fluctuations of the calorimeter signal. The strongest fluctuations are obtained for the electromagnetic fraction f_{em} and the corresponding hadronic fraction $f_h = 1 - f_{em}$. Without any specific measures their signal efficiencies are in general very different (fig. 15.33).

Definition of e/π and e/h. The calorimeter signal generated by a hadron (the symbol π for a (charged) pion is generically used for a hadron) is composed of the fractions f_i of the deposited energy E (see eq. (15.25)) multiplied with the respective signal efficiencies ϵ_i:

$$S(\pi) = \left(f_{em}\,\epsilon_{em} + \underbrace{f_{ion,r}\,\epsilon_{ion,r} + f_{ion,nr}\,\epsilon_{ion,nr} + f_n\,\epsilon_n + f_\gamma\,\epsilon_\gamma + f_B\,\epsilon_B}_{f_h\,\epsilon_h} \right) E\,. \tag{15.61}$$

The different fractions f_i have been discussed in section 15.3.2. The fraction f_{ion} introduced in (15.25) is here separated into the fractions for relativistic and non-relativistic particles, $f_{ion} = f_{ion,r} + f_{ion,nr}$, because the respective signal efficiencies can be very different.

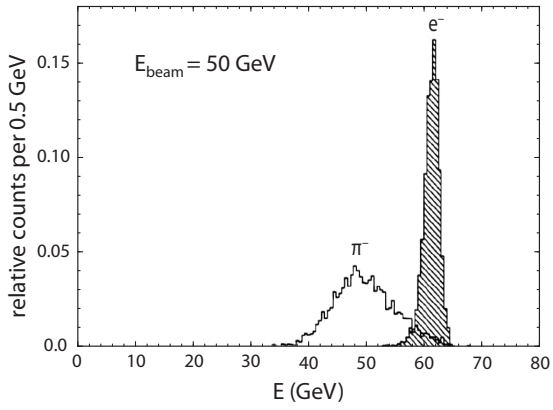


Fig. 15.34 Measurement of e/π of the liquid argon calorimeter of H1 using 50-GeV electron and pion test beams [203]. The measured energy is expressed on the 'hadronic energy scale', such that the mean of the pion distribution lies at 50 GeV. The electron signal is higher by 22%, corresponding to $e/\pi = 1.22$.

In general the calorimeter signals for electrons and hadrons are different:

$$\frac{S(e)}{S(\pi)} = \frac{\epsilon_{em}\, E}{(f_{em}\epsilon_{em} + f_h\epsilon_h)\, E} = \frac{\epsilon_{em}/\epsilon_h}{1 - f_{em}\left(1 - \dfrac{\epsilon_{em}}{\epsilon_h}\right)}\,. \tag{15.62}$$

One distinguishes

$$\begin{aligned} \frac{S(e)}{S(\pi)} &\overset{\text{def}}{=} \frac{e}{\pi} && \text{determined with } e^{\pm} \text{ and } \pi^{\pm} \text{ test beams (fig. 15.34),} \\ \frac{\epsilon_{em}}{\epsilon_h} &\overset{\text{def}}{=} \frac{e}{h} && \text{intrinsic calorimeter property.} \end{aligned} \tag{15.63}$$

These expressions are pronounced 'e-over-π' and 'e-over-h' ratios. From (15.62) one obtains

$$\frac{e}{h} = 1 \implies \frac{e}{\pi} = 1\,. \tag{15.64}$$

This means that electrons and pions of equal energy deliver equal signals if the calorimeter responds equally to the electromagnetically and hadronically deposited energy fraction, that is, for $e/h = 1$. Such a calorimeter is called 'compensating'. In fig. 15.35 one also sees that e/π approaches 1 for asymptotic energies, independent of the e/h ratio because the electromagnetic fraction f_{em} in a hadronic shower increases with the energy, approaching 1, see (15.26).

Resolution and linearity of hadron calorimeters. The fluctuations, in particular between the electromagnetic and hadronic components, deteriorate resolution and linearity of a hadron calorimeter. For the resolution one can make an ansatz [979]:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E/\text{GeV}}} \;+\; b\left(\frac{e}{h} - 1\right)\,. \tag{15.65}$$

Sometimes both terms are quadratically added. The first term, the stochastic contribution, determines the limit of the resolution of a fully 'compensating' calorimeter with $e/h = 1$. For example, in compensating calorimeters $a \approx 0.3$–0.35 has been reached; the parameter b is of the order 1, with a possible energy dependence [982].

A ratio $e/h \neq 1$ also influences the linearity of a calorimeter because the electromagnetic fraction f_{em} is energy dependent (see eq. (15.26)). The non-linearity with respect to an energy E_0 can be estimated as follows [982]:

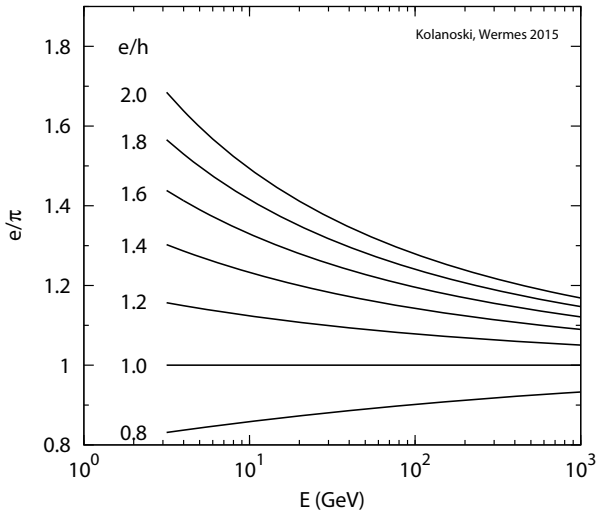


Fig. 15.35 The energy dependence of e/π, the ratio of the calorimeter signals from electrons and hadrons, for different values of the parameter e/h. The ratio e/π was calculated according to (15.62) with an energy dependence of f_{em} corresponding to (15.26) on page 597 with the parameters $E_0 = 1\,\text{GeV}$ and $k = 0.82$.

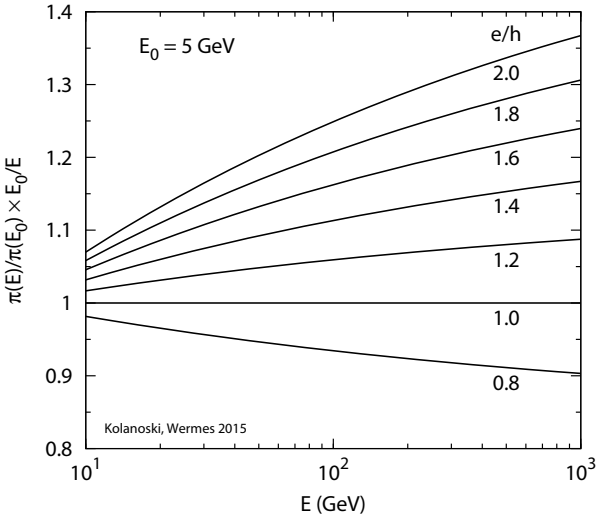


Fig. 15.36 The nonlinearity estimated using (15.66) for hadron calorimeters with different e/h values (the reference energy E_0 is always 5 GeV).

$$\frac{S(\pi(E))/E}{S(\pi(E_0)/E_0} = \frac{f_{em}(E) + (1 - f_{em}(E))\,(e/h)^{-1}}{f_{em}(E_0) + (1 - f_{em}(E_0))\,(e/h)^{-1}}\,. \tag{15.66}$$

In fig. 15.36 the energy dependence of the nonlinearity is shown for different values of e/h using the parametrisation (15.26) of f_{em} (page 597).

The mip signal as a reference. Since form and size of the signals are detector-specific it is customary to use as a reference the signal of a minimum-ionising particle (mip). With the definition of the efficiency ϵ_{mip} a mip which deposited in the considered detector an energy E generates the signal

$$S(mip) = \epsilon_{mip}\, E\,. \tag{15.67}$$

However, this signal is not directly measurable because a particle that is indeed minimum-ionising would no longer be minimum-ionising after having traversed a

calorimeter and also the deposited energy E is generally not known. This problem can be overcome by determining a calibration signal, for example with a muon beam, from which the interesting quantity ϵ_{mip} can be derived with the aid of simulation calculations.[8] The signals obtained with other particle beams can then be related to this mip signal. Expressed in 'mip units' the signal of a hadron given in (15.61) is

$$\frac{S(\pi)}{S(mip)} = \frac{\pi}{mip} = f_{em}\frac{\epsilon_{em}}{\epsilon_{mip}} + f_{ion,r}\frac{\epsilon_{ion,r}}{\epsilon_{mip}} + f_{ion,nr}\frac{\epsilon_{ion,nr}}{\epsilon_{mip}} + f_n\frac{\epsilon_n}{\epsilon_{mip}} + f_\gamma\frac{\epsilon_\gamma}{\epsilon_{mip}}\,. \quad (15.68)$$

As common in the literature we will abbreviate the ratios of efficiencies analogously to e/h in (15.63):

$$\frac{S(\pi)}{S(mip)} = \frac{\pi}{mip} = f_{em}\frac{e}{mip} + f_{ion,r}\frac{r}{mip} + f_{ion,nr}\frac{nr}{mip} + f_n\frac{n}{mip} + f_\gamma\frac{\gamma}{mip}\,. \quad (15.69)$$

In sampling calorimeters the individual signal contributions exhibit the following behaviour (mostly adapted from [982]):

e/mip: According to the simple model for electromagnetic showers electrons completely deposit their energy through ionisation when they reach the critical energy. Consequently one would expect that the signal efficiency is about that of mip particles. In reality the ratio e/mip can be significantly below 1, the more so the higher the atomic number Z of the passive layer. The reason is as follows. A large fraction of the energy is deposited by electrons with less than 1 MeV of energy. These electrons are preferentially generated by MeV photons via Compton and photoeffect. In particular because of the strong increase of the photoeffect cross section with Z (section 3.5.3 and fig. 3.49) the photons interact preferentially in the passive layer where they were generated. The low-energy electrons also have little chance to leave the passive layer because of their short range at low energy (e.g. 691 µm for 1-MeV electrons in lead [559]).

For scintillators as active medium one finds for different passive layers [258]: uranium and lead $e/mip \approx 0.6$; iron and copper $e/mip \approx 0.9$, corresponding to the expectation that the ratio e/mip approaches 1 as the Z values of the passive and active layers become more similar. These numbers are nearly independent of the relative thickness of the passive and the active layer for a passive layer thickness above about 3 mm. For copper with liquid argon readout, hence higher Z of the active layer than for scintillator, e/mip is close to 1 [258].

r/mip: For relativistic charged hadrons, which are generated in a high-energy cascade, one can assume that they behave about as mips, hence yielding $r/mip \approx 1$.

nr/mip: The ionisation contribution by non-relativistic particles mainly comes from spallation protons which preferentially have kinetic energies below 1 GeV and are thus non-relativistic. Heavier fragments have a minor chance to change over to the passive medium, thus barely contributing to the signal. In the next item we separately talk about neutrons.

[8] The extraction of ϵ_{mip} in this way is not quite simple because with increasing energy muons create increasingly more energetic δ electrons which can also lose energy through bremsstrahlung and pair production, thus forming electromagnetic cascades. Therefore, even with mip signals the efficiencies ϵ_{ion} and ϵ_{em} cannot be clearly separated.

Protons which have an energy sufficient to cross several active layers exhibit a similar behaviour as mips because the ratio of the energies deposited in the passive and active layers is similar to that for mips. Below an energy of about 50 MeV the signal yield becomes smaller because the proton range becomes similar to or smaller than the thickness of the passive layer. For example, the range of 50-MeV protons in uranium is 2.8 mm, in lead 4.5 mm and in iron 4.3 mm. In scintillators the contribution of low-energy, highly ionising particles is limited by saturation effects (see Birks' law in section 13.2.3).

n/mip: Neutrons with energies around 100 MeV emerging from spallation reactions in similar numbers to protons can elastically scatter (see details below) or can trigger nuclear reactions, including fission of heavy nuclei. A much larger number of neutrons, however, arises in the subsequent evaporation process yielding neutrons with energies of a few MeV. They can excite nuclei to emit charged fragments like protons or α particles and gamma rays. Again, the charged nuclear fragments contribute nearly nothing to the signal because of their short range. The gamma radiation will be separately discussed under the next item.

The most frequent neutron process is elastic scattering off nuclei. The relative energy transfer in the scattering of a neutron off a nucleus with mass number $\mathcal{A}$ is (see e.g. [317]):

$$\frac{\Delta E}{E_n} = \frac{2\,\mathcal{A}\,(1-\cos\theta)}{(\mathcal{A}+1)^2}\,, \tag{15.70}$$

where θ is the scattering angle in the centre-of-mass system. Under the assumption that the angular distribution is isotropic in the centre-of-mass system, averaging over the scattering angles yields

$$\left\langle \frac{\Delta E}{E_n} \right\rangle = \frac{2\,\mathcal{A}}{(\mathcal{A}+1)^2}\,. \tag{15.71}$$

For heavy nuclei the neutron's relative energy loss is proportional to $1/\mathcal{A}$ yielding for uranium and lead about 1% and for iron about 3%, but for the light hydrogen 50%. Therefore, if the active medium contains hydrogen, the neutron energy can very efficiently be transferred to protons which subsequently deposit their energy by ionising the active medium. This ionisation is very local because of the low proton energy and hence short range. Since little energy is lost in this way, in the passive medium values of $n/mip \gg 1$ can be reached.

The surprising magnitude of this effect can be seen by the following quantitative example. In [982] an infinitely extended sampling structure is considered which only consists of (liquid) hydrogen and lead. A volume ratio H_2(liquid) : Pb of 1 : 99 yields for 1-MeV neutrons $n/mip = 1630$, and a ratio 50 : 50 yields $n/mip = 45$. Even if hydrogen is only contained as a fraction in the active medium, for example in a scintillator, really large n/mip values can be reached. This opens the possibility for a targeted compensation of the non-detectable energy in hadron showers, as will be discussed in more detail in the next section.

If neutrons are thermalised by elastic scattering off light nuclei the cross section for neutron capture becomes large; the capture is delayed due to the thermalisation time. The nuclei preferably release the excess binding energy through gamma radiation, which is also delayed relative to the primary signal.

γ/mip: Signals by gamma radiation which arise from various nuclear reactions are suppressed like the electromagnetic part of the shower, thus yielding $\gamma/mip < 1$. Specific to this radiation is that it occurs partially strongly delayed. This can be purposefully exploited for active media with a relatively high Z by longer signal integration (for this technique liquid argon, for example, is better suited than scintillator). In practice, however, during a longer integration time more background may accumulate under the signal, thereby deteriorating the energy resolution.

15.6.3 Compensation

We have seen that resolution and linearity of hadron calorimeters can be substantially improved if the signals from electromagnetic and hadronic energy depositions are equalised, that is, if e/h is close to 1. It turns out that it is indeed possible to reach $e/h = 1$. The different methods developed to reach this goal are summarised by the term 'compensation', the meaning of which will be explained in the following paragraphs.

The idea of compensation of the non-detectable energy in nuclear reactions, which lead to smaller calorimeter signals of hadrons relative to those of electrons (fig. 15.34), was first proposed and studied by Fabjan, Willis and co-workers [384]. The proposal was scrutinised by comparing the behaviour of an iron and an uranium calorimeter with liquid argon readout. The tested hypothesis was that due to the higher cross section for neutron production in ^{238}U as compared to iron the additionally produced neutrons, protons and gammas from nuclear fission deposit energy in the active medium, at least partly compensating for the non-detectable (mainly binding) energy. As a matter of fact, with uranium an approximate equality of the signals from electrons and hadrons was found, resulting in a better resolution than with iron.

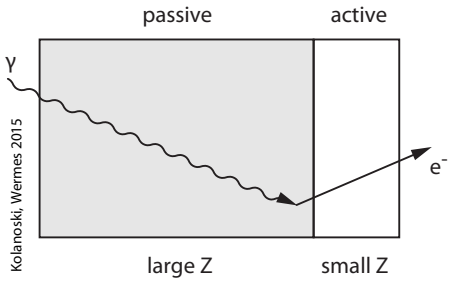
These investigations subsequently led to the construction of the first uranium calorimeter [70] for the Axial Field Spectrometer (R807) at the proton–proton collider ISR at CERN [468]. In contrast to the test calorimeter with liquid argon readout, here a scintillator readout was chosen, obviously not for reasons related to the concept of compensation. Meanwhile it is known that the combination uranium–scintillator offers better prospects for achieving compensation than uranium–liquid-argon, as will be elaborated in the following.

The development of compensation methods requires detailed knowledge of signal generation by the various energy contributions. Besides the contributions from the high-energy particles and the neutrons, significant contributions notably come from low-energy particles produced at the end of a cascade chain where they deposit energy through strong ionisation and photoeffect. Fundamental work in the 1980s essentially led to an understanding of compensation mechanisms, see [979, 258].

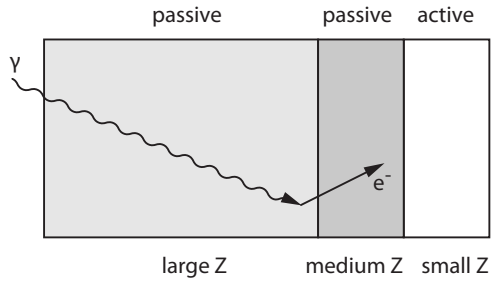
Compensation can be achieved through the specific construction of a calorimeter as well as by the proper choice of the materials for the passive and active layers and their thicknesses. But it is also possible to achieve compensation by software corrections of the deposited energy if the depositions are spatially resolved with sufficient accuracy. In the following, the different possibilities to achieve compensation will be introduced and discussed.

15.6.3.1 Hardware compensation

The ratio $e/h \approx 1$ yielding compensation can be obtained by design of the calorimeter. Since without targeted measures one usually gets $e/h > 1$, one attempts either to



(a) Calorimeter with a normal sequence of passive and active layers.



(b) Calorimeter with an intermediate medium-Z layer.

Fig. 15.37 Suppression of the electromagnetic signal component in a hadron calorimeter by inserting an intermediate layer with medium Z between the passive (high Z) and active layer (low Z) (schematic).

decrease ϵ_{em} or to increase ϵ_h, including a combination of both:

ϵ_{em} *decrease:* The discussion of e/mip in the previous section showed that for passive layers with high Z the detection of the electromagnetic component is less efficient than for lower Z; on page 638 we quoted the values for uranium and lead $e/mip \approx 0.6$ and for copper and iron $e/mip \approx 0.9$ (both with scintillator readout). This behaviour was attributed to the contribution of low-energy photons (below 1 MeV) which preferentially generate electrons in the passive layer where most of them are also absorbed without reaching the active layer. Thus to make a calorimeter compensating the Z of the passive layer should be high.

Photons preferentially interact in the passive layer because of the high Z. Electrons from the Compton or photoeffect only yield a signal if they are generated close to the transition to the active medium. The signal can even be further suppressed by covering the passive layers with a material of lower Z (fig. 15.37). The cover should be thick enough to efficiently absorb the electrons, while producing relatively less additional electrons than in the high Z material.

This effect was observed with the uranium–scintillator calorimeter of the ZEUS experiment (section 15.7.1) where for environmental reasons the depleted uranium had to be wrapped into a thin stainless steel foil (see fig. 3.29 in [982]). Systematic tests with 3.2-mm uranium layers and 2.6-mm scintillator layers showed that the largest suppression of e/mip is reached with 0.5 mm iron sheets.

ϵ_h *increase:* The fractions f_n, f_γ and f_B are determined by the nuclear spallation process and they are therefore strongly correlated. Here it is important for compensation that the ratios between the hadronic contributions are fairly energy independent, as shown in fig. 15.13 on page 599. Therefore the loss of binding energy can be compensated by efficiently recording neutrons and photons from nuclear reactions.

The elastic neutron scattering in the active medium turns out to be the best tunable process for compensation of hadron calorimeters yielding the best energy resolutions. Because of its fundamental relevance, this compensation method will be discussed in more detail in the following.

Compensation in passive layers with high nuclear masses. The high neutron production rate in passive layers with heavy nuclei, like uranium and lead, can be exploited for an enhancement of the hadronic signal. Neutrons can transfer their energy

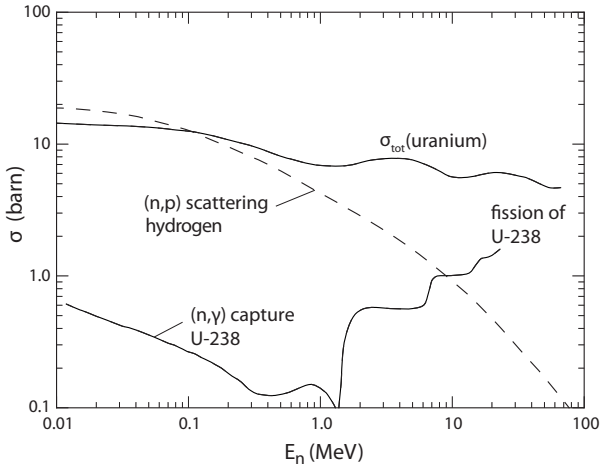


Fig. 15.38 Cross section for neutrons in ^{238}U (from [258], with kind permission of Elsevier). Besides the inelastic (n, γ) and the fission reactions primarily the elastic scattering with little energy transfer contributes to the total cross section. For comparison also the cross section for elastic neutron scattering in hydrogen is displayed where the energy can very efficiently be transferred.

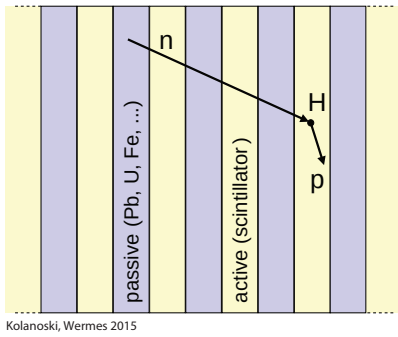


Fig. 15.39 Neutron detection by energy transfer to a proton in elastic (n, p) scattering in the active layer of a calorimeter.

to ionising particles by elastic scattering off light nuclei in the active medium where the ionisation energy is then detected.

For a neutron spectrum, as in fig. 15.10 on page 595, the cross section for elastic neutron scattering in uranium-238 lies in the range of some barns (fig. 15.38), yielding free path lengths of a few centimetres. With such path lengths neutrons can cross several passive and active layers. In the passive medium they do not lose much energy because the energy transfer depends on the nuclear mass number $\mathcal{A}$ as $1/\mathcal{A}$ according to (15.70). If now the active medium is enriched with hydrogen most of the energy can be transferred to the hydrogen nuclei, on average half of the neutron energy in one collision (fig. 15.39). Active media containing hydrogen are, for example, organic scintillators (section 13.2) or hydrogen-enriched gases.

An important point is the fact that the transferred energy will nearly completely be deposited in the active medium because the proton range is very short, at 1 MeV for example, only 20 µm in scintillator material. This effect, and the fact that neutrons cross several layers essentially losing energy only in the active medium, makes it possible to tune the e/h ratio by varying the sampling fraction f_S, the ratio of the energies deposited in the active and passive medium as defined by (15.39). For given passive and active media the sampling fraction depends on the ratio R of the geometrical thicknesses d_p and d_a of the passive and active layers, respectively:

Table 15.10 The table shows for uranium and lead calorimeters with scintillator readout the ratio R of the thickness of the passive to that of the active layer for which compensation, that is, $e/h = 1$, is reached (simulation calculations [982]). The scintillator thickness is always 2.5 mm. The corresponding layer thickness d_p and the measured intrinsic resolutions are also displayed. The last line reports the number of neutrons with energies below 20 MeV produced per GeV deposited energy in a high-energy cascade [258].

	Uranium	Lead
R	1 : 1	≈4 : 1
d_p [cm]	0.25	1.0
$\frac{\sigma_{intr}}{E}\sqrt{E/\mathrm{GeV}}$	22%	13%
neutrons/GeV	33	22

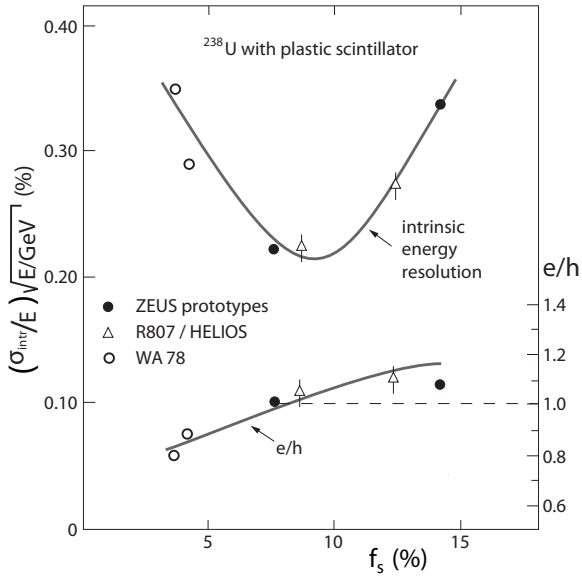


Fig. 15.40 Intrinsic energy resolution as defined in (15.74) (left axis) and e/h ratio (right axis) of uranium–scintillator calorimeters as a function of the sampling fraction f_s. Adapted from [979], with kind permission of Elsevier.

$$R = \frac{d_p}{d_a} \propto \frac{1}{f_S}\,. \tag{15.72}$$

The signal that is due to the electromagnetically deposited energy becomes smaller with increasing R because the total length of ionising tracks in the active medium becomes smaller. In addition the effect is over-proportionately amplified because the low-energy photons produced at the end of a shower are, as described before, preferentially absorbed by photoeffect in the material with high Z and the electrons produced mostly do not reach the active medium. Therefore, by varying R one changes the efficiencies ϵ_{em} and ϵ_n, thus also ϵ_h, relative to each other (see also eq. (15.61) on page 635). In this way $\epsilon_{em}/\epsilon_h = e/h = 1$ can be reached. Table 15.10 shows for lead and uranium calorimeters with scintillator readout examples for R values for which the targeted $e/h = 1$ is reached.

A variation of R changes both e/h and also the sampling fraction f_s. With the sampling fraction $f_s = E_{vis}/E_{dep}$ and approximating the total deposited energy E_{dep}

by the primary E, we obtain the energy deposited in the active layer as $E_{vis} = f_s\, E$. Then the sampling contribution to the relative energy resolution is

$$\frac{\sigma_{samp}}{E} \propto \frac{1}{\sqrt{f_s\, E}} \propto \sqrt{\frac{R}{E}}\,, \tag{15.73}$$

where the proportionality $1/\sqrt{f_s} \propto \sqrt{R}$ only holds for one specific combination of passive and active media. For further studies of compensation it is convenient to define an 'intrinsic calorimeter resolution' by correcting the measured resolution by the sampling contribution:

$$\sigma^2_{intr} = \sigma^2 - \sigma^2_{samp}\,. \tag{15.74}$$

Figure 15.40 shows examples of intrinsic resolutions as a function of f_S, simulated for different calorimeter assemblies [979]. The e/h ratio obtained is shown on the right vertical axis. Compensation, $e/h = 1$, and thus optimal resolution is achieved for a certain f_S value. For the uranium–scintillator combination in fig. 15.40 one finds an optimal intrinsic resolution of about $22\%/\sqrt{E}$, the corresponding intrinsic resolution for a lead–scintillator combination is even $13\%/\sqrt{E}$. The better value for lead can be explained by the closer correlation of the neutron number to the binding energy than in uranium. The reason is that in lead less neutrons originate from the subsequent fissions, which are less connected to the break-up of nuclei in the particle cascade [982]. In fact, in the subsequent nuclear fissions the predominant part of the energy turns into the kinetic energy of the fragments, which deposit their energy nearly exclusively in the passive medium and thus do not contribute to the compensation of the binding energy.

The resolutions for a 3.2 mm uranium absorber with 3.0 mm scintillator readout and for 10 mm lead absorber with 2.5 mm scintillator readout, in both cases corresponding to the R value required for the optimal intrinsic resolution, are measured to be around $35\%/\sqrt{E}$ and $44\%/\sqrt{E}$, respectively [364]. Thus for uranium the intrinsic resolution has about the same magnitude as the sampling contribution, whereas for the same thickness of the active layer the resolution of a lead calorimeter (with $e/h = 1$) is dominated by the sampling fluctuations. To make a lead–scintillator calorimeter compensating requires a thickness ratio of R = 4:1 according to table 15.10. For a scintillator thickness of 2.5 mm this corresponds to a sampling fraction $f_s^{Pb} \approx 3\%$ (for ^{238}U one obtains $f_s^U \approx 10\%$ at the required R = 1:1) [980]. An improvement of the energy resolution can only be achieved by increasing the sampling frequency. In order to obtain at the same time compensation the scintillator layers have to be made thinner, but for technical reasons a scintillator thickness below about 2 mm does not seem possible.

For lead this problem can be circumvented. With thin scintillating fibres as active medium, an $R = 4:1$ ratio and simultaneously a continuous readout along the shower development can be achieved. For this purpose about 1 mm thick fibres are laid in the longitudinal direction into the lead, as in fig. 15.29(b) on page 620, making it possible to optically readout from the back of the calorimeter. With such a device the SPACAL Collaboration [37, 38] achieved a suppression of the shower fluctuations by about a factor of four relative to the values quoted above for a sandwich assemply. Measurements of the corresponding energy resolutions for electrons and pions yielded [981]

$$\left(\frac{\sigma_E}{E}\right)_e = \frac{12.9\%}{\sqrt{E/\mathrm{GeV}}} \oplus 1.2\% \qquad \text{and} \qquad \left(\frac{\sigma_E}{E}\right)_\pi = \frac{27.7\%}{\sqrt{E/\mathrm{GeV}}} \oplus 2.5\%\,. \quad (15.75)$$

Delayed signals for tuning e/h. Hadronic cascades develop on a time-scale of a few nanoseconds. For recording of the signals the characteristic times of the readout detectors have to be added; for scintillators these are also a few nanoseconds and for liquid argon readout some 100 ns.

In addition to the main component of a cascade, delayed energy depositions also occur, preferentially from gamma transitions of excited nuclei and from thermalised neutrons. Typical delays last up to about 1 µs. Since these delayed contributions are also correlated to the binding energy loss it is in principle possible to influence the e/h ratio by extending the integration time of the signals. On the other hand, the extension of the integration time increases the background contribution, in particular from radioactive decays of uranium. In practice it was found that the related degradation of the energy resolution is not compensated by the gain in e/h (fig. 3.23 in [982]).

Construction criteria for achieving $e/h = 1$. As we have seen there are different possibilities to influence the ratio $e/h = 1$ when constructing a calorimeter. The most important construction parameters and their impact are as follows:

Atomic number Z of the passive medium: Complete compensation can be best achieved with high atomic number Z of the passive medium, as for example with uranium, lead and tungsten. The reason is the relative suppression of the electromagnetic signals and the high production yield of neutrons, which can be exploited for an adjustable compensation of the binding energy losses.

Hydrogen in the active medium: If a high neutron yield is to be efficiently exploited the active medium should contain hydrogen because neutrons can most effectively transfer their energy to those light nuclei which then, due to being charged, contribute through ionisation to the signal.

Ratio of the thickness of the passive to the active layer: For a specific active medium with fixed hydrogen content (e.g. scintillator) the value of e/h can be adjusted by the relative thickness of the active and passive layers.

One should always experimentally verify that full compensation, that is, $e/h = 1$, can indeed be reached.

Pros and cons of hardware compensation. Until now the best resolutions for high-energy hadrons and jets have been achieved with compensating calorimeters. Here the most prominent examples are the SPACAL prototype with lead–scintillator structures (achieved resolutions in (15.75)), which, however, has never been deployed in a running experiment, and the uranium–scintillator calorimeter of the ZEUS experiment (see table 15.11), which has been operated in a large detector. However, as yet only few completely compensating calorimeters have been built. For example, none of the LHC experiments has chosen a compensating calorimeter. There are two essential reasons for this fact:

Neutron diffusion: The diffusion of the low-energy neutrons broadens the showers. On the other hand, as much as possible of the neutron signals has to be included in the shower detection in order to reach compensation. Therefore the signals have to be collected within an enlarged region around the shower axis, leading to an increased background and thus to a degradation of particle and jet separation, in particular

at high particle fluxes.

Electromagnetic resolution: The compensation concept requires the same detector technology for the electromagnetic and hadronic calorimeter components, which usually follow each other in shower direction. This restriction prevents an optimal electromagnetic resolution from being reached. For the construction of the ATLAS and CMS detectors, for example, it was decisive that the best resolutions for photons should be reached, optimised mainly for the detection of the two-photon decay of a Higgs boson with small mass (120–130 GeV).

Further reasons in favour of a non-compensating calorimeter can be costs and, for uranium calorimeters, also environmental aspects.

15.6.3.2 Software correction

In several experiments it was shown that resolution and linearity can be improved by applying software corrections to the energy depositions, also for calorimeters that are not a priori compensating by construction (usually with $e/h > 1$). The corrections exploit the fact that high local energy densities are mainly caused by electromagnetic depositions because the hadronic energy depositions have wider spatial distributions. In a calorimeter with good readout segmentation the energy depositions in single cells can be weighted for each individual shower such that the signals from the electromagnetic and hadronic components are on average equal. In this way resolution and linearity are improved. However, the figures of calorimeters with good intrinsic compensation will not be reached, because the energy depositions have statistical fluctuations and thus the software correction can only be correct on average.

Amongst others, the weighting method was developed for the iron–scintillator calorimeter of the CDHS experiment (neutrino experiment at CERN) [31]. The weighting formula,

$$E_i' = E_i \left(1 - C \frac{E_i}{\sqrt{E_{tot}}}\right), \tag{15.76}$$

contains the unweighted energies E_i and E_{tot}, measured in a cell i and in all cells, respectively, and E_i' is the corrected energy. The weighting term $C/\sqrt{E_{tot}}$ takes into account that the correction becomes smaller with increasing energy. The parameter C is empirically determined such that the resolution becomes optimal. Figure 15.41 shows that for hadrons the resolution is improved by about 40% after the weighting and that the corrected energy distribution follows a normal distribution about the expectation value very well.

15.6.3.3 'Particle flow' concept

In the framework of studies for a detector at a high-energy electron–positron collider (at the time of writing 0.5–3 TeV are discussed for the projects 'International Linear Collider (ILC)' [555] and 'Compact Linear Collider (CLIC)' [314]) a concept was developed with the goal to reach the highest resolutions for jets: the 'particle flow analysis' (PFA). The concept aims at making the resolutions of two-jet decays of vector bosons comparable with or smaller than their mass widths, requiring an energy resolution of better than $\sigma_E/E \approx 3$–4% above $E_{jet} \approx M_Z/2$. The necessary hard- and software is currently under development and experimentally tested by the CALICE Collaboration (see e.g. [52, 262]).

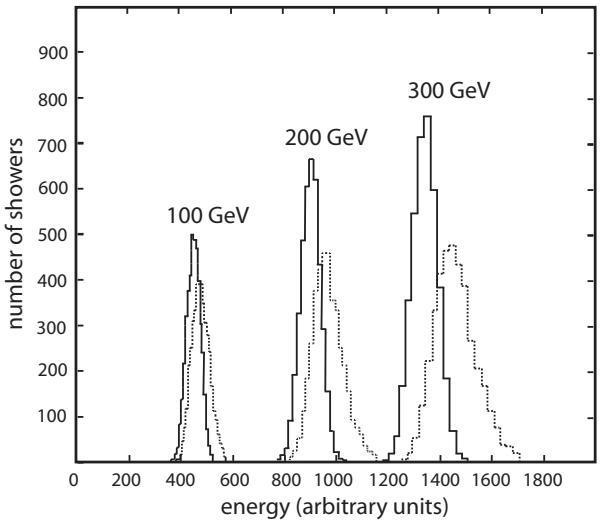


Fig. 15.41 Software compensation. Distributions of recorded energies (dotted histograms) of mono-energetic pions measured with an iron–scintillator calorimeter [31]. The solid line histograms represent the same distributions after applying weighting according to (15.76) [218]. Source: [613], with kind permission of Springer Science+Business Media.

In principle the idea of PFA is an advancement of the weighting method with particularly high calorimeter segmentation such that energy clusters in the calorimeter can be assigned to single particles. A first step is the linking of associated clusters to charged particles that are reconstructed with tracking detectors and of the electrons and photons that were recognised as electromagnetic showers. In this way most of the deposited energy is already captured since for jet energies below 250 GeV the contributions to the deposited energy are about 60% from charged particles, 30% from photons and 10% from neutral hadrons.

Simulation studies show that $\sigma_E/E \approx 3.4\%$ for jet energies of 45–250 GeV can be reached [940]. For lower energies ($E_{jet} \approx M_Z/2$) this is significantly better than reached with current calorimeters. However, for this method the relative energy resolution is not proportional to $1/\sqrt{E}$, because with increasing jet collimation at higher energies it becomes more difficult to correctly associate the particles to clusters. For $E_{jet} = 250\,\text{GeV}$, for example, the resolution is estimated to be about $50\%/\sqrt{E/\text{GeV}}$.

15.6.3.4 Dual readout

The differing detection efficiencies ($e/h \neq 1$) and the fluctuations of the electromagnetic and hadronic components of a hadron shower are the main reasons for the limited energy resolution of hadron calorimeters. The hitherto discussed methods to overcome these limitations always have certain disadvantages:

Hardware compensation ($e/h = 1$): As a rule the optimisation of compensation leads to relatively bad resolution for purely electromagnetic showers; the signal has to be integrated over a large calorimeter volume if diffusing neutrons contribute to compensation.

Software correction: The weighting algorithms do not reach the optimal resolution that is possible with hardware compensation.

Particle flow: The separate detection of particles in jets is no longer possible at energies above some 100 GeV.

To overcome these limitations the so-called 'dual readout' has been proposed, which implies that the signals from the electromagnetic and hadronic components of the shower are recorded by different readout schemes. The concept is based on the observation that the energy deposition in electromagnetic showers is mainly due to relativistic electrons and positrons, which emit Cherenkov light, whereas this contribution amounts to only about 10–15% in the hadronic component. A combination of readout techniques with unequal sensitivities for relativistic and non-relativistic particles, like the simultaneous detection of Cherenkov radiation and of ionisation, allows for a separation of both components. The total energy,

$$E = E_{em} + E_h \,, \tag{15.77}$$

can be determined by solving the following equations for the components E_{em} and E_h [1001]:

$$\begin{aligned} S_c &= \eta_c^h \, E_h + \eta_c^{em} \, E_{em} \,, \\ S_i &= \eta_i^h \, E_h + \eta_i^{em} \, E_{em} \,. \end{aligned} \tag{15.78}$$

In these equations $S_{c,i}$ are the measured Cherenkov and ionisation signals, respectively; the η factors with the respective indices determine the relative contributions of the deposited energies E_{em} and E_h to the corresponding signals. The system of equations (15.78) is obviously unambiguously solvable for E_{em} and E_h only if the ratio of the η factors for both components is not equal (corresponding to the requirement that the determinant does not vanish).

The concept has been studied in the framework of a research and development project, called 'DREAM' (*Dual REAdout Method*), see for example [983]. Tests included readout systems according to the SPACAL principle (see section) with simultaneous readout of scintillating fibres for the detection of ionisation and of non-scintillating fibres for the detection of the Cherenkov light [983]. Also considered are homogeneous calorimeters which employ materials offering simultaneous measurement of both Cherenkov light and ionisation, such as heavy liquid noble gases. With such an arrangement one could come close to the theoretical resolution limit of about $15\%/\sqrt{E}$ for hadrons, which is governed by the intrinsic shower fluctuations within both components [983].

15.6.4 Calibration of hadron calorimeters

Operating hadron calorimeters one mostly has to rely on calibration with test beams, even more than for electromagnetic calorimeters. Usually this is only possible before the installation of modules into the detector and then only exemplarily. Therefore one has to find calibration processes which can be run during data taking or sometimes also in special calibration runs.

For hadron calorimeters, mostly in combination with electromagnetic calorimeters, the energy scale of jets (JES) is an important quantity. It can be determined by combining measurements of single particles at test beams, the extrapolation to jets by means of simulations and the use of kinematical constraints in the experiments. At the electron–proton collider HERA the kinematic relation between the scattered electron and the hadronic final state could be used. At high-energy hadron colliders the hadronic decays of the vector boson $W^{\pm}$ and Z^0 with two jets in the final state offers very convenient possibilities for calibration. A precision of the JES of about 5% can be reached.

15.6.5 Energy resolution of hadron calorimeters

We have seen that hadronic showers are in general measured with much worse energy resolution than electromagnetic showers (see e.g. fig. 15.34). This is due to the following:

- the diversity of contributions to the energy depositions with large fluctuations in the partitioning;
- the generally different efficiencies for different shower components;
- the relatively large fraction of non-visible energy (binding energy and neutrinos);
- the potentially larger leakage losses due to the hadronic absorption length being much larger than the radiation length and due to the stronger fluctuations in the shower profile.

The difference between the detection efficiencies of the electromagnetic and the hadronic fractions is important. Therefore the best energy resolutions have been obtained with compensating calorimeters having the ratio of both efficiencies, e/h, close to 1.

Usually a detector at an accelerator features an electromagnetic calorimeter in front of the hadronic one. Since the depth of an electromagnetic calorimeter is in the order of magnitude of a hadronic absorption length the hadron shower can start in the electromagnetic section. Therefore the resolutions of hadrons have to be determined for the combination of both calorimeter parts. In compensating calorimeters both parts should form an entity with the same mechanical and electronic properties.

Furthermore one has to consider that hadron calorimeters mostly serve in the energy measurement of jets. In test beams the absolute energy scales and resolutions for single particles, for example for electrons and pions, are usually determined. Starting from such measurements the corresponding scales and resolutions for jets are derived by means of simulations. However, such extrapolations have then to be experimentally verified, for example by analysing over-constraint kinematics. In experiments at high energies the precise knowledge of the JES is one of the most important prerequisites for the analysis of data containing hadronic final states (section 15.6.4).

The energy dependence of the resolution can in general be described by three quadratically added terms, as given in (15.33) for the case of electromagnetic calorimeters:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\,. \tag{15.79}$$

As for electromagnetic calorimeters the dominant sources for the single contributions are (labelled by their coefficients a, b, c): (a) stochastic fluctuation, (b) electronic noise and (c) various imperfections in the realisation of the calorimeter. See also the general discussion of resolutions in section 15.4.3.

Stochastic contribution. As already introduced in (15.74), when discussing compensation, the stochastic term can be divided into a term for the intrinsic calorimeter resolution and a term for the sampling resolution:

$$\left(\frac{\sigma_{stoch}}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 = \left(\frac{a_{intr}}{\sqrt{E}}\right)^2 + \left(\frac{a_{sampl}}{\sqrt{E}}\right)^2 . \tag{15.80}$$

The intrinsic resolution has been discussed in section 15.6.3 and the optimally reachable values for uranium and lead are listed in table 15.10.

The sampling resolution can usually be given as a function of the energy ΔE_{mip} deposited by a minimum-ionising particle in a sampling layer [364]:

$$\frac{a_{sampl}}{\sqrt{E}} = 11.5\% \, \frac{\sqrt{\Delta E_{mip}/\mathrm{MeV}}}{\sqrt{E/\mathrm{GeV}}} \, . \tag{15.81}$$

For electromagnetic calorimeters a corresponding dependence is given in (15.54) with the difference that there $t\,E_c$ instead of ΔE_{mip} parametrises the energy loss per layer.

In principle the sampling resolution can be nearly arbitrarily improved by increasing the sampling frequency or, equivalently, decreasing the thickness of an absorber layer. In practice, however, this is limited by costs and in the case of compensating calorimeters by the compliance with the correct thickness ratios yielding $e/h = 1$ (see table 15.10). In general one would try to assure that the sampling resolution does not dominate over the intrinsic resolution.

Noise. The contributions to the noise term come from electronic sources and possibly also from an instrumentally or experimentally conditioned background. For example, in calorimeters with liquid argon readout electronic noise dominates, which is proportional to the input capacitance of the readout channel (see section 17.10). By contrast, in uranium–scintillator calorimeters fluctuations due to radioactive decays tend to dominate. The size of the contribution of noise to the resolution strongly depends on the number of summed individual channels. Therefore, in reconstructing showers one should take care that the number of contributing channels is restricted to a minimum (e.g. by requiring that the height of a signal should exceed a certain multiple of the width of the noise distribution). For example, reconstructing a hadronic shower in the liquid argon calorimeter of H1 the noise contribution is about 900 MeV, mainly due to electronic noise [98,35], whereas in the uranium calorimeter of ZEUS it is less than 500 MeV, mainly from radioactive decays [1020]. For scintillator readout of iron calorimeters, as employed in the LHC experiments ATLAS and CMS, the typical noise contribution is about 50 MeV per readout channel, summing up to numbers in the GeV range for jets.

In the LHC experiments the background from particles contributes in a similar way to electronic noise. The dominant background comes from soft inelastic reactions where each single particle has an average transverse momentum of about 500 MeV. At design luminosity with about 25 reactions per beam crossing, this background piles up yielding about 1 TeV total energy in the whole calorimeter, roughly evenly distributed in the rapidity η.

Constant term. The energy-independent part of the relative energy resolution is caused by mechanical tolerances, intercalibration errors, lacking compensation, leakage losses and others. This constant term is typically of the order of a few per cent. Since it limits the resolution at high energies one tries to minimise it. For example, for a stochastic term of 35% and a constant term of 2% the constant contribution already dominates above 300 GeV. Halving the constant term to 1%, it would dominate only above the fourfold energy, hence above 1.2 TeV.

In hadron calorimetry leakage losses are particularly important because on the one hand the absorption lengths are large and leakage losses considerably fluctuate, while on the other hand the length of a calorimeter is limited for practical and cost reasons. In practice a calorimeter has hardly more than about 7 absorption lengths, yielding for iron about 1.2 m for the passive layers only.

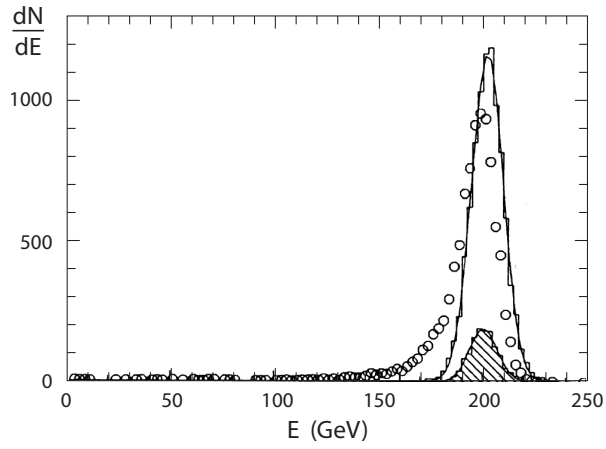


Fig. 15.42 Leakage loss and resolution: distributions of reconstructed energies of 205-GeV pions measured with a test module of the liquid-argon–iron calorimeter of the H1 detector (adapted from [35], with kind permission of Elsevier). In this module the liquid argon calorimeter with a total of 6 absorption lengths was followed by a tail catcher with additional 4.5 absorption lengths. Displayed are the energy distributions for reconstruction without tail catcher (open points), with tail catcher (open histogram) and for the case that the shower is fully contained in the calorimeter, without any energy deposition in the tail catcher (hatched histogram). The plotted curve is a normal distribution corresponding to the resolutions given in table 15.11.

The fluctuations can be substantially reduced by a relatively coarse measurement of the leakage losses with a tail catcher (section 15.4.2.1), as demonstrated in the example of fig. 15.42. The figure shows that leakage losses lead to an asymmetric energy distribution with long tails towards lower energies (open points, reconstruction with signals from the calorimeter only). The full resolution can be reached either by requiring that no energy is seen in the tail catcher (hatched histogram) or that the energy measured in the tail catcher is included in the reconstruction (open histogram). Obviously, only by taking the leakage losses into account is a Gaussian resolution curve achieved. An energy deposition in the tail catcher can also be used as a veto in order to obtain a better resolution for the remaining showers (hatched histogram in fig. 15.42).

The leakage fluctuations are strongly correlated with the starting point of the shower development. In section 15.4.2 we had already pointed out that a measurement of the location of the first interaction allows us to better estimate the leakage losses and thus to improve the energy resolution.

Examples. Table 15.11 shows examples of resolutions of different calorimeter systems. Non-compensating hadron calorimeters feature energy resolutions of more than $50\%/\sqrt{E}$ to be compared to the energy resolutions achieved with a compensating hadron calorimeter (see section 15.6.3):

$$\begin{aligned}
\text{Pb/scintillator (sandwich):} &\quad \tfrac{\sigma_E}{E} \approx 45\%/\sqrt{E/\text{GeV}} + 1\%\,, \\
\text{Pb/scintillator (SPACAL):} &\quad \tfrac{\sigma_E}{E} \approx 28\%/\sqrt{E/\text{GeV}} + 2.5\%\,, \\
\text{U/scintillator:} &\quad \tfrac{\sigma_E}{E} \approx 35\%/\sqrt{E/\text{GeV}} + 1\%\,.
\end{aligned}$$

Table 15.11 Properties of calorimeter systems in experiments at the colliders LEP (e^+e^-), SLC (e^+e^-), HERA ($e^\pm p$), Tevatron ($p\bar{p}$) and LHC (pp), in each case for the electromagnetic (EM) and the hadronic (HAD) part of the central calorimeter. The energy resolutions are listed for electrons and photons ('EM') as well as for hadrons (combination of 'EM' and 'HAD'; single particles or jets). The e/h values are given after weighting corrections if applicable. The resolution numbers are only valid as examples because they can widely vary depending on the specific calorimeter region and the applied selection and reconstruction algorithms.

Experiment & reference	Cal.	Structure	e/h	Resolution $\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c$ a (% $\sqrt{\text{GeV}}$)	b (MeV)	c (%)
ALEPH	EM	Pb/PWC		18	n. s.	1.9
[346, 264]	HAD	Fe/LST	n. s.	85	n. s.	n. s.
DELPHI	EM	Pb/TPC		23	n. s.	4.3
[11, 33]	HAD	Fe/LST	n. s.	120	n. s.	n. s.
L3	EM	BGO		2.2	n. s.	0.7
[50, 253]	HAD	U/PWC	n. s.	55	n. s.	n. s.
OPAL	EM	Pb glass		6.3	11	0.2
[63]	HAD	Fe/LST	n. s.	120	n. s.	n. s.
SLD	EM	Pb/LAr		15	n. s.	n. s.
[136, 22]	HAD	Pb/LAr, Fe/LST	n. s.	60	n. s.	n. s.
ZEUS	EM	U/scin.		18	n. s.	n. s.
[1020, 96]	HAD	U/scin.	1.00	35	< 500	2.0
H1	EM	Pb/LAr		11	250	1.0
[35]	HAD	Fe/LAr	1.4	51	900	1.6
CDF	EM	Pb/scin.		14	n. s.	n. s.
[21]	HAD	Fe/scin.	n. s.	80	n. s.	n. s.
D0	EM	U/LAr		16	n. s.	0.3
[13]	HAD	U/LAr	1.08	45	1300	4.0
CMS	EM	$PbWO_4$		2.8	120	0.3
[298]	HAD	brass/scin.	1.40	125	560	3.0
ATLAS	EM	Pb/LAr		10	245	0.7
[4]	HAD	Fe/scin.	1.30	56	1800	3.0

Abbreviations: EM, HAD: electromagnetic, hadronic calorimeter; PWC: proportional wire chamber; LST: limited streamer tubes; LAr: liquid argon; n. s.: not specified.

In various projects new concepts are developed to surpass these resolutions, such as 'particle flow analysis' (section 15.6.3.3) or 'dual readout' (section 15.6.3.4).

15.7 Calorimeter systems for measurements of hadrons and jets

While electromagnetic calorimeters for the measurement of electrons, positrons and photons can typically be employed at relatively low energies (starting from the MeV range for photons), hadron calorimeters only become important for energies of hadrons and jets above about 50–100 GeV. Actually, for the electron–positron storage rings of

the 1980s, PETRA, PEP and TRISTAN (table 2.2 on page 13), with centre-of-mass energies up to about 60 GeV, hadron calorimeters did not play any role. At these energies hadron calorimeters do not yield a better resolution for jets than can be achieved through measurement of the momenta of charged particles and the energies of photons.[9] In contrast, all experiments at the electron–positron collider LEP, with centre-of-mass energies up to about 200 GeV, featured hadron calorimeters, though with relatively bad resolution because their optimisation was not the central goal. Hadron calorimetry was more vigorously developed at hadron accelerators and colliders like PS (proton synchrotron, CERN), ISR, SPS, HERA, Tevatron and LHC (table 2.2). In particular, experiments at high-energy neutrino beams belong to the protagonists of such detectors. In such experiments the nucleon structure has been investigated in charged and neutral currents:

$$\nu + \text{nucleus} \to \mu + X \qquad \text{respectively} \qquad \nu + \text{nucleus} \to \nu + X\,. \tag{15.82}$$

Since neutrinos are practically not observable it is pivotal for the kinematical reconstruction of the processes to have a precise measurement of the hadronic final state X.

In the following we want to present some examples of calorimeter systems.

15.7.1 HERA: calorimetry at H1 and ZEUS

The electron–proton storage ring HERA in Hamburg was operated until 2007 at beam energies of nominally 920 GeV for protons and 27.5 GeV for electrons or positrons. The unequal energies required asymmetric detector configurations optimised for high particle energies in proton direction ('forward direction'). With the time between beam crossings being 96 ns the signal processing for the correct assignment of the crossing time and for the triggering posed a certain challenge. A detailed description of the HERA machine and physics programme can be found in [670] and a compact summary thereof in [995].

The most important goal of the HERA experiments H1 and ZEUS was the measurement of the proton structure in deep-inelastic electron–proton scattering,

$$e + p \to e' + X\,. \tag{15.83}$$

The typical topology of the hadronic final state X consists of a quark jet, possibly accompanied by gluon radiation, and the remnants of the dissociated proton which are heading into the proton direction and mostly disappear in the beam pipe (fig. 15.43). The reconstruction of the hadronic system with typical jet energies around 100 GeV requires hadron calorimetry with good energy and direction resolution with a tight coverage as close to the beam as possible. For the reconstruction of the scattered electron a corresponding electromagnetic calorimetry is necessary. The partially redundant measurements of the electron and the hadrons could successfully be used for calibration of the calorimeter, in particular for fixing the jet energy scale (JES) (section 15.4.7). Energy resolutions of both experiments are listed in table 15.11.

The ZEUS [96,350] was conceived such that compensation, that is, $e/h = 1$, should already be achieved at the hardware level. Based on the experience from the uranium–scintillator calorimeter of the Axial Field Spectrometer (R807) at the Intersecting

[9] An exception is the detection of the neutral hadrons K_L, n and $\bar{n}$, for which at low energies, for example at B-factories, muon systems have already been used as coarse calorimeters.

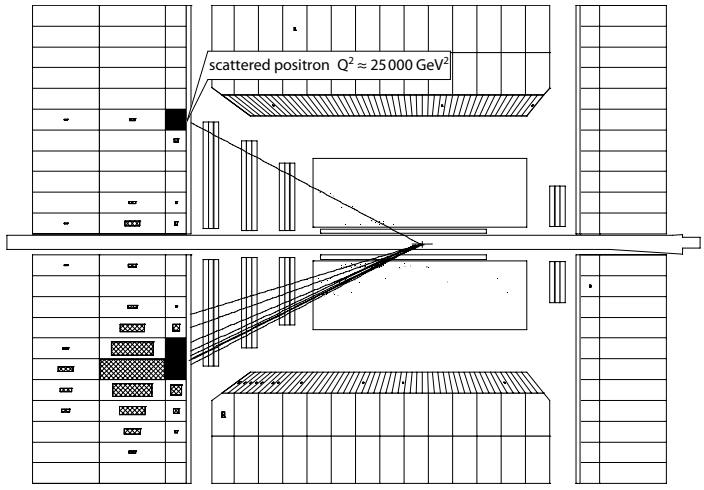


Fig. 15.43 Cut through the ZEUS detector with calorimeter and inner tracking chambers (the muon system is not displayed) [1020] showing a deep-inelastic electron–proton scattering event with particularly high transverse momentum. The calorimeter is divided in three parts which cover the forward, central and backward regions, respectively. The forward–backward asymmetry of the calorimeter takes the kinematic at HERA into account. The inner calorimeter regions have a finer granularity for the measurement of purely electromagnetic showers. The different depths are optimised such that 90% of the jets with the highest kinematically possible energy deposit at least 95% of their energy in the calorimeter. The depths correspond at perpendicular incidence to 7, 5 and 4 absorption lengths in the forward, central and backward regions, respectively. In each case 1 absorption length belongs to the electromagnetic section.

Storage Ring (ISR) at CERN [70], the decision was made in favour of a sampling calorimeter with depleted uranium (^{238}U) as passive layer and 2.6 mm thick scintillator in order to reach compensation (section 15.6.3).

Figure 15.43 shows a longitudinal cross section of the calorimeter and the tracking detectors inside the calorimeter. The different components, which have a larger depth in the forward direction of the proton, almost completely surround the interaction point. The calorimeter comprises in all regions an electromagnetic section which has a finer segmentation than the hadronic section and can be separately read out. In the central and forward region the hadronic part is once more divided into two longitudinal readout sections. The readout proceeds through wavelength shifters which are sidelong attached to the modules and which guide the light to the photomultiplier (corresponding to the arrangement in fig. 15.27(b)).

In electron and pion test beams full compensation, that is, $e/\pi \approx 1$, could be shown for this calorimeter construction, yielding a very good energy resolution for hadrons and jets of

$$\frac{\sigma_E}{E} = \frac{35\%}{\sqrt{E/\mathrm{GeV}}} \oplus (1\text{–}2)\% \tag{15.84}$$

and linearity up to the highest energies occurring in the experiment. In addition, the readout scheme offers capabilities for electron–hadron separation, a direction resolution

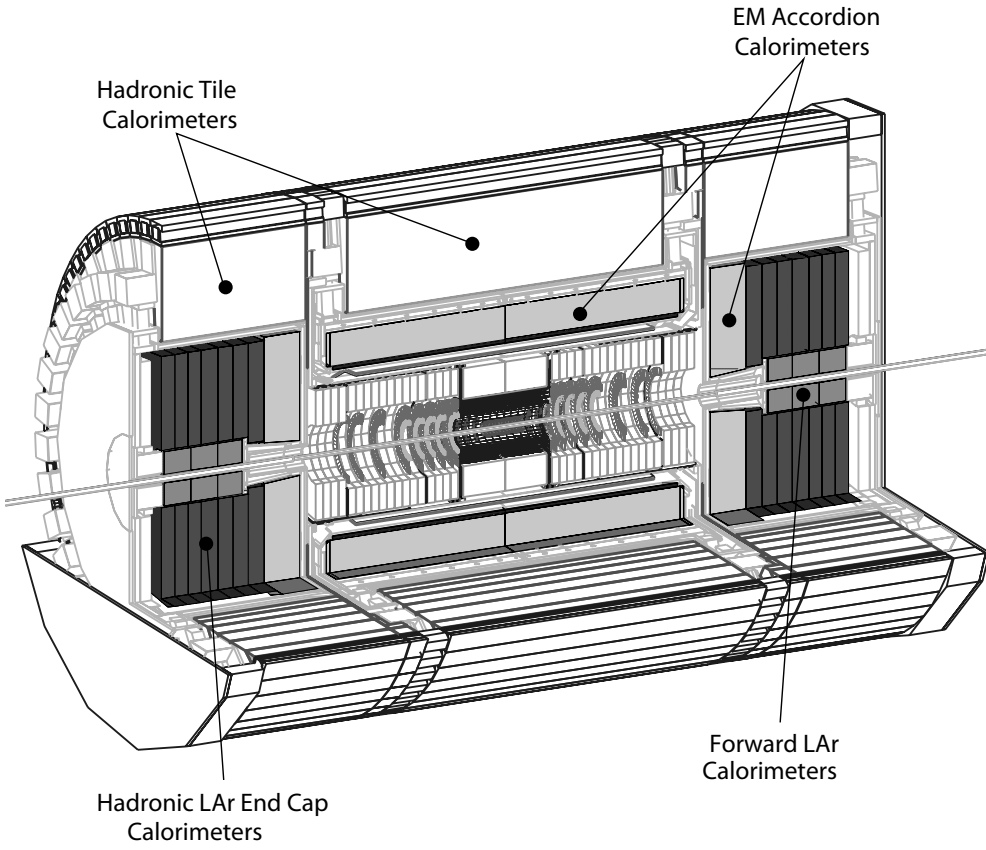


Fig. 15.44 The ATLAS calorimeter system [4], see the description in the text. Source: CERN/ATLAS Collaboration.

for jets of better than 10 mrad and a time resolution of about 1 ns due to the fast scintillator readout.

15.7.2 LHC: calorimetry at ATLAS and CMS

The Large Hadron Collider (LHC) was designed to reach centre-of-mass energies of 14 TeV. At these energies calorimetry plays a central role, in particular for the *general purpose detectors* ATLAS and CMS. At a time spacing between beam crossings of 25 ns and about 20 inelastic proton–proton interactions per crossing the calorimeters, together with the muon system, are particularly suited to provide fast triggers for high-energy particles, jets and missing transverse energy (the latter being an indication for non-interacting particles). The precise measurement and identification of electrons, photons, jets and non-interacting particles is a precondition for the discovery of 'new physics'.

As an example we want to elaborate on the ATLAS calorimeter system [4]. The different calorimeters of the ATLAS detectors (fig. 15.44) follow different construction principles. Most of the solid angle is covered by an electromagnetic sampling calorimeter with lead absorbers and liquid argon readout. This calorimeter with its specific accordion structure has been presented in section 15.5.3.2. The hadron calorimeters cover

the rapidity range (see section 2.5.3) of $|\eta| < 4.9$ with an iron–scintillator calorimeter in the central region, a copper–liquid-argon calorimeter in the endcap region and a forward calorimeter with copper/tungsten–liquid-argon sampling. The argon in the different calorimeter components has to be kept liquid in cryostats. The endcap and forward calorimeters, both hadronic and electromagnetic, are housed in a common cryostat, one in each endcap region. The central electromagnetic calorimeter has its own cryostat. At $\eta \approx 0$, that is, perpendicular to the beam, the hadron calorimeter has about 8 absorption lengths and the electromagnetic calorimeter adds about 1 absorption length.

The central hadron calorimeter, the so-called *tile calorimeter*, has three components (one *barrel* and two *extended barrels*, see fig. 15.44) which are cylindrically arranged around the beams. The passive layers are iron plates defining the mechanical structure which is radially oriented in the shower direction (perpendicular to the beams). The active layers are 3 mm thick scintillator *tiles* inserted into slots in the iron structure. The volume ratio of iron to scintillator is 4.7:1. This unconventional orientation of the scintillators allows for a simpler longitudinal and transversal segmentation with a particularly efficient readout through wavelength shifting fibres (see section 13.4.3). Segmentation is essential for a software compensation, as described in section 15.6.3.2, which yields the energy resolution [71]:

$$\frac{\sigma_E}{E} = \left(\frac{0.42}{\sqrt{E/\mathrm{GeV}}} + 0.018 \right) \oplus \frac{1.8}{E/\mathrm{GeV}} \,. \tag{15.85}$$

For this calorimeter the energy dependence of the ratio e/π (see section 15.6.2) was determined from electron and pion test beam measurements [54]. Using (15.62) this yielded the ratio $e/h \approx 1.36$ (see also fig. 15.35).

16

Detectors for cosmic particles, neutrinos and exotic matter

16.1 Introduction

In recent decades astroparticle physics has been established as an independent field for the investigation of the cosmic radiation [438].[1] The roots of this field go back to the discovery of cosmic radiation by Viktor Hess in the year 1912. During balloon flights (fig. 16.1) he realised that besides radiation emerging from the Earth, there is also an ionising radiation with an intensity increasing with altitude. In the course of further investigations it was found that the radiation is preferentially due to high-energy, positively charged particles which come from outer space, generating particle showers in interactions with the Earth's atmosphere (Pierre Auger, 1938; see [131] and references therein).

Up to the 1950s this particle radiation was the most abundant source for the discovery of new particles (see also section 2.1). The measurements were performed on high mountains and during balloon flights, mainly employing photoemulsions (section 6.3, fig. 16.2) and cloud chambers (section 6.1, fig. 15.26). The discovery of a continuously growing 'zoo of particles' in the cosmic radiation had boosted the research field of particle physics. However, since the 1950s particle accelerators, offering controllable experiments and high rates at continuously growing energies, became gradually more important for new-particle searches. But still today searches are performed in cosmic radiation for—mostly exotic—particles with names like WIMPS, axions or magnetic monopoles.

The main objective of astroparticle physics is to answer the question of origin and composition of the high-energy cosmic radiation. From outer space the Earth is reached by nucleons, nuclei, electrons, positrons, photons and neutrinos. The percentage of antimatter is low and consistent with being generated in secondary interactions with the interstellar matter. The fluxes of all components strongly decrease towards high

[1] An introduction to astroparticle physics can be found, for example, in the book by C. Grupen [487].

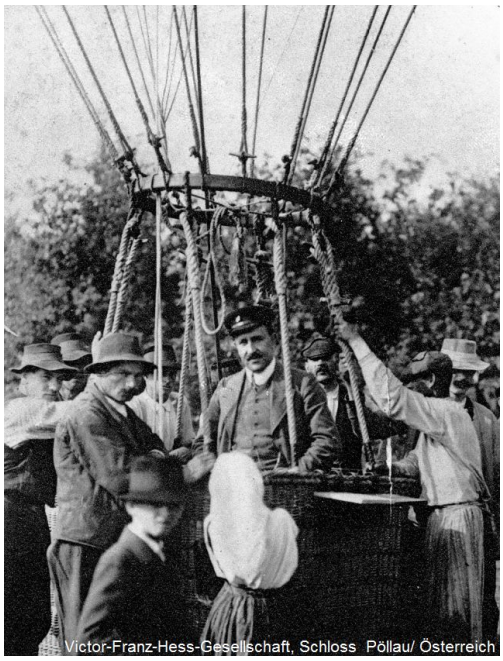


Fig. 16.1 The discoverer of the cosmic radiation, Viktor Hess (in the centre of the picture), after the landing of his balloon in Bad Saarow (near Berlin) on 7 August 1912 [873]. Source: Victor-Franz-Hess-Gesellschaft, Schloss Pöllau/ Austria.

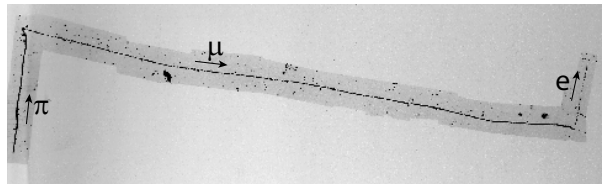


Fig. 16.2 Particle traces generated by cosmic rays in a photoemulsion (section 6.3). Decay of a charged pion into a muon and a neutrino; the muon further decays into an electron and two neutrinos ($\pi\mu e$ decay). All produced neutrinos remain invisible. Source: Powell [794], with kind permission of IOP Publishing.

energies, such that the detector dimensions have to grow accordingly with increasing energy in order to achieve sufficient event rates.

The low transparency of the Earth's atmosphere aggravates the detection of cosmic radiation for nearly all kinds of radiation. Only for neutrinos is the atmosphere nearly fully transparent due to their low interaction probability with matter which, however, for the same reason makes their detection very difficult. While for electromagnetic radiation windows exist in the visible and the radio ranges, primary charged particles of all energies have practically no chance to reach the Earth. In the range of low energies, for which the flux is high enough, particles can be directly measured by balloon or satellite experiments at the periphery or outside the atmosphere. If this is no longer possible at higher energies because of the low fluxes and the limited detector dimensions, one employs secondary interactions in the atmosphere for the detection of the interaction products.

Frequently only the charged component is denoted as 'cosmic rays'. The measured flux of the charged cosmic radiation shows a steep fall-off over the huge energy range

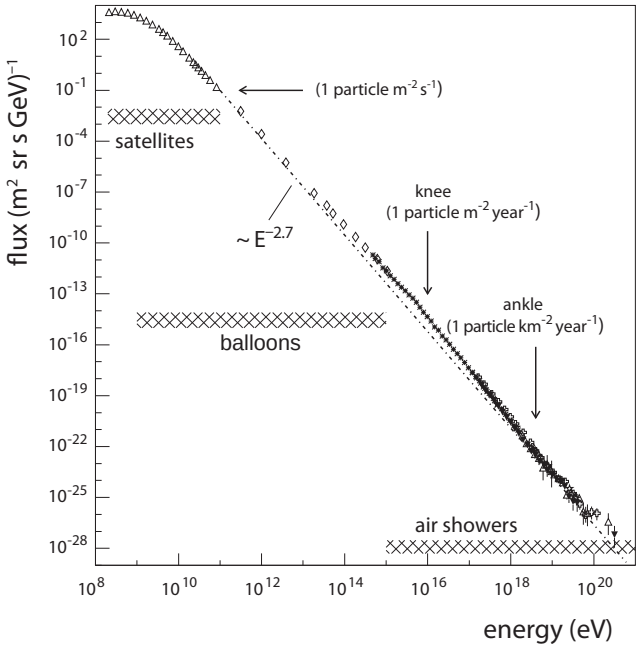


Fig. 16.3 Energy spectrum of the charged component of cosmic radiation. The dashed line is only drawn to guide the eye. Depicted are the ranges from which the data are collected predominantly by satellite, balloon or air shower experiments, respectively (adopted from [241]).

from 100 MeV to 10^{21} eV by about a factor 10^{30} (fig. 16.3). At low energies, up to about 10 GeV, the influence of the Earth's magnetic field leads to a shallower gradient because, depending on energy and direction, the radiation is deflected and may not reach the Earth. Towards higher energies the flux falls off very steeply with a distinct change of the slope above some 10^{15} eV, called the 'knee' and another slope change, the 'ankle' around 10^{19} eV before it breaks off around 10^{21} eV. The integrated rates are about $1/(\mathrm{m}^2\,\mathrm{year})$ above 10^{15} eV and about $1/(\mathrm{km}^2\,100\,\mathrm{year})$ above 10^{20} eV. While up to nearly 10^{15} eV direct measurements of the primary particles are possible with balloon and satellite experiments, at higher energies and lower fluxes the respective detector areas are no longer sufficient. In this range one exploits the fact that high-energy particles generate showers in the atmosphere which can be detected by large-area detectors on the Earth's surface.

In the last century the measurement of the electromagnetic component of cosmic radiation was extended from the visible spectrum to radio waves on one side and to gamma radiation on the other side, in total to a wavelength range between about 10 m and 10^{-21} m. Gamma radiation is measured with the aid of balloons and satellites. For the expected fluxes this is possible up to some 100 GeV. Above that, as in the case of charged cosmic radiation, the atmosphere is used as a calorimeter. In this case an electromagnetic shower develops, which is detected on the Earth's surface.

Neutrino physics received a boost with the discovery of oscillations amongst neutrino flavours and the non-vanishing masses of the neutrinos, which initiated many new experiments. The detection of neutrinos poses a big challenge because of their weak interactions with detector materials, which has to be overcome with different techniques for different energy ranges. The ranges can be divided as follows: neutrinos with keV to MeV energies, which originate from nuclear processes in stars or nuclear reactors or also from nuclear decays in the Earth; GeV neutrinos from the interaction of

cosmic radiation in the atmosphere (atmospheric neutrinos) and from neutrino beams at accelerators; high-energy neutrinos of cosmic origin with TeV to EeV energies.

In the following we first introduce the detection of cosmic radiation with relatively low energy above the atmosphere. Then we describe the properties of the atmosphere as a calorimeter and the corresponding detection instruments. The chapter is closed with a description of detectors for the search for dark matter. The search for dark matter is an example for experiments in particle physics which are not done at accelerators. Other non-accelerator experiments are for example the search for neutrino-less double β decay, for proton decay, for axions and magnetic monopoles. The detectors for the latter searches deserve special attention. In this book, however, they will not be discussed, mostly to limit the book's volume.

16.2 Balloon and satellite experiments

With balloon and satellite experiments for the investigation of cosmic radiation the absorption by the atmosphere can be largely avoided. However, the respective detector areas can be at most a few square metres. Because of the strongly falling fluxes (fig. 16.3) the energy range is effectively limited to a few 100 TeV.

Compared to space missions, balloon flights are relatively cost-efficient but have the disadvantage of a relatively short lifetime of the experiments, some days to at most few weeks. On satellites or space stations, in contrast, the duration of an experiment can be substantially extended, in the order of years.

16.2.1 Balloon experiments

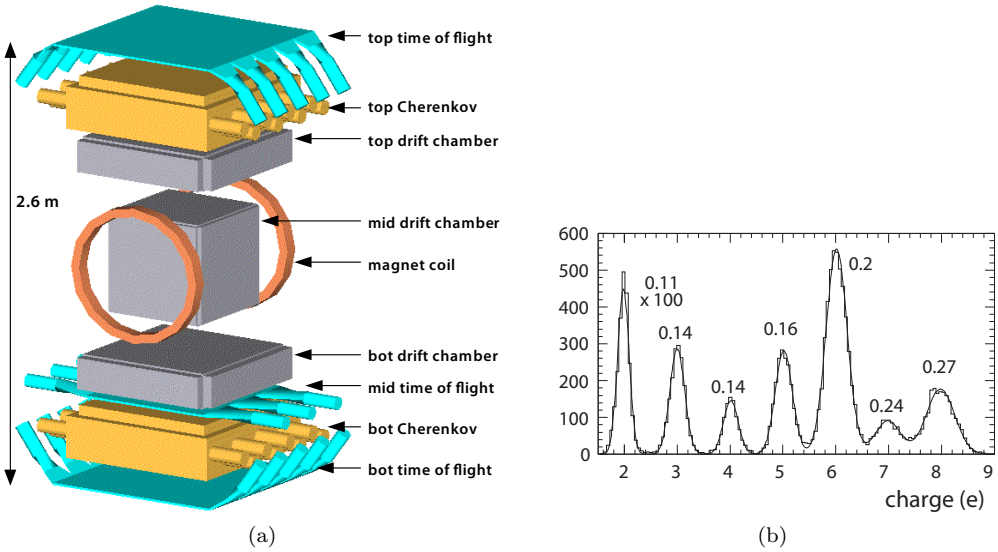
The balloons used for experiments (fig. 16.4) can have volumes up to a million cubic metres and carry payloads of up to 3 tons at altitudes of up to 40 km. On the ground the balloon envelope is only minimally filled and expands at high altitude to enormous volumes, for example at a pressure of 5 hPa by about a factor of 200. A balloon filled with helium under standard conditions yields a buoyancy of about $1\,\mathrm{kg/m^3}$, meaning that $1\,\mathrm{m^3}$ of helium bears a mass of 1 kg at ground level. For a payload of 2000 kg this requires a filling of $2000\,\mathrm{m^3}$ helium at ground level, corresponding to about $400\,000\,\mathrm{m^3}$ at high altitude at a pressure of 5 hPa.

In the achievable balloon altitudes the residual atmospheric depth[2] above the experiment is only 3–5 $\mathrm{g/cm^2}$. Comparing this figure with the mean free path of a proton in air of about 90 $\mathrm{g/cm^2}$ (in units of column density according to (3.28)), one realises that at balloon altitudes space conditions are almost fulfilled. The typical flight durations of such balloons reach from about 24 hours to few weeks. The longest flight durations are achieved by flights circumnavigating the South Pole. If a balloon is launched at the Antarctic coast station McMurdo it can almost reach the starting point after about 10–14 days due to the westward stratosphere current (fig. 16.4).

The composition of the incident cosmic radiation and the corresponding energy spectra were determined for the first time by balloon experiments. The charged component can be measured up to some 100 TeV. The separation of nuclei with different masses can be achieved by combining measurements of different observables, for example energy loss, range, time-of-flight, Cherenkov and transition radiation as well as the momentum (see also the discussion of particle identification in chapter 14).

[2]That is, the density integrated over the real geometrical depth given in units of mass per area, see section 16.3.1.

Fig. 16.4 Flight of the balloon experiment CREAM in Antarctica (from [887], with kind permission of Elsevier). Left: Launching the balloon the apparatus is moved on a vehicle following the floating balloon until the apparatus lifts off. Right: A flight track on which the CREAM apparatus has circled the South Pole three times in 37 days.



(a) (b)

Fig. 16.5 The balloon experiment ISOMAX [502]. (a) Schematic 3D view of the ISOMAX apparatus. The detector consists of a central drift chamber in a magnetic field which is surrounded symmetrically on top and bottom by time-of-flight counters, Cherenkov detectors and further drift chambers. (b) Determination of the charge of light nuclei by ISOMAX using time-of-flight and dE/dx measurements with scintillators. The charges are given in units of the elementary charge e. The numbers at the maxima denote the respective charge resolution. The first maximum at $2\,e$ (helium) is scaled down by a factor of 100. Source: NASA/T. Hams.

If the deployment of a magnet for momentum measurements is possible, matter and antimatter can also be distinguished. For the measurement of electron and gamma radiation, in general calorimeters are employed (chapter 15).

Figure 16.5(a) shows the typical size and complexity of a modern balloon experiment using the example of ISOMAX ('Isotope Magnet Experiment') [502]. The apparatus is 2.6 m high and has a mass of 2000 kg. A mass spectrometer with a superconducting magnet and drift chambers for measurement of the deflection is placed

in the centre of the experiment. Actually, from the measured radius of deflection ρ in a magnetic field B (section 9.1) only the momentum p per charge ze, denoted *rigidity* R, can be determined by the magnet spectrometer:

$$R = \frac{p}{ze} = \rho\, B. \tag{16.1}$$

Corresponding to the momentum unit eV/c the magnetic rigidity can be given in units of V/c. For the identification of the incident particles, that means for the determination of mass and charge, the spectrometer is supplemented by scintillation counters (chapter 13) and aerogel Cherenkov detectors (here $n = 1.14$, see chapter 11). The scintillation counters account for the measurement of energy loss (dE/dx) and time-of-flight (TOF). For particles that are accepted by the detector and can be reconstructed one defines the 'geometry factor' (G) as the integral over the product of the area element $d\vec{S}$ of the entrance window and the respectively accepted solid angle element $d\Omega$:

$$G = \int d\vec{S}\, d\Omega\,. \tag{16.2}$$

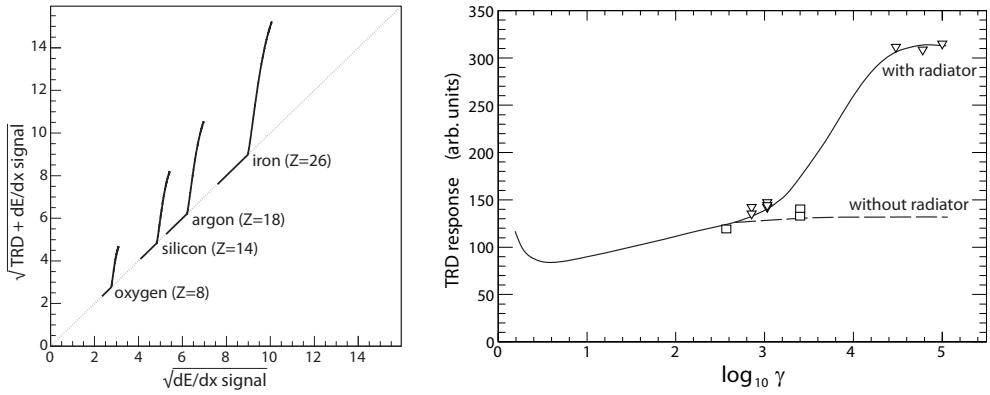
The geometry factor of ISOMAX is $G = 450\,\mathrm{cm}^2\,\mathrm{sr}$, limited mainly by the dimensions of the magnet. The rate to be expected per energy band is

$$\frac{dN}{dt dE} = F\, G\,, \tag{16.3}$$

where F is the flux. Taking for example the flux in fig. 16.3 at 10 GeV, yields for ISOMAX a rate of about 3×10^5 per week in an energy bin of 1 GeV. In this region hydrogen dominates, followed by helium and, with far lower rates, by heavier elements. For example, the element group from lithium to boron (Z = 3–5) accounts for less than 0.1% of the flux [164].

Combining the measurements of the magnetic rigidity by the magnet spectrometer, of the velocity by the TOF and the Cherenkov counters and of the energy loss in the scintillation counters allows one to determine energy, mass and charge (see chapter 14). An example for the charge separation of light elements is shown in fig. 16.5(b). In this case the momentum measurement was not used. For each observed particle the charge is derived from β (by TOF) and dE/dx measurements using the scintillators for both. For a given β the average energy loss is proportional to the square of the particle's charge (see eq. (3.25) on page 30).

In order to be able to measure the very small fluxes up to particle energies close to 10^{15} eV the geometry factor has to be substantially increased. Since a magnet with a correspondingly large aperture is too heavy for a balloon experiment, the energy and mass determination by large-area detectors has to be achieved by dE/dx and β measurements. For $\beta \to 1$ both TOF and Cherenkov measurements become infeasible. Here transition radiation (TR) can be employed for the measurement of a particle's Lorentz factor γ (see chapter 12 and specifically for applications in cosmic ray experiments [964]). The principle is shown in fig. 16.6(a). The square root of the signal of a TR detector is plotted against the square root of the specific ionisation, as measured by another detector (the square root of the ionisation signal is proportional to the charge). In general a TR detector measures the sum of the ionisation and TR signals (see also fig. 12.12 on page 497). Points on the delineated diagonal correspond to energies below the TR threshold with $\sqrt{dE/dx}$ values which are proportional to the



(a) (b)

Fig. 16.6 (a) The square root of the (TR+dE/dx) signals plotted against the square root of the dE/dx signals for different elements (schematic, adopted from [539]). The TR signal cannot be separately measured but only summed with the dE/dx signal. Signals without TR contribution lie on the diagonal. The TR signal sets in at a certain element-dependent energy. The Z^2 dependence of the dE/dx signal is used to separate the different elements and the γ dependence of the TR signal determines the energy. (b) Measurement of TR signals in a test module of the CREAM TR detector as a function of the Lorentz factor γ (from [64], with kind permission of Elsevier). In test beams with different particles (protons, pions, muons, electrons) measurements have been performed with (triangles) and without (squares) radiator. The measurements are compared with corresponding simulations (curves).

charge of the respective nucleus:

$$Z \propto \left(\sqrt{dE/dx}\right)_{\text{TR threshold}} . \tag{16.4}$$

The signals above threshold allow for a determination of the Lorentz factor γ and thus also of the energy if the nucleus' mass for the measured Z is assumed, for example as the average mass of the isotopes with that Z. It is also possible to exploit the logarithmic increase of the dE/dx curve below the TR threshold for a rough γ measurement (see eq. (3.25) on page 30).

The method is shown by a realistic example in fig. 16.6(b). In this figure test measurements are plotted as function of the Lorentz factor γ and compared to simulations of the signals of the CREAM TR detectors (TRD) (fig. 16.7). Measurements and simulations are shown with and without TR radiator such that one can well distinguish the signal contributions from ionisation and from TR radiation. For the CREAM detector the threshold lies at $\gamma \approx 10^3$ (about 1 TeV for protons and 50 TeV for iron) and saturation is reached at $\gamma \approx 10^{4.5}$ (about 30 TeV for protons and 1.5 PeV for iron). For the heavier elements the measurements are not limited by saturation but rather by a too low flux.

As an example for a balloon experiment for the detection of particles up to the highest possible energies fig. 16.7 shows the CREAM detector [64]. Also referring to this experiment are fig. 16.6(b) with test measurements for CREAM and fig. 16.4 relating to CREAM flights. In order to obtain a geometry factor as large as possible, at a low total weight, a magnetic field is abandoned. The energy measurements are performed

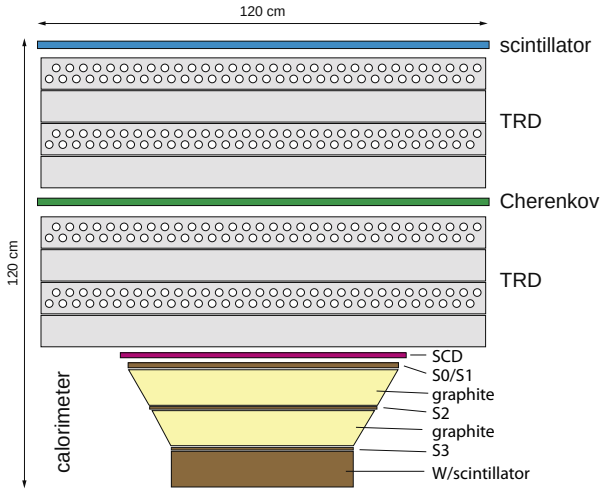


Fig. 16.7 Detector of the balloon experiment CREAM experiment (adopted from [65], with kind permission of Elsevier). The upper part with the larger acceptance consists of a scintillation counter, a Cherenkov detector and TRD modules. The lower part consists of a silicon detector (SCD), several planes of scintillating fibres (S0–S3), two graphite layers and a tungsten–scintillator calorimeter. See description in the text.

by a total of eight layers of TR detectors. A TR layer consists of a bloc of polyethylene foam as radiator and proportional counter tubes filled with a xenon–methane mixture (95:5) for the readout. The TRD acceptance ($G \approx 2.2\,\mathrm{m^2\,sr}$) is additionally covered by scintillation counters providing a precise time and pulse height measurement and a Cherenkov detector (acrylic glass, $n = 1.5$). Both detector components are important for the suppression of backscattering signals from the subjacent calorimeter. All detectors above the calorimeter have a total column density of only $7\,\mathrm{g/cm^2}$, still acceptable as compared to the low density of the overlying atmosphere. For the minimisation of weight and area density it is also important that the detectors are self-supporting and that a pressure vessel is not required.

The calorimeter (chapter 15) is optimised for a good energy resolution up to energies of some 100 TeV at a weight as low as possible. Because of the weight limitation its acceptance ($G \approx 0.3\,\mathrm{m^2\,sr}$) is smaller than for the TRD. The acceptance of the calorimeter is covered by a silicon detector with $2.12\,\mathrm{cm^2}$ sized readout pads. The calorimeter consists of two blocs of compacted graphite with a total thickness of 0.42 interaction lengths or about one radiation length. A portion of the incident nuclei initiates a shower in the graphite such that the shower can be measured by the calorimeter. In front and between the graphite modules scintillating fibre detectors (section 13.6) are inserted each providing two-dimensional position measurements which allow for direction reconstruction. The active layers of the calorimeter also consist of scintillating fibres. In this way the calorimeter reaches about 20 radiation lengths with a geometrical thickness of only 20 cm. The TRD–calorimeter combination can be used for a relative energy calibration.

The CREAM experiment has substantially contributed to filling the energy gap between the direct measurements of cosmic radiation and the indirect measurements by ground-based air shower detectors.

16.2.2 Satellite experiments

Experiments on satellites can be operated over many years, such that small fluxes, like those of antimatter in the cosmic radiation, become measurable. The first satellite experiments for measurements of galactic cosmic rays, hence not only particles coming

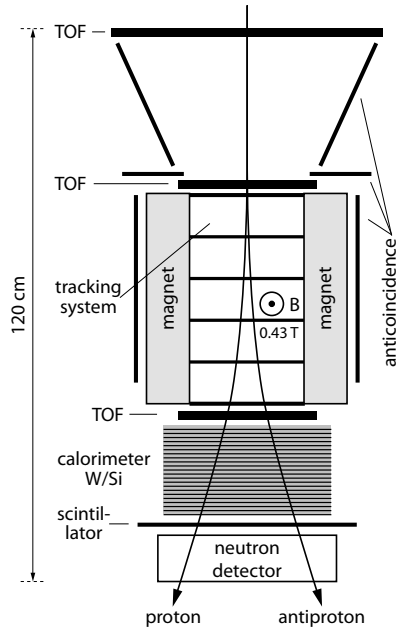


Fig. 16.8 Cross section of the PAMELA apparatus (from [55], with kind permission of Elsevier). The detector is composed of a magnetic spectrometer with a permanent magnet, a silicon-strip tracker, a silicon—tungsten calorimeter, time-of-flight (TOF) counters, anticoincidence counters and a neutron detector. An event is triggered by the TOF counters with the anticoincidence as veto. The detector has a geometrical acceptance (geometry factor) of $G = 20.5\,\mathrm{cm}^2\,\mathrm{sr}$ and weighs about 400 kg.

from the sun, are PAMELA [758] and AMS-02 [93], launched in 2006 and 2011, respectively (before, only particles from the sun had been measured). Both experiments are designed as magnetic spectrometers, which is an essential improvement compared to earlier experiments since a momentum determination permits a charge determination and thus the search for antimatter. In particular the measurements of antiprotons and positrons have previously suffered from small statistics and the influence of the Earth's atmosphere, since until then all such measurements had been accomplished with balloon experiments. For the measurement of cosmic gamma radiation up to hundreds of GeV the Fermi satellite, launched in 2008, is currently (2017) dominating the field. As a further example for satellite experiments we will introduce the gamma-ray telescope on the Fermi satellite in this section.

PAMELA. PAMELA is an acronym for 'Payload for AntiMatter Exploration and Light-nuclei Astrophysics' [55, 758]. The scientific goal of the PAMELA mission is above all the measurement of antiprotons and positrons and the search for antinuclei and dark matter.

PAMELA orbits since June 2006 on board the Russian satellite Resurs-DK1 in a nearly polar orbit with an inclination of 70.4° and an altitude between 350 km and 600 km. This orbit allows for measurement of the low-energy galactic cosmic particles near the poles where the influence of the Earth's magnetic field is small.

Figure 16.8 shows a schematic of the experimental components of PAMELA. The essential parts are a magnet spectrometer for the measurement of the rigidity of charged tracks, a silicon–tungsten calorimeter (16.3 radiation lengths, 0.6 hadronic absorption lengths), time-of-flight and anticoincidence counters. The calorimetric measurements are supported by a scintillation counter serving as a tail catcher (see section) and a neutron counter. The detection of the neutrons leaving the calorimeter improves the electron–hadron separation because hadrons produce about 10–20 times more neutrons than electrons [758].

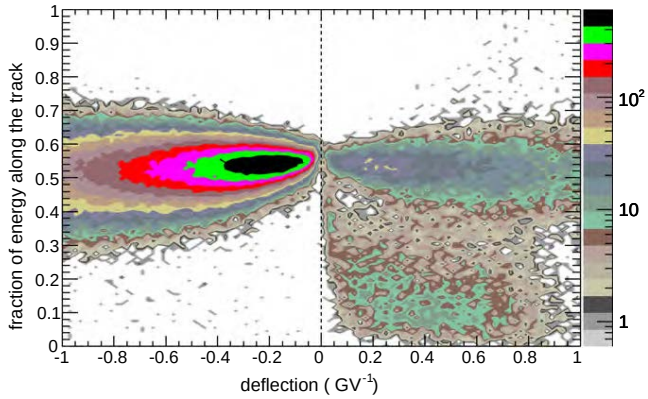


Fig. 16.9 Positron–proton separation with the PAMELA detector (from [57], with kind permission of Elsevier). Plotted is the ratio of the energy deposited close to the particle trajectory to the total calorimeter energy as a function of the magnetic track deflection (here expressed as inverse rigidity in units of inverse voltage). The colour code indicates the particle density. High-energy particles with little deflection in the magnetic field lie close to the zero-point which separates negative charges (left) and positive charges (right).

The PAMELA collaboration has observed an unexpectedly high positron flux relative to the electron flux [56]. The experimental method of positron identification, requiring in particular the separation from the strongly dominating protons, is demonstrated in fig. 16.9. The plot displays the fraction of the calorimeter energy that is deposited within 0.3 Molière radii around a particle trajectory as a function of the track deflection (here expressed as inverse rigidity, see eq. (16.1)). The tight region of 0.3 Molière radii takes into account that electromagnetic showers are in general narrower than hadronic showers. The energy ratio (ordinate) should be a constant for electrons and positrons, independent of the curvature. Most of the entrances on the left side of the plot are due to electrons, the horizontal continuation to positive charges correspond to positrons. Below the positrons the proton background is recognisable, while on the negative side there is little evidence for antiprotons.

AMS experiment. Since May 2011 the Alpha Magnetic Spectrometer (AMS) has recorded data on the International Space Station (ISS) [93]. This experiment pursues similar physics goals as the PAMELA experiment: determination of the spectrum and composition of the cosmic radiation up to energies of some 100 GeV, detection of antimatter and search for signs of dark matter.

However, the AMS detector (fig. 16.10) is much larger than PAMELA and more sensitive to small fluxes and to high particle energies. The particle separation, in particular the positron–proton separation, also employs Cherenkov and TR detectors in addition to the TOF counters and the calorimeter. The magnetic field of 0.125 T is provided by a permanent magnet (1.2 t $Nd_2 Fe_{14} B$), which replaced the originally foreseen superconducting magnet [673]. Since a permanent magnet does not need any utilities, like helium in the case of a superconducting magnet, it guaranties a longer lifetime of the experiment (up to 20 years). There are eight layers of silicon-strip detectors for tracking in the field volume. The 6.7 t detector has a geometrical acceptance of 0.45 m^2 sr. After on-board reduction of the internal data stream of 10 Gbit/s, the remaining stream of 2 Mbit/s is transferred to Earth.

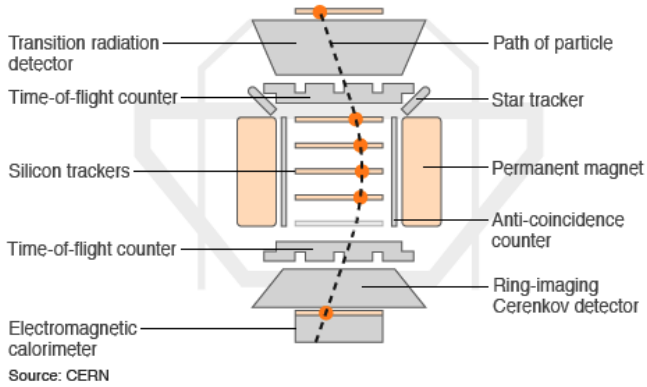


Fig. 16.10 Schematic drawing of the Alpha Magnetic Spectrometer (AMS-02) as installed on the ISS (Source: CERN). See description in the text.

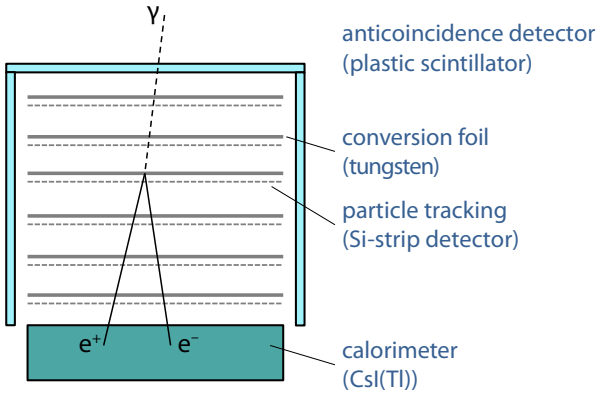


Fig. 16.11 Fermi LAT: simplified drawing of the LAT detector on the Fermi satellite (adopted from [399]). See explanation in the text.

Fermi satellite. The Large Area Telescope (LAT), the main instrument of the satellite mission 'Fermi Gamma-ray Space Telescope' (Fermi), is an imaging gamma radiation telescope covering the range from less than 20 MeV to more than 300 GeV [127]. With the telescope gamma radiation from discrete and diffuse sources is measured, transient events are followed and if necessary fast alarms are sent to other experiments.

The principle of the detector and the measurement of photons is depicted in fig. 16.11. A photon passes through a plastic scintillator which functions as anticoincidence detector in order to discriminate against charged particles, and converts in a foil which belongs to a stack of 16 conversion foils and 18 silicon-strip detectors. The energy of the created electron–positron pair is measured in the subsequent CsI(Tl) crystal calorimeter (8.6 radiation lengths). The silicon-strip detectors allow for a vertex and direction reconstruction of the pair (mostly without resolving the two tracks).

The detector has a modular structure with 16 'towers' oriented in the direction of incoming particles and 89 modules of the plastic scintillators. The modularity is essential for rejection of hadron showers, which are in general broader than electromagnetic showers. The detector has an area of 1.8 m × 1.8 m and a height of 0.72 m yielding a field of view (FOV) which corresponds to nearly 20% of the whole sky.

16.3 The atmosphere as a calorimeter

Due to a steeply falling energy spectrum the fluxes of cosmic rays above about 10^{15} eV become so small that they cannot be measured with detectors fitting onto satellites or balloons. The integrated particle flux (fig. 16.3) at energies above 10^{15} eV is about $1/(\mathrm{m}^2\,\mathrm{year})$ and above 10^{20} eV about $1/(\mathrm{km}^2\,100\,\mathrm{year})$. Nevertheless, cosmic ray detection at these high energies is possible with good acceptance because a primary cosmic nucleus generates an extensive air shower (EAS) which can reach the Earth's surface and can be measured with large area detectors.

In this and the following section we consider hadronic air showers which are initiated by primary nucleons and nuclei. Electromagnetic showers generated by high-energy photons in the atmosphere are treated in section 16.5.

16.3.1 Atmospheric depth

The interaction probability of cosmic rays in the atmosphere depends on the so-called *atmospheric depth* (also called *overburden*), that is, the integrated mass density ρ of the air from outside the atmosphere to an altitude h:

$$X(h) = \int_h^\infty \rho(z) dz\,, \tag{16.5}$$

where the altitude h is usually related to sea level. The atmospheric depth corresponds to the column density of the atmosphere at the altitude h and is given in units of mass per area. At sea level the whole atmosphere has the atmospheric depth

$$X_N = 1033\,\mathrm{g\,cm^{-2}}\,. \tag{16.6}$$

The numerical value is based on the so-called 'International Standard Atmosphere' (ISA)[3] with the parameters at sea level $p_0 = 1013\,\mathrm{hPa}$, $\rho_0 = 1.225\,\mathrm{kg/m^3}$, $T_0 = 15\,^\circ\mathrm{C}$ and an average temperature profile as described in [989].

An estimation for the density profile of the atmosphere can be obtained with the help of the barometric formula as follows. The weight increase of the atmosphere with elevation, $\rho\, g\, dh$, produces a pressure change of

$$\frac{dp(h)}{dh} = -\rho(h)\, g\,. \tag{16.7}$$

The ideal gas law connects pressure and density at a given temperature:

$$\rho(h) = \frac{p(h)\, M}{R\, T}\,, \tag{16.8}$$

where M is the molar mass of air ($M = 28.96\,\mathrm{g/mol}$ for dry air), $R = 8.31\,\mathrm{J\,mol^{-1}\,K^{-1}}$ the universal gas constant and T the absolute temperature. Assuming that temperature, air composition and gravitational acceleration are constant over the whole atmosphere, the insertion of (16.8) in (16.7) yields the barometric formula for the altitude dependence of the pressure:

[3] The ISA was defined by the International Civil Aviation Organization (ICAO) and is mainly used in aviation (see e.g. [989]). On the web interactive atmosphere calculators based on ISA can be found, for example [923].

$$p(h) = p_0\, \mathrm{e}^{-h/H}, \quad \text{with } H = \frac{R\,T}{M\,g} = \frac{p_0}{\rho_0\, g} = 8.4\,\mathrm{km}\,. \tag{16.9}$$

For g the *standard gravitational acceleration* $g_N = 9.80665\,\mathrm{m\,s^{-2}}$ [765] was used. Since the gravitational acceleration changes only by about 1% up to heights of 30 km, $g = const$ is a good approximation for calculating atmosphere properties. Because of the linear relation (with the assumptions above) between $p(h)$ and $\rho(h)$ the same exponential dependence on altitude also holds for the density:

$$\rho(h) = \rho_0\, \mathrm{e}^{-h/H}\,. \tag{16.10}$$

In the literature one finds different values for the *scale height* H, dependent on the assumptions for p_0, ρ_0 and the reference temperature. Since the temperature in the atmosphere varies with height one often approximates the atmosphere by a concatenation of exponential functions of the type (16.9) or (16.10). The scale height H then becomes a function of the temperature T, according to (16.10):

$$H(T) = \frac{R\,T}{M\,g} = 0.02928\,T\,\mathrm{km/K}\,. \tag{16.11}$$

For example, for a typical temperature in the lower stratosphere of 220 K, the scale height is about 6.5 km [438].

With (16.10) the integration of (16.5) yields for the atmospheric depth at the height h the estimate

$$X(h) = X_N\, \mathrm{e}^{-h/H}\,. \tag{16.12}$$

Because of (16.7) the atmospheric depth in (16.5) can also be expressed as an integral over the pressure profile:

$$X(h) = \int_0^{p(h)} \frac{dp}{g} \qquad \Longrightarrow \qquad X(h) = \frac{p(h)}{g} \quad \text{for } g = const\,. \tag{16.13}$$

For a realistic determination of the density distribution one uses measured pressure and temperature profiles $p(h)$, $T(h)$, for example from weather balloons.

A convenient parametrisation of the atmospheric depth for different height ranges is given in [438] (equation 5.59 of that reference). For example, in the range $25 < X(h) < 230\mathrm{g/cm^2}$, corresponding to $11 < h < 25\,\mathrm{km}$, which is a range most relevant for the first interactions of cosmic rays, the atmospheric depth is described by:

$$X(h) = X_0\, \mathrm{e}^{-h/H} = 1308.5\,\mathrm{g/cm^2} \times \exp\left(-\frac{h}{6.34\,\mathrm{km}}\right)\,. \tag{16.14}$$

The program CORSIKA [515], which is widely used for the simulation of air showers (see also section 16.4.1), provides atmosphere parametrisations for different locations and seasons [516]. The atmospheric depths are described by a concatenation of five functions of the type

$$\begin{aligned} X(h) &= a_i + b_i\, \mathrm{e}^{-h/c_i} \quad (i = 1, \ldots, 4)\,, \\ X(h) &= a_i - b_i\, h/c_i \quad (i = 5)\,. \end{aligned} \tag{16.15}$$

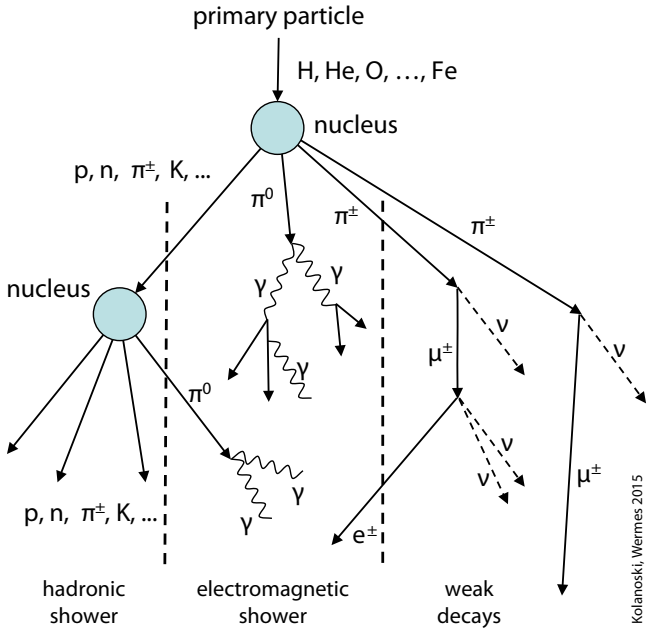


Fig. 16.12 Schematic representation of the different components of an air shower initiated by primary nuclei (in general fully ionised nuclei). The figure supplements the scheme in fig. 15.8 by weak decays which are relevant for thin media.

The first four functions describe the atmosphere up to 100 km, the fifth function above 100 km. The parameters a_i, b_i and c_i are tabulated for different geographic regions and different times of the year in the CORSIKA User's Guide [516].

16.3.2 Development of hadronic air showers

The protons and nuclei of the cosmic radiation interact with the atomic nuclei of the atmosphere through the strong interaction. The mean free path is obtained from the cross section σ per target particle and a density n of the target particles (see (3.7) in section 3.1) as

$$\lambda = \frac{1}{n\,\sigma} \quad \text{with } n = \frac{\rho\, N_A}{A}\,. \tag{16.16}$$

Then the mean free path can be given in the same units as the atmospheric depth in (16.5), namely as column density:

$$\lambda' = \lambda\,\rho = \frac{A}{N_A\,\sigma}\,. \tag{16.17}$$

As the characteristic length scale for the development of hadron showers we had introduced the *nuclear absorption length*, which is defined as the mean free path of hadrons with respect to inelastic scattering (see section 15.3.3). For protons the nuclear absorption length in air is $\lambda'_a \approx 90\,\mathrm{g\,cm^{-2}}$ [765,762] (see table 15.3 on page 599). Hence the atmosphere has a thickness of $1033/90 \approx 12$ absorption lengths which practically makes it impenetrable for cosmic hadrons. The absorption length λ'_a corresponds to a mean height h_1 for the first (inelastic) interaction:

$$\lambda'_a = X(h_1) = X_0\,\mathrm{e}^{-h_1/H} \quad \Longrightarrow \quad h_1 = H\,\ln\frac{X_0}{\lambda'_a} \approx 17\,\mathrm{km}\,. \tag{16.18}$$

Table 16.1 Properties of particles that are mainly produced in air showers (besides nuclear fragments).

Particle	Mass (MeV/c^2)	Dominant decays	Lifetime (s)	$c\tau$ (m)
p	938.27			
n	939.57	$pe\nu$	885.7	2.6×10^{11}
$\pi^{\pm}$	139.57	$\mu\nu$	2.6×10^{-8}	7.8
π^0	134.98	$\gamma\gamma$	8.4×10^{-17}	25.5×10^{-9}
$K^{\pm}$	493.68	$\mu\nu,\ \pi\pi$	1.24×10^{-8}	3.7
K^0_S	497.61	$\pi\pi$	8.95×10^{-11}	2.6×10^{-2}
K^0_L	497.61	$\pi e\nu,\ \pi\mu\nu,\ 3\pi$	5.12×10^{-8}	15.3
$e^{\pm}$	0.51			
$\mu^{\pm}$	105.66	$e\nu\nu$	2.2×10^{-6}	658.6
$\nu_e,\ \nu_\mu$	≈ 0			

For the numerical evaluation we have chosen the parameters $X_0 = 1308.5\,\mathrm{g\,cm^{-2}}$ and $H = 6.34\,\mathrm{km}$ according to the approximation (16.14) which covers the height range between about 11 and 25 km and thus the relevant heights of first interactions.

The absorption length λ'_a is a fixed parameter corresponding to the mean free path of a hadron at some energy which is relatively low in the context of cosmic ray physics (see footnote 2 on page 599). Because the cross sections of hadrons rise at high energies (see [765]) the actual mean free path is shorter and the height of the first interaction correspondingly increases with energy. For example, at $10^{16}\,\mathrm{eV}$ the mean free path of protons in air is about 60% of the fixed parameter λ'_a (see e. g. [17]), leading to $h_1 \approx 20\,\mathrm{km}$. For heavier nuclei the cross section increases with the mass number $\mathcal{A}_p$ of the projectile roughly as $\mathcal{A}_p^{2/3}$ at low energies, approaching a proportionality to $\mathcal{A}_p$ at high energies (as before we distinguish mass number $\mathcal{A}$ from atomic mass A, see footnote 1 on page 24). Since $\mathcal{A}_p$ corresponds to the number of nucleons the $\mathcal{A}_p$ proportionality corresponds to incoherent scattering off single nucleons.

The hadronic components of the primary cosmic radiation, mainly protons and nuclei, generate particle showers by inelastic reactions in the atmosphere (fig. 16.12). We distinguish between primary and secondary cosmic radiation, where primary radiation designates the particles hitting the atmosphere, while secondary radiation emerges from interactions of the primary radiation with the atmosphere.[4]

The development of hadronic showers is discussed in section 15.3. The atmosphere differs from a compact calorimeter by its low density (at least three orders of magnitude lower), its lower atomic number and its density gradient. In the following we emphasise the corresponding differences, which mainly result from the fact that weak decays can no longer be neglected in air showers.

At a certain stage of shower development part of the hadronic energy remains in the hadronic shower component, another part turns via decays into photons, electrons and positrons which initiate electromagnetic sub-cascades and yet another part reaches the detector on the surface as muons and neutrinos. One can easily describe the char-

[4] In another context one calls 'primary particles' those which are accelerated in cosmic sources and 'secondary particles' those which are produced in reactions of primaries on their way through the cosmos.

acteristic behaviours of these three components by assuming that mostly pions are produced. The neutral and charged pions (π^0, $\pi^\pm$, table 16.1) decay dominantly into the following channels:

$$\begin{aligned} \pi^0 &\to \gamma + \gamma \\ \pi^+ &\to \mu^+ + \nu_\mu \\ \pi^- &\to \mu^- + \bar{\nu}_\mu \,. \end{aligned} \tag{16.19}$$

The neutral pions, contributing a fraction of about one third, almost spontaneously decay into two photons (electromagnetic interaction). The photons then trigger an electromagnetic shower, that is, a cascade of photons, electrons and positrons (section 15.2). Since in such a cascade the energy is relatively quickly split up, the electromagnetic component is also called *soft component* of the air shower. The absorption length of an electromagnetic cascade is given by the radiation length (section 3.3.2) of air $X_0^{\text{air}} = 36.6\,\text{g}\,\text{cm}^{-2}$. The *hadronic component* of the air shower (fig. 16.12) continuously transfers energy into the electromagnetic component due to the production of neutral pions and other mesons decaying into photons (see the corresponding discussion in section 15.3).

In not too dense media, as in the atmosphere, the decay probabilities of weakly decaying hadrons, like pions and kaons (table 16.1), can compete with the reaction probability in the medium via strong interactions (fig. 16.12). The dominant decay modes are

$$\pi^\pm \to \mu^\pm \, \nu_\mu \, (\bar{\nu}_\mu) \,, \tag{16.20}$$

$$K^\pm \to \mu^\pm \, \nu_\mu \, (\bar{\nu}_\mu) \,. \tag{16.21}$$

The produced neutrinos carry energy away which virtually cannot be detected. The ratio of the probabilities for decay and interaction are strongly dependent on the energy of the particle and on the density of the medium because of the energy dependence of the mean decay length (with the Lorentz variables $\gamma = E/(mc^2)$, $\beta = v/c$ and the lifetime τ),

$$\lambda_\tau = \gamma \, \beta \, c \, \tau = \frac{|\vec{p}|}{mc} \, c \, \tau \,, \tag{16.22}$$

and because of the density dependence of the absorption length,

$$\lambda_a = \frac{A}{\rho N_A \sigma_{inel}} \,. \tag{16.23}$$

At low energies and low densities decays dominate, at high energies and high densities interactions dominate. For example, the mean decay length of a charged pion at $E = 1\,\text{GeV}$ is $\lambda_\tau = 55\,\text{m}$, corresponding to only 1% of an absorption length at the height of the first interaction. Hence the decay is about 100 times more frequent than an interaction. Since λ_τ is at not too low energies proportional to the energy (or γ in (16.22)), we conclude that at this height, decay and interaction probabilities become equal for pion energies around 100 GeV.

Muons and neutrinos, mainly originating from pion and kaon decays, constitute the *hard component* of the shower (fig. 16.12). Above an energy of some GeV muons have a very high chance to reach the Earth. This is due to the long lifetime and thus long decay length which is relativistically stretched according to (16.22) and the small

interaction probability of muons with matter, such that energy is essentially only lost by ionisation. For example, the mean decay length of muons with more than 3 GeV energy is longer than the mean height of the first interaction of about 20 km. Therefore most muons created somewhere in the atmosphere with such an energy or higher, reach ground level since a minimum-ionising particle (section 3.2.1) loses less than 2 GeV in the whole atmosphere through ionisation.

Muons decay into electrons and neutrinos:

$$\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu \,, \tag{16.24}$$

$$\mu^- \to e^- + \nu_\mu + \bar{\nu}_e \,. \tag{16.25}$$

Thus, if muons decay they contribute to both the soft and the hard component of a shower. In summary, a shower features

– a soft electromagnetic component (electrons, positrons, photons),
– a hard muonic component with associated neutrinos,
– a hadronic component.

Each component can be separately detected and can jointly be exploited for the measurement of cosmic ray particles.

Most of the energy goes into the electromagnetic component because after the initial development of a hadronic shower again and again neutral pions are produced in the cascade which decay spontaneously into photons and thus drop out of the hadronic cascade (see the respective eq. (15.26) for compact calorimeters for which, however, weak decays can mostly be neglected). While at high altitude the electromagnetic component is mainly built up by photons from π^0 decay, at the end of the shower development mainly electrons and positrons from muon decays contribute.

Because the spectrum of the cosmic radiation strongly falls off with energy (fig. 16.3), low-energy atmospheric showers dominate such that for most of the showers the hadronic and electromagnetic components are already absorbed when the shower reaches the ground. The secondary particles of the cosmic radiation arriving at the ground consist mainly of muons and electrons and positrons from the muon decays. At sea level the flux of vertical muons above 1 GeV/c is [164]

$$\frac{dN}{dA\,dt\,d\Omega} \approx \frac{70}{\mathrm{m^2\,s\,sr}} \,, \tag{16.26}$$

with a zenith angle distribution[5] proportional to $\cos^2\theta$. The freely available cosmic muons, usually referred to as 'cosmics', are often used for detector tests. The flux (16.26) integrated over a hemisphere leads to the well-known rule-of-thumb for the rate

$$\frac{dN}{dA\,dt} \approx \frac{1}{\mathrm{cm^2\,min}} \,. \tag{16.27}$$

16.3.3 Properties of hadronic air showers

In the following we discuss some characteristic properties of air showers, which we back up by formulae whenever possible. However, it has to be understood that most formulae are just relatively coarse approximations which are usually not suited for quantitative data analyses. For data analyses one usually refers to simulations.

[5] The flux at ground level is dominated by GeV muons; at very high energies the angular distribution becomes flatter and eventually decreases towards small zenith angles [164].

Longitudinal shower development. The electromagnetic component dominates the number of shower particles and the deposited energy over most of the development of a shower. The longitudinal distribution of the number of electrons N_e as a function of the atmospheric depth (eq. (16.5)) is given by the Gaisser–Hillas formula [439]:

$$N_e(X) = N_{e,max} \left(\frac{X - X_1}{X_{max} - X_1} \right)^{\frac{X_{max} - X_1}{\Lambda}} \exp \frac{X_{max} - X}{\Lambda}, \tag{16.28}$$

where $N_{e,max}$ is the number of electrons in the shower maximum, X_{max} the atmospheric depth at the shower maximum and X_1 the depth of the first interaction. The parameter Λ is an effective radiation length, amounting to about $70\,\mathrm{g/cm^2}$. For the description of individual showers $N_{e,max}$, X_{max}, X_1 and Λ can be fitted as free parameters. The fluctuations of these quantities from shower to shower are relatively large, as typical for hadronic showers.

The Gaisser–Hillas formula resembles the Longo formula (15.17) in section 15.2.2.1 describing the development of electromagnetic showers, while the Gaisser–Hillas formula describes the electromagnetic component of a hadronic shower in the atmosphere. Here as well one finds that the average number of electrons in the shower maximum is proportional to the primary energy. For showers initiated by protons we use a parametrisation according to [439]:

$$\langle N^p_{e,max} \rangle = S_0 \, \frac{E}{E_c} \,, \tag{16.29}$$

where $E_c = 87.9\,\mathrm{MeV}$ is the critical energy in air (section 3.3.3, table 3.4) and $S_0 \approx 0.045$ an empirical parameter[6] [439]. The position of the shower maximum is linearly dependent on $\ln E$ as known for purely electromagnetic showers (see eq. (15.19)) [585]:

$$\langle X^p_{max} \rangle = c + d^p \, \ln E \,. \tag{16.30}$$

The parameters c and d^p, the so-called elongation rate, are determined through simulations as $c \approx 200\,\mathrm{g/cm^2}$ (for E in GeV) and $d^p \approx 25\,\mathrm{g/cm^2}$, corresponding to an increase per decade of $d^p_{10} = \ln 10\, d^p \approx 60\,\mathrm{g/cm^2}$.

When heavier nuclei initiate a shower one can as an approximation assume that a nucleus with mass number $\mathcal{A}$ and energy E interacts with the medium as $\mathcal{A}$ independent nucleons each with energy $E/\mathcal{A}$. Then one gets for the average shower parameters [585]:

$$\langle N^{\mathcal{A}}_{e,max} \rangle \approx \mathcal{A} \left(S_0 \, \frac{E/\mathcal{A}}{E_c} \right) = \langle N^p_{e,max} \rangle \tag{16.31}$$

and

$$\langle X^{\mathcal{A}}_{max} \rangle = c + d^p \, \ln(E/\mathcal{A}) = \langle X^p_{max} \rangle - d^p \, \ln \mathcal{A} \,. \tag{16.32}$$

For protons and iron nuclei the difference is $d^p \, \ln 56 \approx 100\,\mathrm{g/cm^2}$. In the general case the cosmic radiation is composed of different nuclei and $\ln \mathcal{A}$ has to be replaced by $\langle \ln \mathcal{A} \rangle$, an average over the abundance distribution of primary nuclei. The scatter of

[6] In electromagnetic showers the total number of produced electrons can be estimated by $N_{tot} = E/E_c$ (see eq. (15.8) in section 15.2.1). The parameter S_0 takes into account that in hadronic showers only a part of the energy goes into the electromagnetic component and that electrons and positrons are produced and destroyed before and after the shower maximum.

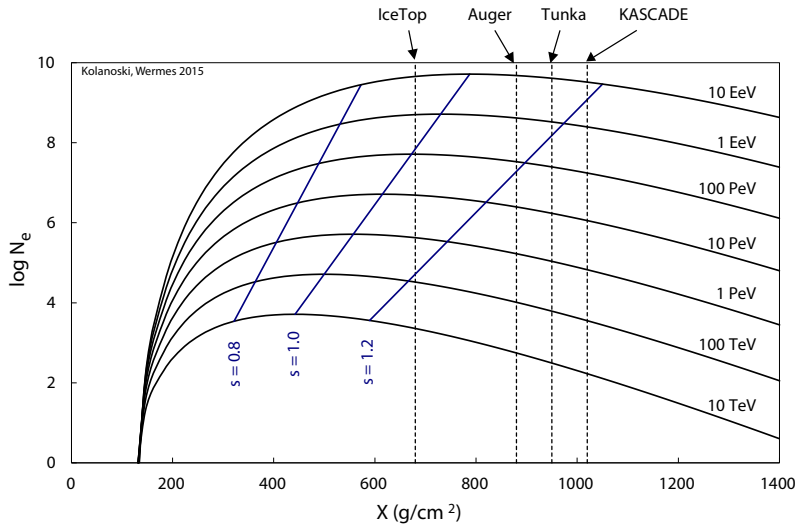


Fig. 16.13 The average longitudinal shower profile for primary protons as represented by the electrons generated in a shower. The number of electrons is plotted versus the atmospheric depth for different primary energies. The straight solid lines connect the points of equal shower age s (eq. (16.33)), for example, $s = 1$ connects the shower maxima. As examples, the atmospheric depths of four experiments are specified: IceTop, the surface component of the Neutrino Observatory IceCube at the South Pole, at an altitude of nearly 3000 m above sea level ($X \approx 680\,\mathrm{g\,cm^{-2}}$), the Pierre Auger Observatory in Argentina at an altitude of about 1350 m ($X \approx 880\,\mathrm{g\,cm^{-2}}$), Tunka near Lake Baikal at an altitude of about 675 m ($X \approx 950\,\mathrm{g\,cm^{-2}}$) and KASCADE in Karlsruhe at an altitude of 110 m, which is nearly at sea level ($X \approx 1020\,\mathrm{g\,cm^{-2}}$).

X_{max} becomes smaller with increasing $\mathcal{A}$ since approximately $\mathcal{A}$ independent subshowers, which can be averaged, contribute to the full shower. For the measurement of the shower maximum X_{max}, see for example fig. 16.21 and the corresponding text. The mean and the standard deviation of the measured distributions of X_{max} values in an energy bin are employed for the determination of the cosmic ray mass composition at that energy (see e.g. [1]).

In fig. 16.13 longitudinal profiles of the number of electrons are shown for protons with energies from 10 TeV to 10 EeV. Also plotted are the lines of constant *shower age* s which specifies the depth relative to the shower maximum:

$$s = 3\left(1 + \frac{2X_{max}}{X}\right)^{-1}. \tag{16.33}$$

The longitudinal shower development is described by a shower age between 0 and 3, where the value $s = 1$ is the value at the shower maximum.

The altitude of the detector location influences what part of the shower can be seen. Different ranges of the shower development can be investigated by detectors at different altitudes. The shower profile can also be scanned by measuring showers of a given energy at different zenith angles because at slant incidence the effective atmospheric depth (*slant depth*) increases with increasing angle.

How the different measurement quantities of an air shower are influenced by the

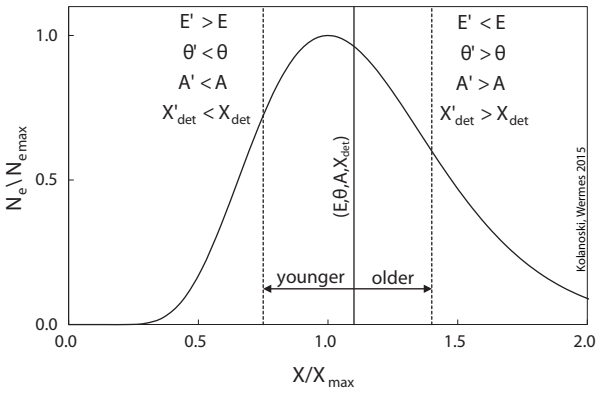


Fig. 16.14 Schematic representation of the detector position relative to the shower maximum under variation of the shower parameters energy E, zenith angle θ and mass A of the primary particle, measured at an atmospheric depth X_{det}. The full line indicates the position of a reference detector relative to the shower profile. The dashed lines show the changes of the relative position for the specified parameter variations. Adapted from [615].

detector position relative to the maximum is depicted schematically in fig. 16.14. We take as a reference quantity the detector signal of a shower with energy E, zenith angle θ and mass A of the primary particle measured at an atmospheric depth X_{det}. Any change of one of the parameters changes the shower age at which the shower is observed. For example, at a larger zenith angle a shower is observed at a later stage of its development than at a smaller zenith angle.

Lateral shower development. For the lateral extension of electromagnetic and hadronic showers in compact calorimeters we have only given coarse parametrisations for percentage containment (see sections 15.2.2 and 15.3.3, respectively). In compact calorimeters showers can mostly be reconstructed from the full energy deposition and the extension parameters are only used for estimations of losses. In contrast, air showers are usually measured in one detector plane only and the measurement of the lateral shower distribution becomes an essential ingredient for the reconstruction of cosmic rays from air showers.

The electron distributions are particularly important at energies for which the shower maximum is close to the detector, that is, for shower ages around $s = 1$. For 'old' showers the electromagnetic component dies out and the muonic component becomes more important. The lateral electron density of an air shower at the atmospheric depth X as a function of the distance r from the shower axis is parametrised (based on analytical calculations) by the so-called NKG function (NKG = Nishimura–Kamata–Greisen) [584, 475]:

$$\rho(r, X) = C\,\frac{N_e(X)}{R_M^2}\left(\frac{r}{R_M}\right)^{s-2}\left(1+\frac{r}{R_M}\right)^{s-4.5}. \tag{16.34}$$

Here $N_e(X)$ is the number of electrons according to (16.28), R_M the Molière radius (see eq. (15.21) on page 590) and s is the shower age at atmospheric depth X. Normalising the density to the number $N_e(X)$ at fixed X fixes the constant C:

$$N_e(X) = 2\pi\int_0^\infty r\,\rho(r, X)\,dr = 2\pi\,C\,N_e(X)\,B(s, 4.5-2s)$$

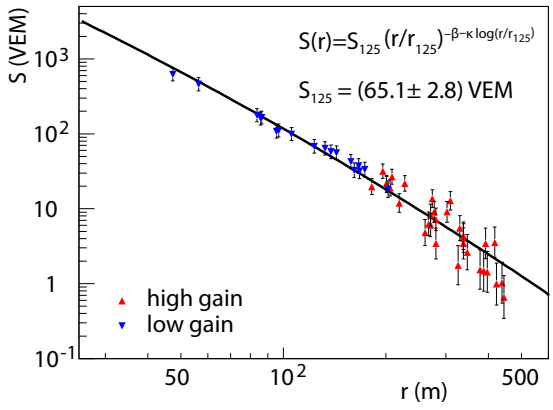


Fig. 16.15 Lateral signal distribution in the IceTop array [16] (section 16.4.3) of an event with a primary energy of about 65 PeV. The signals, given in units VEM (see section 16.4.2), are fitted by a lateral distribution function as displayed in the legend. The parameter S_{125} is the signal expectation value at a reference radius $r = 125\,\mathrm{m}$, which serves as an energy proxy for the reconstructed event. Source: IceCube Collaboration.

$$\Rightarrow \quad C = \frac{1}{2\pi\, B(s, 4.5-2s)} = \frac{\Gamma(4.5-s)}{2\pi\,\Gamma(s)\,\Gamma(4.5-2s)}\,. \tag{16.35}$$

Here B and Γ are the beta and gamma functions [976]. The Molière radius in units of a column density is for air $\rho\, R_M = 8.83\,\mathrm{g/cm^2}$. Since the shower development is determined by the atmosphere densities above the detector it is not quite obvious what radius should be used at detector level. In [475] it is proposed to refer to the density at about two radiation lengths above the detector.

The NKG function (16.34) is used to describe the lateral distribution of electrons in hadronic showers, with the Molière radius R_M usually adjusted as free parameter. If the exponents in (16.34) are kept free as well the hadronic and muonic components can be described by the same functional form [219].

The measured signals of an air shower are, for detectors which count particle passages (thin scintillation detectors, wire chambers and similar), roughly proportional to the number of charged particles per area, independent of whether they are electrons, muons or other particles. If instead the deposited energy is calorimetrically measured (e.g. through Cherenkov radiation in water or ice tanks), the electromagnetically interacting particles are weighted with their energy while muons deliver approximately energy-independent signals. Furthermore, since the particle composition in a shower depends on the altitude of the detector, the optimal description of the signals by a lateral distribution function (LDF) is different for each experiment. An example of a fit to the lateral distribution of signals in the IceTop array (see section 16.4.3) is shown in fig. 16.15. The different LDFs have in common that for not too small distances from the shower core the signals approximately fall off with a power of r, the distance from the shower axis.

Dependences on the mass of the primary particle. The determination of the mass composition of the primary cosmic radiation is an important goal of air shower experiments. Besides the astrophysical significance the mass composition also plays a role for shower reconstruction because in general the determination of the energy from the detector signals requires the knowledge of the primary mass. Mass dependences of shower properties can be traced back to the already mentioned fact that a nucleus with mass number $\mathcal{A}$ and the energy E interacts like $\mathcal{A}$ independent nucleons, each with energy $E/\mathcal{A}$. In order to extract the mass of a primary cosmic ray from the initiated air shower, which is in general only possible on a statistical average, one can exploit

the following mass-dependent observables:

– The position of X_{max}: According to (16.32) for primaries with a higher number of nucleons $\mathcal{A}$ the shower maximum is reached earlier and the spread is smaller because $\mathcal{A}$ primary interactions occur.
– Muon abundance: With a larger number of nucleons more pions are generated in the first interactions which have a high probability to decay before they interact because in the high, thin atmosphere decay as compared to interaction is more frequent than in a denser atmosphere. In particular, this applies to high-energy muons from the first interactions which appear in a narrow 'muon bundle' collimated around the shower axis (see section 16.4).

Methods for mass determination are exemplary described in section 16.4 for the air shower detectors KASCADE and IceTop, which exploit the muon abundance, while Tunka and Auger mainly use the position of X_{max}.

16.3.4 Gamma showers

High-energy cosmic gamma radiation with energies above about 100 GeV can also be detected through their shower development in the atmosphere. The shower develops as an electromagnetic cascade of photons, electrons and positrons, as described in section 15.2. Principally also the photoproduction of hadrons with weak decays into muons is possible if there are shower particles with sufficient energy. However, the respective cross sections are so small that gamma showers are referred to as *muon-poor*, a characteristic feature used for the recognition of gamma showers. Another feature is the smoother development of gamma showers as compared to hadron showers (see section 16.5.2).

Because of the variable density of the atmosphere we will here also employ the atmospheric depth in units of a column depth (mass per area) as length scale according to (16.5). Correspondingly, the longitudinal development of a gamma shower is characterised by the radiation length in air $X_0 = 36.66\,\mathrm{g/cm^2}$ (table 3.4). The first interaction of the photons with the air takes place on average at about $9/7\,X_0$ (see eq. (3.142) on page 84) which is about $47\,\mathrm{g/cm^2}$, corresponding to a height of about 26 km. The shower maximum can be estimated using the approximation (15.19) with the critical energy $E_c = 87.9\,\mathrm{MeV}$ in table 3.4:

$$t_{max} = \ln\frac{E_0}{E_c} + 0.5 \qquad \Longrightarrow \qquad X_{max} = X_0\left(\ln\frac{E_0}{E_c} + 0.5\right) . \tag{16.36}$$

In table 16.2 we list for some energies numerical values for t_{max}, X_{max} and $h_{max} = h(X_{max})$. Here the height of the shower maximum, h_{max}, was calculated from X_{max} using (16.15) together with the parameters of the 'US standard atmosphere' in table 23 of [516].

16.4 Air shower detectors

Extensive air showers (EAS) initiated by high-energy cosmic radiation can be rather cost-efficiently measured by distributing detectors over large surface areas.

16.4.1 Detection principles

Figure 16.16 shows the principle of the development of an EAS and its detection. The shower particles form a widespread shower front which moves nearly with the

Table 16.2 Parameter of the shower maximum of an air shower initiated by a photon with energy E: t_{max} = number of radiation lengths up to the shower maximum, X_{max} atmospheric depth of the shower maximum, h_{max} height of the shower maximum.

E (GeV)	10^2	10^3	10^4	10^5	10^6
t_{max}	7.5	9.8	12.1	14.4	16.7
X_{max} (g/cm^2)	276	361	445	530	614
h_{max} (km)	9.9	8.1	6.6	5.3	4.2

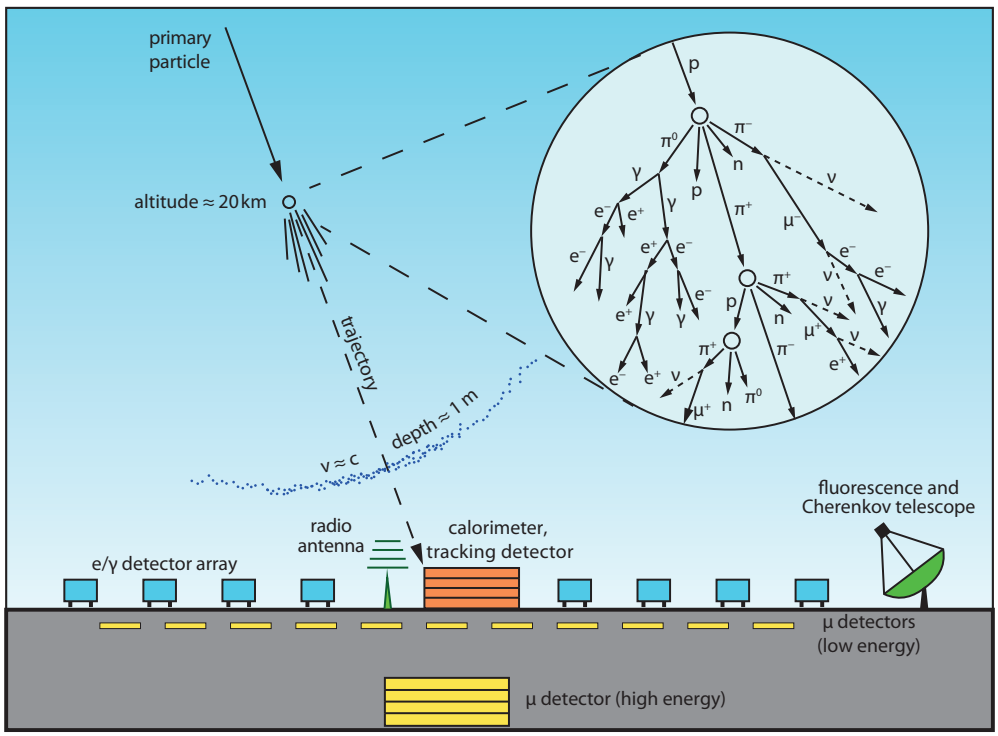


Fig. 16.16 The principle of development and detection of an extensive air shower (adopted from a drawing by K.-H. Kampert in [507]). The distribution of detector modules over large areas is characteristic for air shower experiments. The distances between the detectors and the size of the detector array determine the measurable range of primary energies. With areas of 1 km^2 an energy range up to some EeV can be covered; for the highest energies up to about 10^{21} eV detector areas of some 1000 km^2 are required. Besides the direct detection of the shower particles on the surface one can also record Cherenkov, fluorescence and radio radiation, which the particles emit along the shower propagation in the atmosphere. Some experiments have special detectors for the measurement of muons which are shielded against other, less penetrating particles. For the most abundant low-energy muons (∼GeV) few metres of an Earth layer is sufficient, for the selection of muons at higher energies thicker absorbers are necessary, for example a more than 1 km w.e. (w.e. = water equivalent) thick absorber for the TeV region (see fig. 16.32 on page 700). Sometimes the large area detectors are supplemented by detectors permitting more detailed studies of shower particles within a limited acceptance. Such supplements can be, for example, tracking detectors and calorimeters.

velocity of light. The particle avalanches develop into the direction of incidence of the primary particle. The lateral distribution of the shower particles has a maximum on the shower axis and falls off approximately like a power of the distance from the axis corresponding (see the discussion of lateral distribution functions on page 676f). Thus the highest particle density is found around the impact point of the shower axis on ground. Because of the high particle density in a shower it is sufficient to sample the shower with distributed detectors with sensitive areas that constitute only a small fraction of the total array area. For example, a primary proton with an energy of 10^{15} eV (1 PeV) produces near the surface on average 10^6 secondary particles (80% photons, 18% electrons and positrons, 1.7% muons and 0.3% hadrons).

The interesting quantities, namely direction, energy and mass of the primary particle, can be derived from the properties of the air shower, mainly from the lateral particle densities. The challenges of the measurement increase in the order of the above-named quantities. While the direction can be directly read off the measured data, the determination of the mass requires more or less involved air shower simulations, with the help of which one can infer the original particle by comparison of measured and simulated data. A standard in the field of air shower experiments is the simulation package CORSIKA (COsmic Ray SImulations for KAscade) [515]. The program allows one to choose between different models for hadronic interactions, a feature which is very important for systematic studies (see also the discission of atmosphere models the in CORSIKA in section 16.3.1). The development of interaction models faces the same difficulties as those mentioned in section 15.3 for showers in compact calorimeters related to the complexity of hadron showers and the lack of sufficient input data. At the highest cosmic ray energies the problem is particularly obvious since there are no data to compare with.

16.4.2 Detector concepts

For the measurement of air showers different types of detectors are deployed:

- planar scintillation counters or wire chambers (some square metres per module);
- water or ice tanks registering Cherenkov radiation with a volume of a few cubic metres;
- dedicated muon detectors with absorber and counter layers;
- detectors for Cherenkov radiation in air;
- fluorescence telescopes;
- radio antennas.

While planar counters measure the number of particles, the tanks deliver energy measurements for electromagnetically interacting particles. In counter detectors muons are not distinguishable from other particles. In tanks high-energy muons produce a signal which is approximately energy-independent. It is common practice to use the signal of a muon that passes vertically through the detector for calibration and to define a unit VEM = *vertical equivalent muon* for signal sizes (instead of e.g. charge or voltage).

Since muon production is highly sensitive to the mass of the primary particle one often installs special muon detectors which are able to measure muon rates, even if only in a part of the detector. Typically, energy thresholds lie between somewhat below a GeV and some GeV and the rates are dominated by GeV muons numerously generated in the shower process. Few experiments can also measure the TeV muons in the core

of a shower separately from the other shower particles. For example, for the IceCube Neutrino Observatory [12, 16] the threshold for muons that are measured in the ice is about 500 GeV (see sections 16.4.3 and 16.6.5).

Cherenkov, fluorescence and radio detectors are sensitive to the different radiation mechanisms of the electron/positron component in air. It is thus possible to trace a large part of the shower development in the air such that the shower profile and in particular the shower maximum can be determined, as we will discuss for the different emission types in the following.

While for the directionally concentrated TeV-gamma showers (see section 16.5) Cherenkov detectors are mostly deployed as imaging Cherenkov telescopes, for the broader hadron showers (see fig. 16.22) one usually measures only the lateral distribution of the Cherenkov light in the detector plane without a directional information. The shower parameters are then reconstructed from the lateral distribution, as in the case of the above discussed particle detectors. Additionally the height of the shower maximum can be determined from the shape of the light distribution near the shower axis (see the description of the experiment Tunka in section 16.4.3).

At energies above 10^{17} eV fluorescence light in the wavelength range 300–400 nm can be efficiently detected. It is produced by interactions of charged particles with the nitrogen molecules of the atmosphere and at such energies it can be observed in clear nights at distances up to 30 km. For experiments at the currently highest measured energies fluorescence measurements are of high importance for the absolute energy calibration and for the determination of the mass composition (see the description of the Pierre Auger Observatory in section 16.4.3).

The exploitation of radio emission by high-energy showers is intensively investigated by different experiments. It is expected that with this method very large detection areas can be more cost-efficiently covered than with other methods. In order that charged shower particles emit radio waves the charges have to be separated at least by distances of the order of the wavelength because otherwise the emission of the charges of both signs superpose destructively (note that a positron–electron pair looks neutral from the outside if the distance cannot be resolved). Two mechanisms of charge separation leading to radio emission have been predicted and have been experimentally confirmed. In the atmosphere the prevailing effect is charge separation by the geomagnetic field ('geo-synchrotron effect') [546] and in dense media (ice, salt, rock) it is the build-up of local charge asymmetries through positron annihilation and the production of free electrons by photoeffect and Compton scattering ('Askariyan effect' [121] or 'radio-Cherenkov effect').

Air shower detectors are in general arrangements of many relatively small modules which are distributed at fixed distances on a regular grid. The optimal ratio of sensitive to total area of the detector array and thus the necessary number and size of the modules results from the requirement that the statistical signal fluctuations should not be larger than the natural shower fluctuations. Typically, the minimal size of recorded signals corresponds for counter experiments to one or a few particles and for calorimetric detectors to a fraction of the signal a muon would generate. The distances between the modules are optimised according to the energy range to be covered: about 100 m for detectors which operate in the regime of the 'knee' (about 3×10^{15} eV) and up to more than 1 km for the highest energies around 10^{20} eV. The size of the full array ranges from fractions of a square kilometre to more than 3000 km^2, adjusted to the cosmic ray fluxes to be expected (see fig. 16.3).

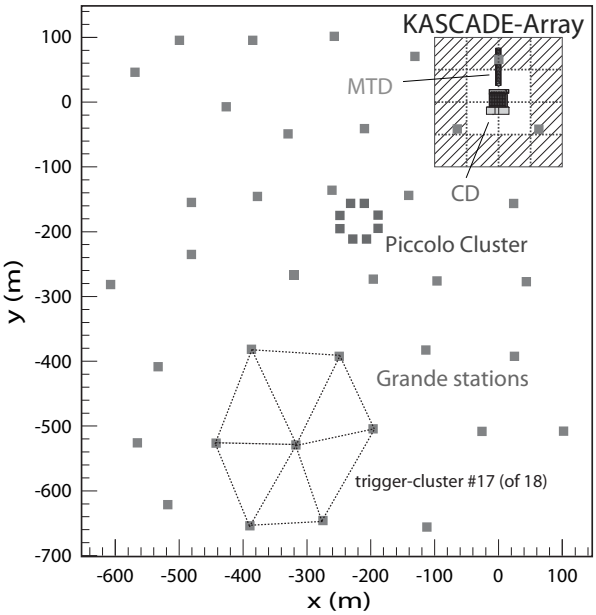


Fig. 16.17 Detector configuration of the experiment KASCADE-Grande (from [110]). KASCADE-Grande consists of the original KASCADE array (top right), the Grande stations, distributed over the whole area, and the trigger array Piccolo. KASCADE is divided into 16 clusters of 15 or 16 detectors each. Additionally it includes a central detector (CD) and a muon tunnel (MTD). Drawn is also one of several possible trigger clusters of seven Grande stations which trigger the data readout if they are hit in coincidence. See the description in the text.

16.4.3 Examples for air shower experiments

In the following we want to discuss three experiments which measure the energy spectrum and the composition of the cosmic radiation in the range of medium energies from about 10^{15} eV (near the 'knee', fig. 16.3 on page 659) to 10^{18} eV employing different detector technologies. As a further example the Auger experiment will be introduced which measures cosmic radiation at the highest energies observed to date of more than 10^{20} eV.

KASCADE and KASCADE-Grande. The air shower detector KASKADE in Karlsruhe [107] has 252 detector stations distributed in a chequered array with relative distances of 13 m on an area of 200 m × 200 m (top right in fig. 16.17). The stations consist of scintillation detectors, mainly for registering the electromagnetic component, and a scintillation detector below a 20 cm thick lead–iron shielding for the detection of muons. The whole array is divided in 16 sub-clusters, from which the outer 12 (hatched squares in the figure) consist of μ and electron/γ detectors and the inner four of electron/γ detectors only. The centre of the facility consists of a 4000 t calorimeter, two layers of multiwire proportional chambers (section 7.8), a further layer of streamer tubes (section 7.6.3) and a trigger plane made of scintillation counters. This system serves for the measurement of hadrons, muons and electrons in the core region of the air shower. A 50 m long tunnel (MTD) in which muons are detected stretches out from this central detector.

The original KASCADE detectors have been augmented by KASCADE-Grande, an array of 37 stations in relative distances of 130 m on an area of 0.5 km^2 (fig. 16.17). Each KASCADE-Grande station is equipped with 10 m^2 scintillation detectors which essentially measure the electromagnetic component of a shower. A small array of scintillation detectors, called Piccolo, provides a fast trigger for all components.

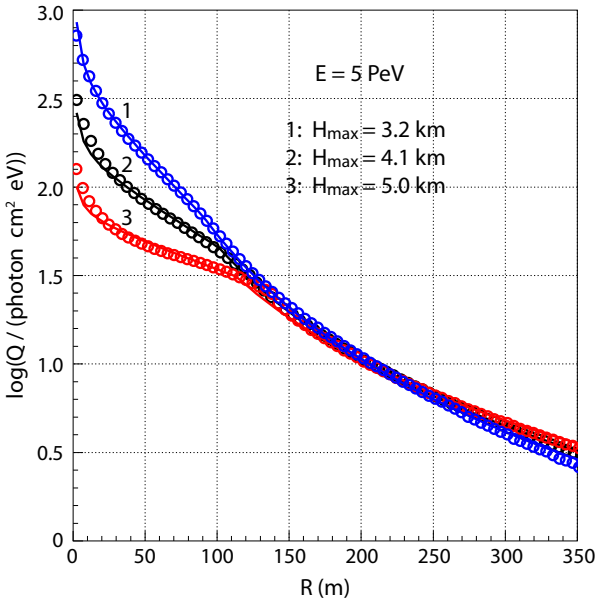


Fig. 16.18 Simulated lateral distribution of Cherenkov light from a 5-PeV shower registered by the TUNKA detector (adapted from [259], with kind permission of Elsevier). The curves for the different heights H_{max} of the shower maximum essentially differ only in the slope at small distances from the shower centre. This is used to determine the shower maximum which is sensitive to the mass of the primary particle. The energy is determined from the light distribution at larger distances R.

TUNKA experiment. The experiment TUNKA-133 [106], situated in the Tunka valley near Lake Baikal, measures air showers which are initiated by charged cosmic rays or high-energy cosmic gamma radiation. The main component of the TUNKA experiment is a 1 km^2 array consisting of 133 photomultiplier detectors which measure in dark, clear nights the Cherenkov light produced by air showers. The system is not imaging as for gamma telescopes (section 16.5), which makes it simple and cost-effective. A disadvantage is that the requirements on darkness and clear sky permit measuring only for about 4% of the time.

From the lateral distribution of the light the direction of incidence, the energy and the mass of the primary particle can be reconstructed, as explained in fig. 16.18. Since the shower particles are predominantly produced at small angles, the Cherenkov light is emitted in a limited region around the shower axis (see the discussion for gamma telescopes in section 16.5.2). The distribution around the shower axis is the broader the higher the shower maximum is located. From that the parameter X_{max} (in terms of atmospheric depths, note that H_{max} in the figure is the absolute height above ground) can be determined for each shower. The X_{max} distribution of all showers measured in an energy interval provides an estimate of the mass composition at this energy.

IceTop. IceTop [16] is a 1 km^2 air shower detector at the South Pole above the IceCube detector, which is situated at a depth around 2000 m in the ice (see section 16.6.5.3 and fig. 16.34). IceTop has 81 stations separated by 125 m from each other. A station has two tanks separated by 10 m which are filled with ice in a volume of 2.7 m^2 × 0.9 m. At the surface of the ice two optical modules with photomultiplier tubes register the Cherenkov light which is generated by air shower particles in the tank (fig. 16.19).

Compared to KASCADE-Grande and TUNKA-133, which have a similar size and cover the same energy range, IceTop has the advantage that the high-energy muon bundle in IceCube can be measured in coincidence with showers in IceTop. Because

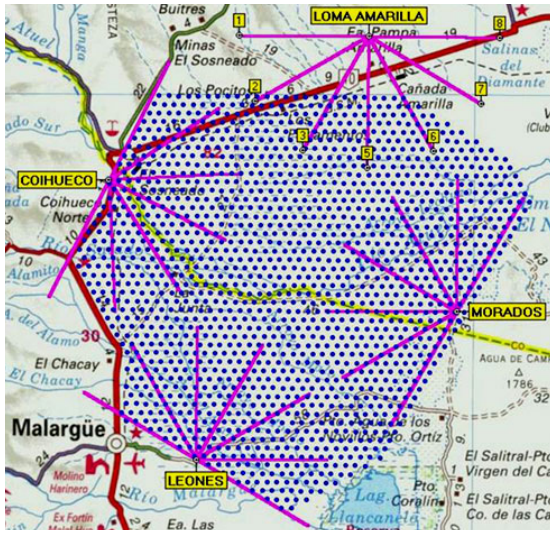
Fig. 16.19 IceTop air shower detector. Left: Two tanks of a station in a distance of 10 m from each other. After the installation work the trench is filled again with snow to the top edge of the tank. The South Pole station is visible in the background. Right: View into the inside of a tank filled with water in which two optical modules wait to be frozen in. Source: IceCube Collaboration/NSF.

of the large depth of IceCube only the high-energy muons above about 500 GeV reach the IceCube detector. Such muons, which come from the first interactions in the high atmosphere, appear in the deep ice in narrow bundles with a muon multiplicity that strongly depends on the mass composition.

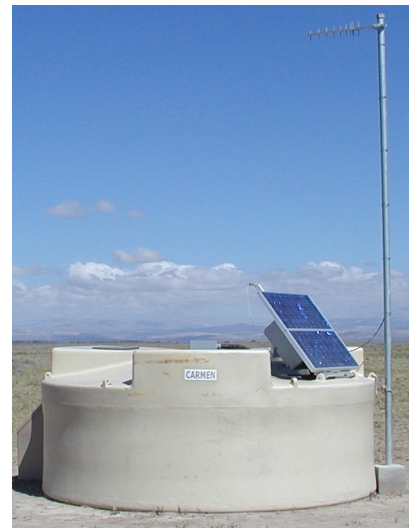
Auger experiment. The Pierre Auger Observatory (PAO) [29] in the province of Mendoza in Argentina covers the so far highest measured energies of cosmic radiation above about 10^{18} eV. The experiment particularly investigates whether the spectrum is cut off at about 5×10^{19} eV, corresponding to the Greisen–Zatsepin–Kuzmin limit (*GZK Cut-off*) [476, 1018]. Such a cut-off was predicted because at these energies the scattering of protons (with some modification also nuclei) off the photons of the cosmic microwave background passes the threshold of pion production. The protons (or nuclei) lose energy in this reaction and are shifted towards lower energies in the spectrum.

The experiment covers an area of about 3000 km^2 (about the size of Rhode Island or Luxembourg) and is thus to date (2017) the largest experiment for the observation of cosmic radiation (fig. 16.20). The 1600 detectors are tanks filled with pure water, each with an area of 11.3 m^2 at distances of 1.5 km from each other. In the water tanks the Cherenkov radiation generated by shower particles is registered (same principle as for IceTop). This detection concept has been chosen because it is cost-efficient for such a large detector and compared to scintillation counters offers some calorimetric information about the shower particles.

Supplementing the water detectors, 30 fluorescence telescopes at four stations observe the atmosphere above. Each telescope has a field of view of $30° \times 30°$. The mirror with an area of about 12 m^2 focuses the fluorescence light of the shower particles onto a camera consisting of an array of photomultipliers (indicated as circles in fig. 16.21(a)). The four stations are distributed along the border of the surface detector array as shown in fig. 16.20(a). In this way they can observe the shower profiles from the side such that the full trace of a shower can be imaged (in fig. 16.21(a) by two adjacent telescopes).

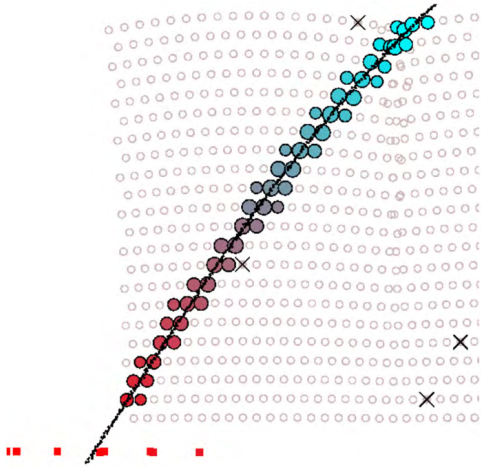


(a) Detector configuration of the Auger experiment with 1600 water tanks and fluorescence telescopes at four stations (yellow labels) at the boundary of the area.

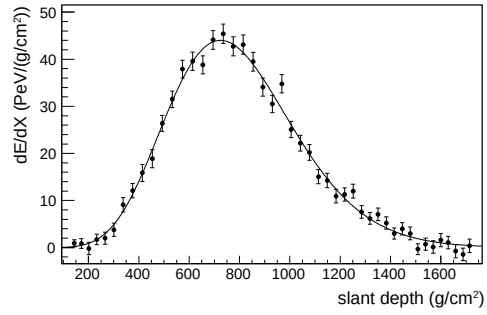


(b) Detector tank of the Auger experiment.

Fig. 16.20 The Pierre Auger Observatory in Argentina [29, 783]. The surface detectors are located at an altitude between 1300 and 1400 m and thus at an atmospheric depth of about 880 g/cm^2.



(a) The light track of an air shower recorded by fluorescence telescopes (the circles indicate the pixels of the telescope cameras). The black line is the reconstructed image of the shower axis (non-linearly distorted). The different colours indicate the temporal sequence of the signals in the pixels of the telescope camera. The squares at the bottom designate the surface stations that were triggered.



(b) The longitudinal profile of the energy deposition of a shower in the atmosphere as reconstructed from measurements of a fluorescence telescope. The curve is a fit of the Gaisser–Hillas function (16.28) to the data. The reconstructed energy is about 3×10^{19} eV.

Fig. 16.21 Principle of the measurement of the shower profile by a fluorescence telescope of the Pierre Auger experiment (from [30], with kind permission of Elsevier).

With the simultaneous detection of the shower by surface detectors and by fluorescence telescopes which both provide an independent energy measurement, systematic uncertainties in the absolute energy calibration can be substantially reduced. Because of the steep spectrum an incorrect energy calibration also leads to a shift of the absolute height of the measured spectrum. The fluorescence telescopes also allow for a determination of the shower maxima (see fig. 16.21(b)) from which one can estimate the mass composition.

16.5 TeV gamma telescopes

16.5.1 Overview

Because of the strongly falling fluxes of cosmic gamma radiation the effective areas of balloon and satellite detectors eventually become too small for high gamma energies. For gamma radiation above about 100 GeV ('very high energy', VHE) one exploits the interaction with the atmosphere leading to the formation of electromagnetic showers (sections 15.2 and 16.3.4).

Photons with energies high enough so that shower particles can reach the ground can in principle also be measured with the earthbound air shower detectors for charged cosmic rays (section 16.4). In this section, however, we want to concentrate on the imaging telescopes specialised in TeV-gamma astronomy. The 'imaging atmospheric Cherenkov telescopes' (IACT) detect the Cherenkov radiation of air showers initiated by VHE photons. Examples for IACTs are H.E.S.S. (High Energy Stereoscopic System) in Namibia [521], MAGIC (Major Atmospheric Gamma-ray Imaging Cherenkov Telescopes) on La Palma [683] and VERITAS (Very Energetic Radiation Imaging Telescope Array System) in Arizona [960]. Hence there are sites both in the northern and southern hemispheres, which is important for the coverage of the full sky. The galactic centre, for example, is only visible from the southern hemisphere. The next generation of VHE-gamma instruments is the project 'Cherenkov Telescope Array' (CTA) [333].

16.5.2 Cherenkov light from hadron and gamma showers

Cherenkov telescopes primarily detect electrons and positrons of the electromagnetic component of an air shower. Since in hadronic showers the electromagnetic component also dominates (see section 16.3.2), other criteria besides the Cherenkov light yield are necessary to separate the VHE gammas from the high background from charged cosmic rays (mostly hadrons). The most important criterion is the more compact lateral particle distribution in (electromagnetic) gamma showers (section 16.3.4) and the consequently more concentrated light distribution on the detector level (fig. 16.22) [184].

Compared to KASCADE-Grande and TUNKA-133, which have a similar size and cover the same energy range, IceTop has the advantage that the high-energy muon bundle in IceCube can be measured in coincidence with showers in IceTop. Because of the large depth of IceCube only the high-energy muons above about 500 GeV reach the IceCube detector. Such muons, which come from the first interactions in the high atmosphere, appear in the deep ice in narrow bundles with a muon multiplicity that strongly depends on the mass composition. The essential relations for the emission of Cherenkov radiation are explained in chapter 11. A particle with the velocity β radiates Cherenkov light in air with the index of refraction n under an angle $\cos\theta = 1/(\beta n)$. The maximum angle θ_{max} is reached by relativistic particles with $\beta \approx 1$ leading to

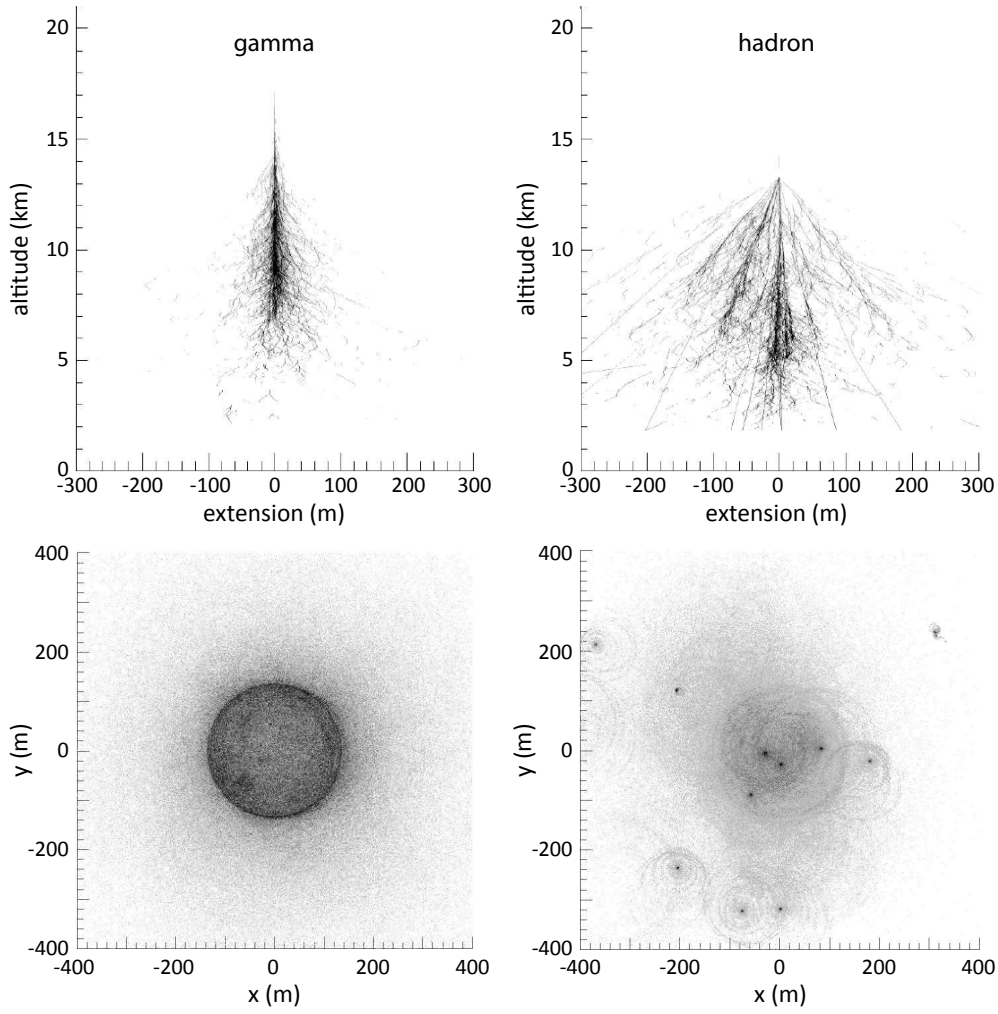


Fig. 16.22 Simulation of a 300-GeV gamma shower (left) and of a 1-TeV hadron shower (right), which both produce about the same amount of light on the ground. The upper pictures show the shower development for the particles which emit the Cherenkov light, the lower pictures show the *light pool* on the ground. Source: [626], using unpublished simulation studies described in [184].

$\cos\theta_{max} = 1/n$. In the atmosphere the index of refraction varies with height h because $n-1$ is proportional to the density:

$$n(h) = 1 + (n_0 - 1)\frac{\rho(h)}{\rho_0} \approx 1 + (n_0 - 1)\,\exp\left(-\frac{h}{8.4\,\mathrm{km}}\right)\,, \qquad (16.37)$$

where we use on the right side the approximation (16.10). The refractive index of air is $n_0 \approx 1.00029$ under standard conditions (STP)[7] (table 11.1 on page 449). An emission at height h has at height h_T of the telescope the maximum radius:

[7] The refractive index of air can be interactively calculated for any conditions on the web page [726].

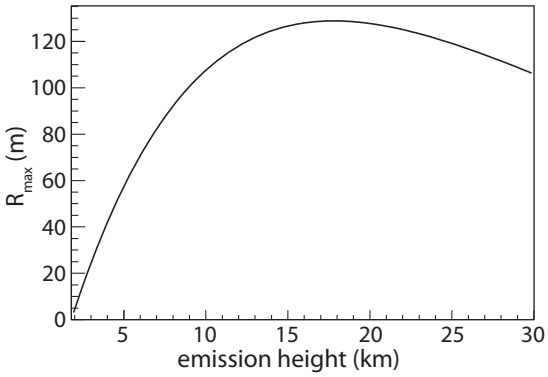


Fig. 16.23 Maximum radius of the Cherenkov light at the height of the H.E.S.S. telescope as function of the emission height according to (16.38) (adapted from [626]).

$$R_{max}(h) = (h - h_T)\,\tan\theta_{max} = (h - h_T)\,(n(h)^2 - 1)\,. \tag{16.38}$$

The dependence of R_{max} on the emission height is shown in fig. 16.23. The geometric effect of a decrease of the radius with decreasing emission height is partly compensated by the increase of the index of refraction so that the curve flattens out above about 10 km, leading to a maximum radius of $R_{max} \approx 120$ m. Most of the Cherenkov photons are emitted at the shower maximum which, for the relevant gamma energies from 100 GeV to several TeV, lies at a height between 7 and 10 km (table 16.2 on page 679), corresponding to $R_{max} \approx 100$ m.

The structures of the light distribution at ground level shown in fig. 16.22 (bottom) result from the limitation of the maximum radius and the strong collimation of electromagnetic showers. This leads for gamma showers to a concentration within a circle with an enhancement at the edge due to photons from the shower maximum. For hadronic showers the many electromagnetic sub-showers lead to a broader distribution with many rings.

Since the velocity of the Cherenkov photons does not deviate much from the particle velocity in the shower the light emitted at all heights reaches ground nearly simultaneously. This leads to short light flashes of a few nanoseconds and allows for an efficient suppression of background from stray light by triggering with a short time window. The measurements are usually only performed in moonless nights with clear sky.

16.5.3 Shower reconstruction

The detection principle is displayed in fig. 16.24. A photon hitting the atmosphere initiates an electromagnetic shower at a height of about 10 km. The electrons and positrons in the shower radiate Cherenkov light in a cone which has, as explained, a radius of about 120 m at ground level. This light is focused by the telescope mirror onto a camera which is positioned near the focal plane of the mirror such that an average shower is optimally imaged. The camera consists of an array of photosensitive detectors, typically photomultiplier tubes (PMTs) which are sensitive to single photons (see chapter 10). A point in the camera plane corresponds to the angle θ under which the photon is seen ($\theta = 0$ corresponds to the symmetry axis of the telescope mirror). Because of the small field of view (about 5°) θ can be expressed in linear approximation by the orthogonal projections (θ_x, θ_y). The covariance matrix of the intensity distribution in

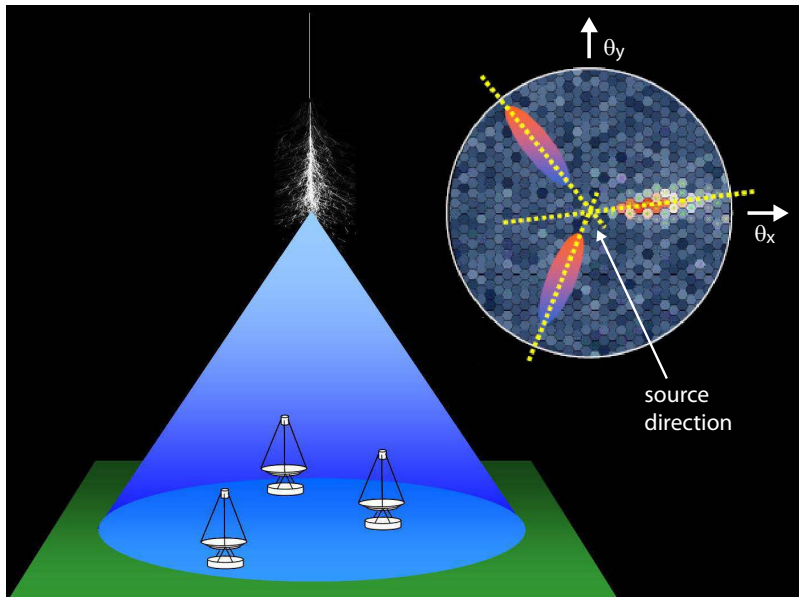


Fig. 16.24 Principle of Cherenkov telescopes for the detection of VHE gammas. In this example three telescopes measure the Cherenkov light permitting the stereoscopic reconstruction of the shower. The picture on the right shows the camera plane of one telescope with the recorded intensity distribution. The intensity ellipses measured by the two other cameras are overlaid in order to visualise the direction reconstruction. Source: H.E.S.S. Collaboration.

the (θ_x, θ_y) plane defines an ellipse whose position and extension in the direction of the principle axes are described by the so-called Hillas parameters [524].

The ratio of width and length of the ellipse is an important criterion for the suppression of background from hadronic showers. However, the background cannot be completely suppressed, but has finally to be statistically subtracted. This is best achieved in observations of point sources or sources with a smaller extent than the field of view of the telescope, because then it is possible to measure simultaneously the background close to the source. Else, signal and background can be measured under the same conditions by alternately pointing the telescope to the source and to the background within some preselected time intervals (*wobble mode*).

A system of several telescopes permits a stereoscopic shower reconstruction, yielding a particularly precise direction determination (about 0.1°). The direction is determined by fitting all long axes of the intensity ellipses to a common intercept point, as can be seen in the inserted picture in fig. 16.24.

16.5.4 Performance characteristics of Cherenkov telescopes

As example for the performance of VHE telescopes we list here the technical data of the H.E.S.S. experiment in the configuration with four telescopes[8] [744]:

[8] H.E.S.S. was upgraded by a fifth telescope with a much larger mirror which mainly reduces the energy threshold to below 50 GeV.

field of view	5°
sensitive area	50 000 m^2
energy threshold	100 GeV
direction resolution	stereoscopic: 0.05°
energy resolution	$\Delta E/E < 20\%$
sensitivity (5σ)	0.01 Crab in 25 h

The 'Crab unit' is defined as the photon rate, which the telescope registers when pointing towards the Crab Nebula. The Crab Nebula is the strongest source of VHE photons, which were first observed in 1989 by the Whipple telescope [973].

16.6 Neutrino detectors

Experiments which study the properties of neutrinos or use neutrinos as probes for the investigation of astrophysical phenomena have significantly contributed to the broadening of our knowledge about particles and the cosmos. Reviews of experiments and the theoretical description of results can be found in [1026, 461].

Neutrinos are detected through weak interactions with matter. The cross section for neutrino–nucleon reactions is very small (fig. 16.25) but up to energies of about 1 TeV it increases linearly with energy, such that with increasing energy the measurement of neutrino reactions becomes simpler. Accelerators provide neutrino beams in the range of some GeV to several 100 GeV neutrino energies, see section 16.6.4.

In this chapter we mainly want to discuss the detection of neutrinos that are not generated at accelerators, such as astrophysical or reactor neutrinos. Neutrinos with MeV energies are generated in nuclear burning processes in stars, in particular also in our sun and in supernova explosions. The highest neutrino energies, which might go up to more than 10^{20} eV, are expected in connection with high-energy cosmic rays (section 16.3) which, on their way through the cosmos, interact with matter and radiation producing, amongst others, mesons (pions, kaons, ...) that can decay into final states containing neutrinos. Both ends of the energy scale pose particular challenges for the detector technology. Characteristic for neutrino detectors are large target volumes, which usually also serve as detector media, and simple, cost-efficient readout procedures making the coverage of the large volumes possible within budget constraints. With such requirements the detectors are often also suited for the detection of rare or exotic phenomena, such as proton decay or the appearance of magnetic monopoles. For example, the observation of neutrino oscillations and of the neutrinos from the supernova 1987A [526, 205] was achieved with detectors that were originally designed for the detection of proton decay.

16.6.1 Detection of solar neutrinos

The phenomenon of flavour oscillation of neutrinos was observed for the first time for solar neutrinos. A description of the experiments and their interpretation can be found for example in [606, 1026, 461].

In the sun neutrinos are produced in different fusion reactions, which partly lead to mono-energetic neutrinos (if they originate from electron capture processes, see appendix A.2) or to a continuous spectrum up to about 10 MeV (fig. 16.26). For each neutrino an additional energy of about 13 MeV is on average dissipated and finally radiated by the sun. This relates the sun's luminosity to the neutrino rate.

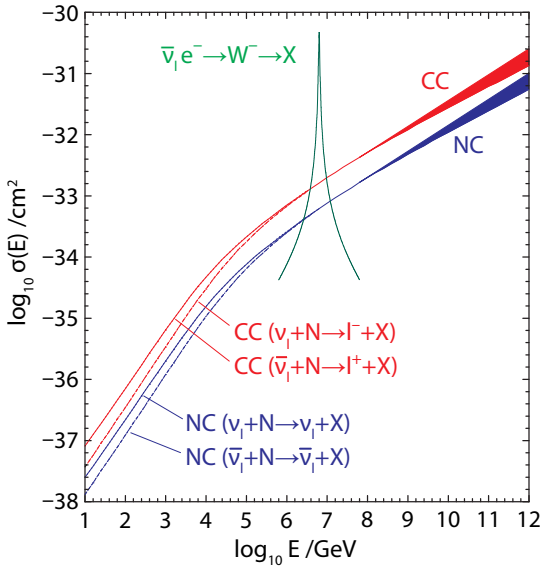


Fig. 16.25 Neutrino–nucleon and antineutrino–nucleon cross sections for charged and neutral currents (CC, NC) as function of the neutrino energy. In charged current reactions the (anti)neutrino turns into the associated charged lepton, while in neutral current reactions the (anti)neutrino reappears in the final state. The plot shows in addition the so-called Glashow resonance arising from the scattering of antineutrinos off an electron at centre-of-mass energies around the mass of the W boson. The plot is based on calculations in [440] (adopted from [942]).

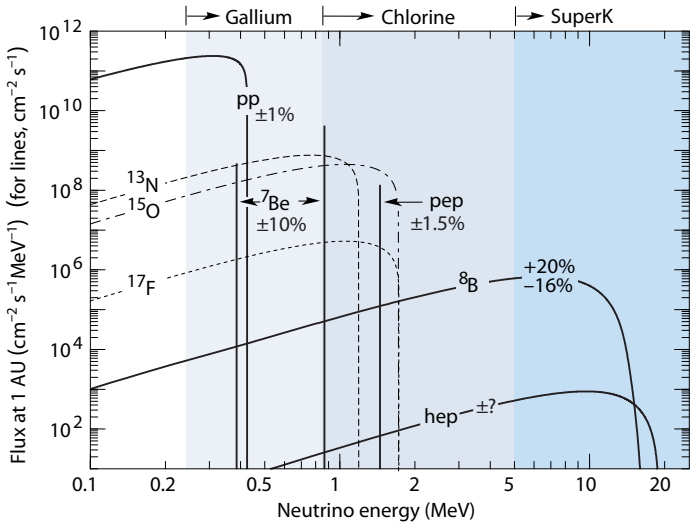


Fig. 16.26 The theoretically calculated neutrino spectrum with contributions from different reactions. Also shown are the threshold energies for neutrinos in different detector media. Further explanations can be found in [723] from where this plot has been taken.

Presuming equilibrium between production and emission of solar energy the neutrino flux from the sun on Earth can be determined from the radiant power per area on Earth, which is the so-called solar constant $S = 8.5 \times 10^{11}\,\mathrm{MeV\,cm^{-2}\,s^{-1}}$ [628]:

$$\Phi_\nu = S/13\,\mathrm{MeV} \;\approx\; 6.5 \times 10^{10}\,\mathrm{cm^{-2}\,s^{-1}}\,. \tag{16.39}$$

The large neutrino flux is confronted with a very small cross section of the order of $10^{-44}\,\mathrm{cm^2}$ at these low energies. This means that one needs more than 10^{26} target particles to detect one single reaction within about a year. Such a number of target particles has, for example, 10 kg chlorine which was used for the first detection of solar neutrinos (see below). If one additionally accounts for energy thresholds, background

and efficiencies one easily arrives at several 100 t in order to obtain significant results. In order to shield the background from cosmic radiation these detectors are in installed in underground caverns, like mines or tunnels (section 2.3.3 on page 18).

Radiochemical neutrino detection. Solar neutrinos were first observed by the experiment of R. Davis in the Homestake Mine in South Dakota employing 615 t perchloroethylene (C_2Cl_4). Data where taken between 1970 and 1994. The detection reaction is the 'inverse β decay'

$$^{37}\mathrm{Cl} + \nu_e \rightarrow {}^{37}\mathrm{Ar} + e^- \,. \tag{16.40}$$

With a neutrino energy threshold of 814 keV the experiment is not sensitive to the most abundant pp process; see fig. 16.26.

The produced ^{37}Ar is volatile in the solution and was extracted once a month by flushing the tank with helium. The extraction interval was chosen as the period in which an equilibrium between production and decay (with a half-life of 35 days) is reached leading to a constant ^{37}Ar concentration. The ^{37}Ar detection predominantly proceeds via the electron capture (EC) process

$$^{37}\mathrm{Ar} + e^- \rightarrow {}^{37}\mathrm{Cl} + \nu_e \tag{16.41}$$

in a proportional counter (section 7.6.1), which is filled with the extracted ^{37}Ar together with natural argon and an admixture of methane. The process is detected through Auger electrons which are emitted after the electron vacancy caused by the EC has been filled again (see also page 77 and fig. 3.40). The available energy of 2.82 keV is usually allocated to several Auger electrons, which thus individually have a shorter range in the gas of the counter tube and hence deposit the energy more locally than for example Compton electrons of the same energy. This yields a short rise time of the ^{37}Ar signals, which can be exploited for the rejection of background. The results are usually expressed in SNU (*solar neutrino unit*):

$$1 \text{ SNU} = 10^{-36} \text{ neutrino captures per target atom and second.} \tag{16.42}$$

The detection rate is only about 15 atoms per month, a number which gives an idea about the experimental difficulties of such a project. The result of the Homestake experiment is $2.56 \pm 0.16 \pm 0.16$ SNU (the uncertainties are statistical and systematic, respectively) where 6–9 SNU have been theoretically expected [312]. This result has been confirmed by other experiments using different methods and was finally explained to be caused by neutrino oscillations, which are the oscillatory conversions of the three neutrino species amongst each other (see the corresponding review article in [765]). The explanation of the deficit of the solar neutrinos is that the electron neutrinos transform into another neutrino species and therefore elude detection as electron neutrinos.

Chlorine experiments with an energy threshold of 814 keV cannot detect neutrinos from the dominant pp process which have energies below 420 keV (fig. 16.26). The experiments GALLEX [501] and SAGE [20] have detected the pp process by the inverse β decay in gallium:

$$^{71}\mathrm{Ga} + \nu_e \rightarrow {}^{71}\mathrm{Ge} + e^- \,. \tag{16.43}$$

The reaction has a threshold for neutrino energies of 233 keV. Germanium returns into gallium by electron capture with a half-life of 11.43 days:

$$^{71}\mathrm{Ge} + e^- \rightarrow {}^{71}\mathrm{Ga} + \nu_e \,. \tag{16.44}$$

The challenge of these radiochemical neutrino experiments are counting rates of few events per week with the corresponding requirement that the background rate has to be significantly smaller. Besides a good shielding it is essential that the counter tube is kept as small as possible with little potentially radiating material. This has led to the development of miniaturised proportional counters [1000].

Real time neutrino detection. At higher neutrino energies, above some MeV, the neutrinos can be directly detected in *real time* through their interaction with the electrons of the target medium:

$$\nu_e + e^- \rightarrow \nu_e + e^- . \tag{16.45}$$

If sufficient recoil energy is transferred to the electron it can be detected, for example through the Cherenkov light it generates in a water tank or through an ionisation trace it leaves in a drift chamber. In both cases the electron direction, approximately corresponding to the neutrino direction, can be measured.

In the 1970s and 1980s facilities for the detection of proton decay were built. Characteristic for these detectors is their large volume (one needs more than 10^{34} protons to reach the current limit of the proton lifetime of more than 10^{34} years) and the possibility that the decay reaction can be kinematically reconstructed, for example the initially favoured decay $p \rightarrow \pi^0 e^+$. So far proton decay has not been discovered. However, at the latest after the observation of the supernova SN1987A in the Small Magellanic Cloud (see e.g. [461, 1026]) with these detectors it became evident that they are also well suited for the detection of solar and supernova neutrinos as well as higher energetic atmospheric neutrinos.

Examples are the experiments Kamiokande and the successor Super-Kamiokande at a depth of 1000 m in the Mozumi mine in Japan. The Super-Kamiokande detector (fig. 16.27) [432] consists of a tank filled with about 50 000 t ultra-pure water in a cylindrical container with a height of 41.4 m and a diameter of 40 m. The sensitive inner volume of the detector (33.8 m diameter and 36.2 m height) is optically separated from an outer volume. Both volumes are surrounded by PMTs, 11 146 in the inner and 1885 in the outer volume. Signals from the outer volume are used as veto against cosmic radiation.

The electrons hit by the solar neutrinos in the reaction (16.45) generate a Cherenkov cone which is projected onto the PMT wall (fig. 16.28(a)). From the measurement of time and pulse height of the PMT signals and their spatial distribution one can determine the interaction vertex and the direction of the charged particles. The sharpness of the Cherenkov ring allows one to draw conclusions on particle properties (see the corresponding discussion of fig. 11.26 on page 473 related to atmospheric neutrinos at higher energies). The electron in fig. 16.28(a) has an energy of about 12.5 MeV yielding a range of 6.1 cm in water. The image of the emitted Cherenkov light has the shape of a ring according to the principle of proximity focusing (see page 453).

For events which are classified as solar neutrinos the measured directions show a peak in the direction of the sun (fig. 16.28(b)). However the number of electron neutrinos in the peak is lower than expected without neutrino oscillations. Hence also for solar neutrinos at higher energies a deficit relative to the expectations without oscillations was found.

A summary of the results of experiments with different energy thresholds can be found in [606] and updated in the review 'Neutrino Mass, Mixing and Oscillations'

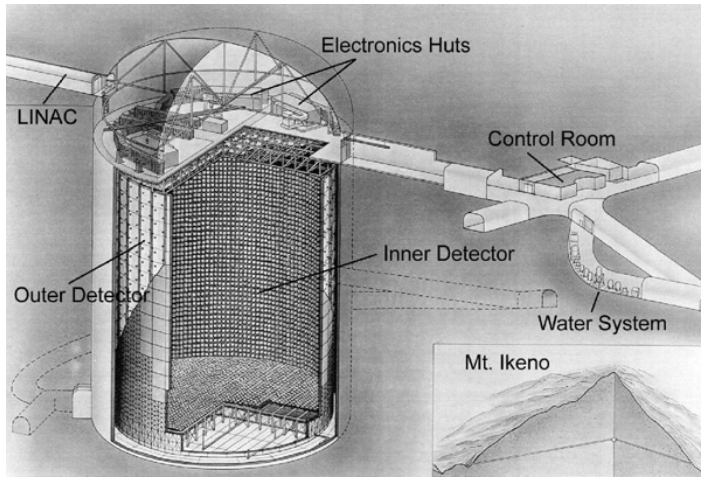
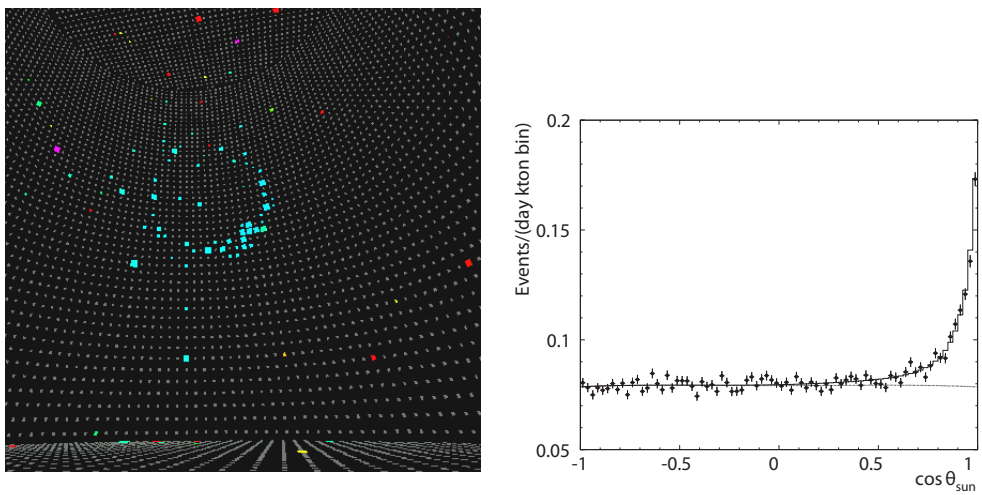


Fig. 16.27 Perspective drawing of the Super-Kamiokande detector displaying the inner and outer detector volume and the electronics and control facilities. The linear accelerator (LINAC) delivers low-energy electrons for calibration. The detector is installed about 1000 m below the peak of the Mount Ikeno (bottom right). Source: Kamioka Observatory, ICRR, University of Tokyo and Super-Kamiokande Collaboration.



(a) (b)

Fig. 16.28 (a) Example of a solar neutrino event in Super-Kamiokande (Source: [159]). An electron neutrino produces an electron in the inner detector whose Cherenkov radiation is recorded by PMTs at the wall of the water tank. The energy of this electron was determined to be about 12.5 MeV (range in water about 6.1 cm). (b) Neutrino rate per kiloton water and per day as function of the direction of incidence relative to the position of the sun, measured by Super-Kamiokande (from [327], with kind permission of the American Physical Society). The peak in the direction of the sun, that is, at $\cos\theta_{\text{sun}} = 1$, accounts for only 40% of the rate expected without oscillations, but is consistent with the observations of other neutrino oscillations experiments.

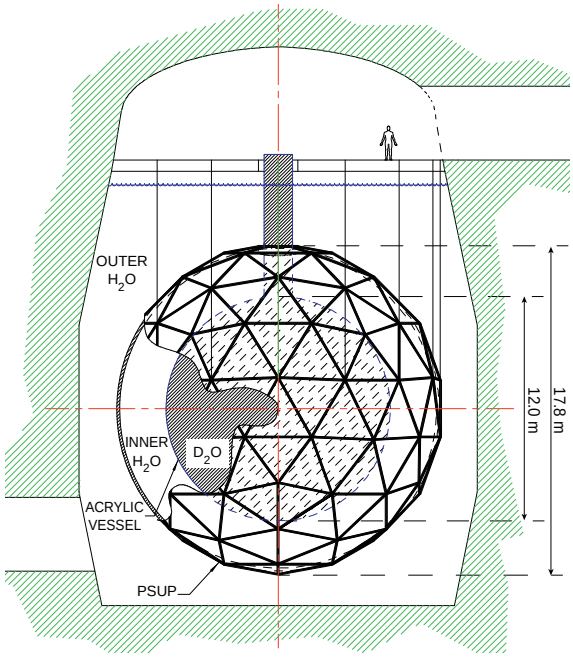


Fig. 16.29 Sudbury Neutrino Observatory (SNO). The picture shows the cavern (22 m wide, 34 m high) filled with light water (H_2O) in which an acrylic sphere filled with heavy water (D_2O) is suspended. An open grid structure, centred around the sphere carries 9438 PMTs (PSUP = *PMT support structure*), each directed towards the inside of the sphere. The acrylic sphere is UV-transparent, matching the photosensitivity of the PMT cathode. From [223], with kind permission of Elsevier.

in [765].

Total flux of all neutrino flavours from the sun. Since it was observed that electron neutrinos disappear (*disappearance experiment*) it was important to show if indeed μ or τ neutrinos have emerged (*appearance experiment*). Measurements of the Sudbury Neutrino Observatory (SNO) in Canada have shown that indeed the sum of the fluxes of all neutrino species corresponds to the expected total flux of solar neutrinos within the measurement uncertainties. The sum of the fluxes was measured through the neutral current (NC) reactions in which all neutrino species can participate.

The SNO experiment consists of 1000 t ultra-pure heavy water (D_2O) in a spherical acrylic container which is surrounded by a H_2O shielding (fig. 16.29). As in Super-Kamiokande the reactions are measured in real time through the Cherenkov radiation emitted by the charged particles. The detector is sensitive to ^{8}B solar neutrinos (high-energy part of the spectrum in fig. 16.26) through the reactions:

$$\begin{array}{lcll} \nu_e + d & \rightarrow & e^- + p + p & (\mathrm{CC})\,, \\ \nu_l + d & \rightarrow & \nu_l + p + n & (\mathrm{NC})\,, \\ \nu_l + e^- & \rightarrow & \nu_l + e^- & (\mathrm{ES})\,. \end{array} \qquad (16.46)$$

The first reaction, deuteron disintegration through the charged current (CC), can only be initiated by electron neutrinos and thus measures the ν_e flux. The second reaction, deuteron disintegration through the neutral current (NC), has the same cross section for all neutrino flavours and therefore measures the sum of neutrino fluxes. The elastic scattering off an electron (ES) is also sensitive to all flavors, but with reduced sensitivity to ν_μ and ν_τ. The experimental signatures are:

CC: The electron emits Cherenkov radiation.

NC: The neutron is captured by the deuterium; the gamma quantum of 6.25 MeV emitted in this process is detected through photoeffect or Compton scattering.

ES: The emitted Cherenkov light of the electron differs from that of the CC electron mostly by peaking of the angular distribution in the neutrino direction. Hence the reaction can at the same time determine the direction to the sun.

The measured total flux is in agreement with the flux of all neutrino flavours expected by the standard solar model (SSM) (details can be found in the review article 'Neutrino Masses, Mixing, and Oscillations' in [765]).

16.6.2 Atmospheric neutrinos

In the air showers initiated by cosmic radiation (section 16.3.2) electron and muon neutrinos are also generated, for example in the reaction chain:

$$p + \text{nucleus} \to \pi^- + X\,, \qquad \pi^- \to \mu^- + \bar{\nu}_\mu\,, \qquad \mu^- \to e^- + \nu_\mu + \bar{\nu}_e\,. \tag{16.47}$$

In the corresponding generation and decay of π^+ mesons, the charge conjugated neutrinos are produced. These neutrinos have typical energies in the range of GeV and can be observed by the above-discussed large-volume detectors like Super-Kamiokande and SNO (fig. 11.26). The spectra of atmospheric neutrinos have been measured up to TeV energies by the telescopes for high-energy neutrinos, which we discuss in section 16.6.5. Super-Kamiokande found that atmospheric neutrinos also feature oscillations. The essential oscillation parameters could be determined by comparing the zenith angle dependence of the electron and muon neutrino fluxes [431].

Neutrinos can be identified by requiring that no signal is found in the veto detectors that typically enclose the sensitive volumes ('contained events') or that the neutrino candidates come from below, which means they passed through the Earth.

16.6.3 Reactor antineutrinos

The discovery of neutrinos. In 1930 Pauli formulated his neutrino hypothesis in order to explain the continuous electron spectra in nuclear β decay (see e.g. [461]). Only in 1953, after nuclear reactors became available, could F. Reines and C. Cowan confirm the existence of neutrinos by direct detection (Nobel Prize 1995)[9] [814]. They observed the inverse neutron decay

$$\bar{\nu} + p \to e^+ + n\,, \tag{16.48}$$

initiated by antineutrinos from a reactor with a flux of $10^{13}\,\mathrm{cm^{-2}\,s^{-1}}$. Employing a water tank of $0.08\,\mathrm{m^3}$ (about 10^{27} protons) as target the expected cross section of $\sigma_{\bar{\nu}} \approx 10^{-43}\,\mathrm{cm^2}$ became measurable (about 4 reactions per hour).

Figure 16.30 shows the principle of the apparatus with which the neutron and the positron of reaction (16.48) were detected (a nice description can be found in [813]). The target medium is water in which cadmium chloride ($CdCl_2$) is dissolved. The two water tanks are placed between three tanks filled with liquid scintillator. The positron generated in the reaction annihilates with an atomic shell electron into a photon pair; the neutron is thermalised by scattering in water on a time-scale of microseconds and is then detected through the capture by a cadmium nucleus, which has a high

[9] The prize was only awarded to Reines, Cowan had deceased already.

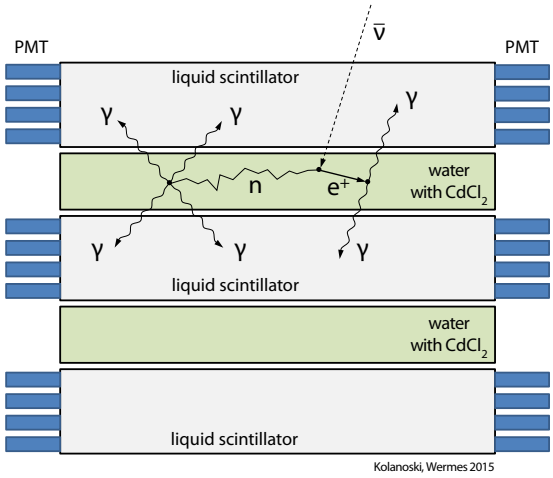


Fig. 16.30 Principle of the apparatus of Reines and Cowan for the detection of antineutrinos. The detection reaction (16.48) is explained in the text. The water tanks as targets are embedded between tanks with liquid scintillator in which the photons from the final state deposit their energy.

capture cross section for thermal neutrons. The cadmium nucleus becomes excited by the capture and releases its energy in the form of photons. Neutrino events are recognised if all photons in the final state deposit their energy in the neighbouring liquid scintillator tanks, usually by photoeffect or Compton scattering. The signature of an antineutrino event is the signal caused by the annihilation photons, which is followed by a signal of the nuclear gamma quanta with a delay of some microseconds. In order to suppress background it was required that, amongst others, no coincident signal be observed in the scintillator tank that is not a neighbour of the water tank where the reaction occurred.

Oscillation experiments with reactor antineutrinos. Neutrino oscillations have also been observed for antineutrinos from reactors (see [1026]). The first reactor oscillation experiment with antineutrinos was KamLand in Japan employing a 1000 t liquid scintillator detector. KamLand is surrounded by 53 reactors at an average distance of 180 km. The expected total flux of antineutrinos and their energy spectra have been very precisely determined.

For a more precise measurement of oscillation parameters detectors are deployed at different distances from one or several reactors. This principle is applied by the experiments Daya Bay in China [343], Double Chooz in France [362] and RENO in Korea [815]. Most importantly, these experiments have quite precisely measured the smallest mixing angle (θ_{13}).

16.6.4 Neutrinos from accelerators

Since the 1960s high-energy muon neutrino beams have been produced at proton accelerators and used for the investigation of neutrino–hadron interactions, in particular for measurements of the structure functions of nucleons. The first focused beam was produced in 1963 using 20-GeV protons from the proton synchrotron at CERN. An overview of the development of such beams is given in [629]. A classical example for a neutrino detector at an accelerator beam is the CDHS detector [532].

High-energy neutrino beams are generated via the decay of high-energy mesons, mostly pions and kaons, according to the following scheme:

- A high intensity proton beam is shot onto a target (lead, tungsten, ...), typical intensities being about 10^{20-21} protons on target (POT) per year.
- In the target, amongst others, charged mesons (mainly pions and kaons) are produced.
- A magnet system, called 'magnetic horn' [456], focuses the mesons of a selected charge and in a certain momentum range towards the beam axis. The horn can be tuned to select a range of momenta to be best focused.
- The mesons decay while travelling through a long channel.
- At the end of the channel the charged particles are absorbed in the ground soil.
- After this filtering a beam of muon neutrinos (with little contamination by electron neutrinos) is available whose energy and angular distribution can be calculated from the generation process and the subsequent focusing and filtering.

With the discovery of neutrino oscillations beams of muon neutrinos have found further application domains. In so-called 'long-baseline' experiments appearance and disappearance of neutrinos of a certain flavour can be tested at different distances to the neutrino source. Two such experiments, OPERA and DONUT (fig. 6.10), have already been discussed in the context of applications of photoemulsions in section 6.3.3.

Another such experiment is T2K ('Tokai-to-Kamioka') [23], a long-baseline neutrino experiment in Japan. In 2011 T2K announced the first indications for $\nu_\mu \rightarrow \nu_e$ oscillations (ν_e appearance). The experiment employs a ν_μ beam with peak intensity around 0.6 GeV produced at J-PARC (Japan Proton Accelerator Research Complex) and compares the ν_μ reaction rates measured in a near and a far detector. The *far detector* is Super-Kamiokande (page 693 and fig. 16.27) at a distance of about 295 km away from J-PARC ('long-baseline'). Both in Europe and USA there are or were experiments with baselines of more than 700 km. From CERN in Geneva a beam (*CERN Neutrinos to Gran Sasso*, CNGS) was directed towards the laboratory in the Gran Sasso (LNGS) near Rome (between 2006 and 2012). From the Fermi Laboratory near Chicago (FNAL) a similar neutrino beam, called NuMI (Neutrinos at Main Injector) [49], is directed towards the Soudan Mine, 735 km from FNAL in Northern Minnesota. Originally it aimed at the far detector of the long-baseline experiment MINOS (running until 2016 [766]) and currently (2018) still serves the experiment NOνA (NuMI Off-Axis ν_e Appearance) [48], located somewhat downstream at a distance of 810 km from FNAL. For future long-baseline experiments FNAL is currently preparing the Long-Baseline Neutrino Facility (LBNF) which will house the Deep Underground Neutrino Experiment (DUNE) with the far detector at the Sanford Underground Research Facility (SURF) at a distance of 1300 km from FNAL [400]. An overview of underground laboratories and facilities can be found in section 2.3.3.

16.6.5 Neutrino detection at high energies

16.6.5.1 Neutrino astronomy

Since the 1990s detectors have been operating with the goal to search for high-energy neutrinos originating from cosmic sources. The construction of such instruments is motivated by the quest for the sources of cosmic radiation up to the highest observed energies and for the understanding of the underlying acceleration mechanisms. For the bulk of the charged cosmic particles there is no direct correlation between the direction of incidence and the direction to the source because of deflections by galactic

and intergalactic magnetic fields. While gamma radiation is not affected by magnetic fields it will be absorbed on intergalactic distances in the order of megaparsecs,[10] in particular through pair production processes in interactions with the background light from stars:

$$\gamma_{\mathrm{HE}} + \gamma_{\mathrm{Stern}} \rightarrow e^+ e^- \,. \tag{16.49}$$

As a probe for cosmic sources gamma radiation has another drawback: it does not provide unambiguous evidence for hadron acceleration, as needed for an indication of the origin of cosmic rays, because high-energy gamma quanta can also be produced by accelerated electrons, for example through the inverse Compton effect (page 81 in section 3.5.4). Neutrinos, in contrast, propagate along straight lines from the source and pass matter nearly without interaction.

High-energy neutrinos originate in or near the accelerators via the production and weak decay of hadrons, in particular charged pions, similar to air showers (fig. 16.12). Such reactions can also occur in interstellar or intergalactic matter or radiation if the respective energy threshold is exceeded. In particular, pion production by scattering of cosmic rays off the cosmic microwave background radiation (CMB) plays an important role at very high energies (GZK cut-off, page 684).

The detection of high-energy neutrinos proceeds via the charged particles that are generated in a neutrino interaction with matter. In charged current interactions, muon neutrinos produce muons which can be reconstructed as tracks while electron neutrinos produce electrons which are observed through electromagnetic showers[11] (here often denoted as *cascades*), and τ neutrinos can be observed both as tracks and cascades. In all cases an additional hadronic shower might be observable at the vertex of the neutrino reaction.

16.6.5.2 Detection of muon neutrinos

The muon neutrinos are mainly detected through the CC interactions of the neutrino with the nuclei of the matter surrounding the detector (rock, water, ice, ...):

$$\nu_\mu + N \rightarrow \mu^- + X \,. \tag{16.50}$$

The muon emerging from this reaction is then registered. In order to suppress signals from particles of cosmic ray air showers, which are many orders of magnitude more abundant, one looks for neutrinos coming from below, which means that the neutrino must have passed through the Earth. In addition one attempts to make the shielding against cosmic rays above the detector as thick as possible by placing the detector in deep mines, under water or in ice. At a water depth of 1 km the ratio of muons coming from above to those coming from below is still about 10^6. The muons coming from below are predominantly generated by atmospheric neutrinos.

Figure 16.31 shows the absorption length of neutrinos in water and in the Earth as function of energy. The range of the muons, as displayed in fig. 16.32, is essential for the probability of neutrino detection through produced muons because the effective target volume for neutrinos increases with the range.

In the currently existing high-energy neutrino detectors muons are detected through the Cherenkov light generated in water or ice with optical sensors (fig. 16.33).

[10] 1 parsec = 1 pc = 3.085 677 581 49 × 10^{16} m = 3.262 ... ly (lightyears).

[11] At very high energies, in ice or water above about 1 PeV, the LPM effect affects the shower development; see appendix G.

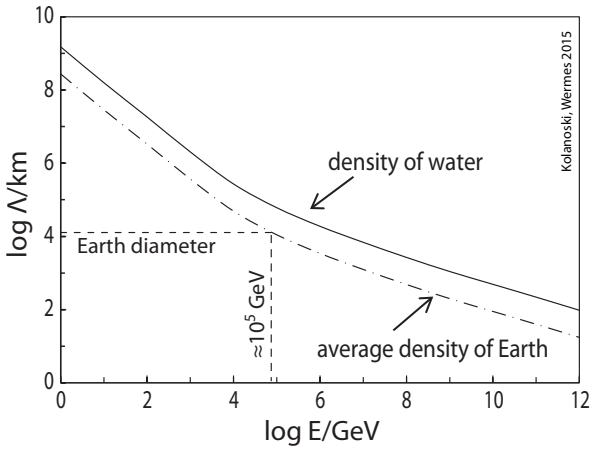


Fig. 16.31 Average range Λ of neutrinos in water and in the Earth (mean density 5.5 g/cm^3) as a function of energy. At about 10^5 GeV the range in the Earth falls below the Earth's diameter, meaning that the Earth becomes opaque for neutrinos as the neutrino energy increases. The curves have been calculated with the cross sections in fig. 16.25.

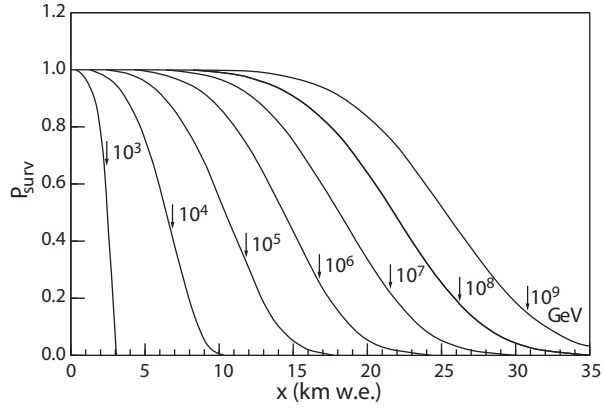


Fig. 16.32 Survival probability of muons as a function of the penetration depth in rock (given in units 'km water equivalent'). The numbers on the curves are the initial muon energies in GeV. For a given energy the ranges are largely fluctuating. The arrows indicate the range calculated from the mean energy loss per path length. The obvious difference between the mean range calculated with and without fluctuations is discussed in [663]. In order to obtain the range in 'standard rock' ($\rho = 2.65\,\mathrm{g/cm^3}$) in units of km the range values given in units km w.e. have to be divide by a factor 2.65. From [663], with kind permission of the American Physical Society.

While the detection principle is based on the Cherenkov effect in the same way as in detectors for lower energies, like Super-Kamiokande, the detectors for cosmic neutrinos have to be much larger than those for solar and atmospheric neutrinos because of the much smaller fluxes. Correspondingly the distances between the individual light sensors are so large that an imaging of the Cherenkov rings is not possible. Instead the events are reconstructed from the spatial distribution of the hit light sensors and the arrival times of the Cherenkov photons. The measurement of the arrival time of the light with a precision of few nanoseconds allows for a reconstruction of the muon direction, which is an approximation for the neutrino direction. The mean angle of the muon relative to the primary neutrino in (16.50) becomes smaller with increasing

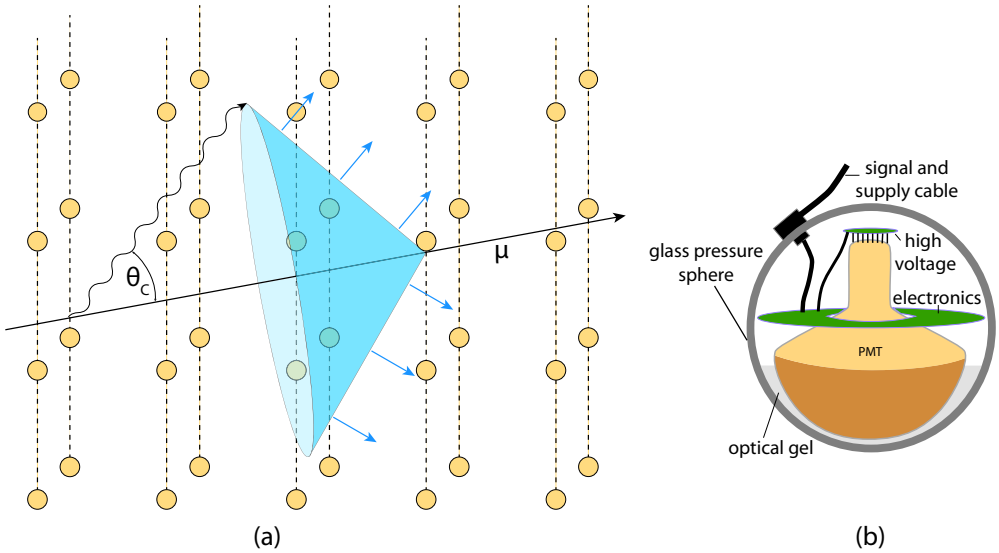


Fig. 16.33 Typical arrangement of a neutrino telescope. (a) A muon produces Cherenkov light in water or ice ($\theta_c(\beta = 1) \approx 40°$) which is then recorded by photomultipliers (PMTs). The depicted cone surface is the wavefront of the Cherenkov light which moves with the muon through the water or ice. From the recorded arrival times of the light photons at the different PMTs one calculates the muon direction. (b) The PMTs are integrated into pressure spheres, called *optical modules*, together with the data acquisition and transfer electronics. Source: DESY.

energy:

$$\langle \theta_{\nu\mu} \rangle \approx \frac{1°}{\sqrt{E_\nu/\mathrm{TeV}}} \,. \tag{16.51}$$

This angle is of similar magnitude to the achievable angular resolution of the detectors. The direction resolution essentially determines the discovery potential for cosmic point sources.

Above the critical energy for muons, which is in ice $E_c^\mu = 1031\,\mathrm{GeV}$, the muon energy can be determined according to (3.95) on page 64 by the radiated energy which linearly increases with the muon energy (section 3.3.4, fig. 3.28).

At lower energies generally only a rough estimate for the energy is possible. If the energy lost before reaching the detector is not known there is no direct relation to the neutrino energy. A good energy determination is possible for tracks which begin and end in the detector (*contained events*) as well as for neutrino-induced cascades (fig. 16.35(b)).

16.6.5.3 Experiments

The proposal for the construction of neutrino telescopes for the detection of cosmic neutrinos employing the Cherenkov radiation in a lake or the ocean was made by Markov in 1960 [696]. A review of the various projects and the status of currently active experiments is given in [591]. The DUMAND project was the first attempt to realise a large-volume neutrino telescope in water (in the Pacific off the coast of Hawaii) according to the just described principle. This project could not master the

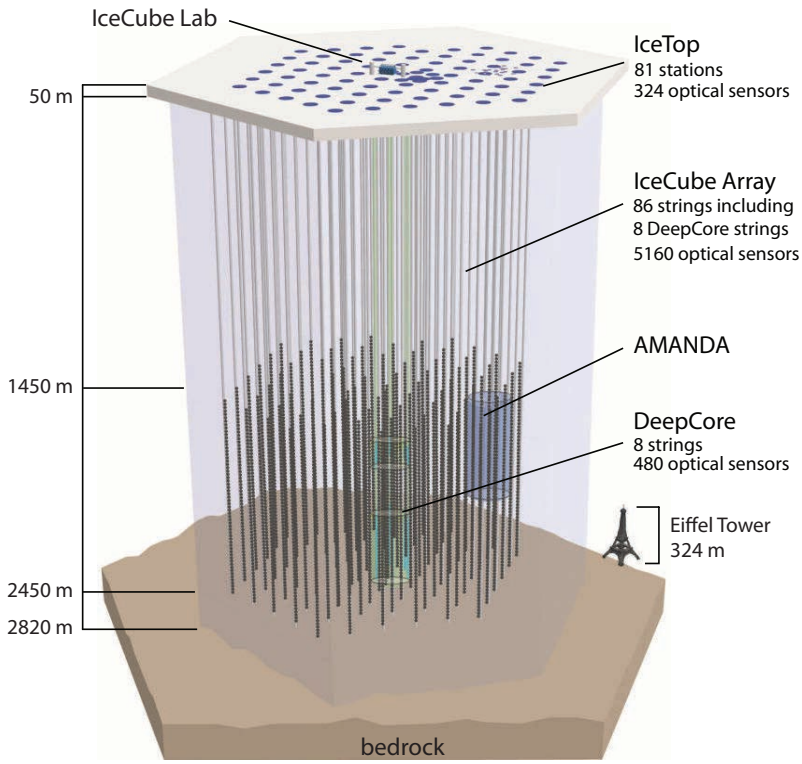
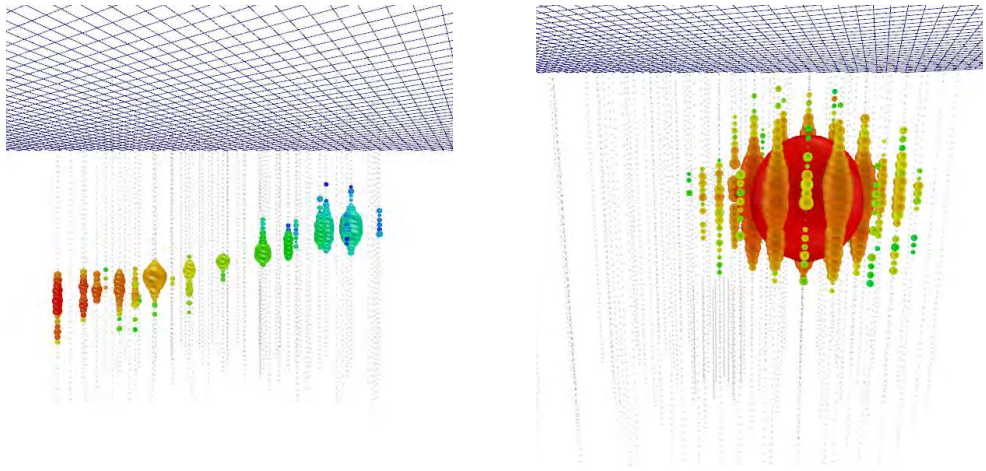


Fig. 16.34 The IceCube detector with its components DeepCore and IceTop. Explanations can be found in the text. Source: IceCube Collaboration/NSF.

problems arising mostly in connection to the salt water, and was officially terminated after about 20 years in 1996. In fact, the first functional telescope was then realised in fresh water. A high-energy neutrino telescope recorded data for the first time in 1993 in Lake Baikal.

A few years later the breakthrough was achieved with the AMANDA experiment, which was installed at a depth between 1500 m and 1900 m in the Antarctic glacier. The AMANDA telescope with a sensitive volume of 0.03 km^3 was the prototype for the IceCube telescope with a volume of about 1 km^3. It was clear from the beginning that only a detector of at least this size would have enough sensitivity to the expected cosmic neutrino fluxes, in particular of point sources. With the experience from AMANDA, IceCube was therefore realised at the same position at the South Pole starting in 2004 (fig. 16.34).

ANTARES, a neutrino telescope of similar size as AMANDA, started operation in 2007 in the Mediterranean Sea off the coast near Toulon. With the experiences gained with the DUMAND project and with additional research and development, the problems due to the salty water could be overcome. The success of ANTARES demonstrates that such a detector can be operated in the sea. However, the sea environment with bioluminescence, wave motion and aggressive salt water raises more problems than the installation and the operation of a detector in ice. With the neutrino telescopes ANTARES and Baikal, neutrinos filtered by the Earth from the southern hemisphere, and thus also from the galactic centre, can also be observed. However, the results of



(a) (b)

Fig. 16.35 Events in the IceCube detector. Shown are the light signals as registered by the optical modules. The size of the coloured areas corresponds to the amount of light and the colours correspond to the arrival time of the light (red stands for the earliest and blue for the latest time). (a) Track of an upward-going muon which was produced by a muon neutrino passing the Earth (time development from bottom to top). (b) A cascade, characteristic for an electron neutrino reaction in the ice (time development from the inside outwards). Source: IceCube Collaboration.

IceCube show that a detector size of at least $1\,\mathrm{km}^3$ is necessary to become sensitive to cosmic neutrinos. In order to obtain a similar coverage in both hemispheres a European initiative developed a respective detector in the northern hemisphere under the name KM3NET. In 2017 the first strings have been deployed and are operational.

In the following we describe as an example the largest currently running detector, the IceCube Neutrino Observatory.

IceCube. The IceCube Neutrino Observatory (fig. 16.34) [12, 16] at the geographic South Pole consists of a $1\,\mathrm{km}^3$ large detector located at a depth between 1450 and 2450 m in the ice, a $1\,\mathrm{km}^2$ large detector at the surface (IceTop, see section 16.4) and a $0.03\,\mathrm{km}^3$ large region with denser instrumentation (DeepCore). The main IceCube component consists of 5160 optical sensors on 86 m long strings for the recording of the Cherenkov light from charged particles which are produced by neutrinos in the ice or the Earth's crust (fig. 16.33). IceCube has primarily been designed for the measurement of neutrinos coming from the direction of the Earth, thus allowing one to observe the northern hemisphere of the sky. The Earth serves as a filter suppressing the overwhelming background of cosmic muons. However, meanwhile the southern hemisphere could as well be included into the observations which is mainly possible by requiring that the event must have started inside the detector volume. Typical event topologies are shown in fig. 16.35.

Each light sensor module, called *digital optical module* (DOM), contains a PMT (10 inch cathode diameter) for the registration of the Cherenkov light and autonomous electronics for recording, digitising and reading out of the signal as well as for triggering, calibrating and performing control tasks [15]. The most important task of the DOM electronics is the analog recording of the pulse shapes (usually referred to as

'waveforms'). The recording is started each time when the threshold, corresponding to 0.25 photoelectrons, is crossed by a photon signal.

Neighbouring DOMs on a string are connected via cables such that for the suppression of background of random cathode noise local coincidences of DOM signals can be required. Otherwise each DOM works autonomously and transmits its own data to the control room at the surface where events can be recognised by a global look at all DOM signals.

Events that pass filters for certain event classes are sent via satellite to the IceCube computing centre in Madison (USA). The complete data are written to storage media which can be transported to the North in the Antarctic summer. In addition fast online processes search for significant temporal enhancements in the rate of neutrino candidates and, if certain criteria are fulfilled, send alarms to other observatories and satellite experiments in order to initiate follow-up observations of various radiation types. This so-called *follow-up programme* is important since it increases the significance of observations of transient events (e.g. 'gamma-ray bursts', which last at most minutes, or the sudden *flare* of an active galaxy).

16.7 Detectors for detection of dark matter

16.7.1 Introduction

Cosmological and astrophysical observations suggest that about 27% of the Universe's density is made up of so-called dark matter (DM) which is subject to gravitational interactions but otherwise does not or only weakly couple to the visible matter. In particular this matter does not radiate, the reason why it is called 'dark'. Hints for DM come from different, independent observations, like the kinematics of stars and galaxies, the structure formation in the Universe, gravitational lensing and the temperature fluctuation of the microwave background radiation.

There are many more or less exotic candidates for particles that could make up the dark matter, for example 'sterile neutrinos', WIMPs (weakly interacting massive particles), axions or topological space-time defects. Searches for DM are performed by many different particle and astroparticle experiments which can be divided into three categories: searches for DM production by accelerator experiments, for DM annihilation by astroparticle experiments and for direct DM detection by specialised experiments. A compact review of indications, candidates and searches for DM is given, for example, by the article 'dark matter' in [765].

In the context of the subject of this book, WIMPs are most interesting because special detectors have been conceived for their direct detection. WIMPs with masses of up to several 100 GeV are most favoured in some common theories. In the following we concentrate on searches using detectors specialised on direct WIMP detection.

16.7.2 Detection of elastic WIMP scattering

The search for WIMPs is based on the assumption that they can elastically scatter off an atomic nucleus, thereby transferring recoil energy. The experiments are extremely difficult because of the small recoil energy and the expected small scattering rate.

16.7.2.1 Experimental conditions

Estimation of the WIMP flux. For the estimation of the recoil energy WIMPS are assumed to move in the centre-of-mass system of our Galaxy with thermal velocities

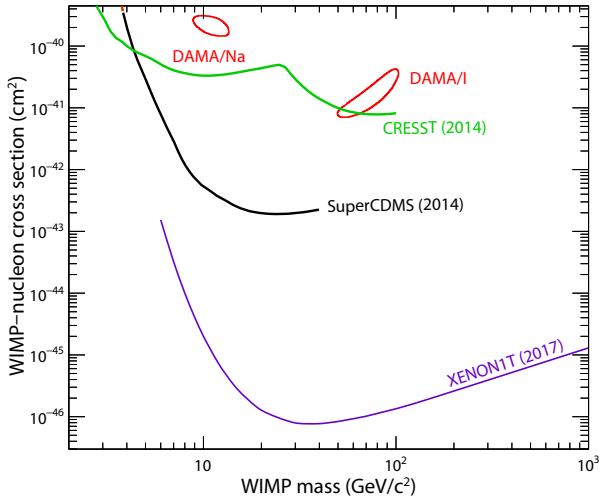


Fig. 16.36 Exclusion limits—respectively evidence in the case of DAMA—for the (spin-independent) elastic WIMP–nucleon cross section as a function of the WIMP mass (compiled by L. Rauch 2017, from the references [183, 101, 61, 113]) These selected examples show the currently best limits from some experiments mentioned in the text. Obviously such a plot can only be a snapshot taken at a given time. For a more complete picture the reader should consult most recent reviews and the corresponding article on the PDG web page [763].

which are small compared to the velocity of the solar system relative to the Galaxy (220 km/s). With this assumption the average WIMP velocity relative to an Earth-bound detector is 220 km/s. This velocity is modulated by the Earth's movement with daily and yearly periods. For WIMP masses between 10 GeV and 10 TeV the expected recoil energies of the nuclei are, for example, in germanium or xenon only 1–100 keV.

The elastic scattering rate is given by the local WIMP flux $j_\chi = n\,v$, where n is the number density and v the velocity, and the cross section $\sigma_{\chi,A}$ for the scattering of a WIMP χ off a target containing N_T nuclei with mass A:

$$R_{elast} = n\,v\,N_T\,\sigma_{\chi,A}\,. \tag{16.52}$$

From the above-mentioned indirect DM indications and cosmological considerations the DM density in our vicinity is estimated to be around $\rho_{DM} = 0.4\,\mathrm{GeV/cm^3}$ [765]. However, the community of direct DM searches commonly uses the value $\rho_{DM} = 0.3\,\mathrm{GeV/cm^3}$ (see e.g. [698]). The justification is that for the comparison of experimental results it is more important to stick to one value rather than to always adapt to the currently best theoretical value. With the number density

$$n = \frac{\rho_{DM}}{m_\chi} \tag{16.53}$$

one finds that the only unknowns in the expression for the rate R_{elast} in (16.52) are the WIMP mass and the elastic WIMP–nucleus cross section. Therefore one presents the exclusion limits obtained from different experiments as functions of these two parameters, as in fig. 16.36. In this figure the cross section is given per nucleon in order to be able to compare the results obtained with different targets. For the conversion one assumes a coherent sum of scattering amplitudes for all nucleons and the same amplitudes for protons and neutrons.

Background suppression. The largest difficulty that the WIMP experiments are facing is the expected extremely small WIMP scattering rate. At the time of writing

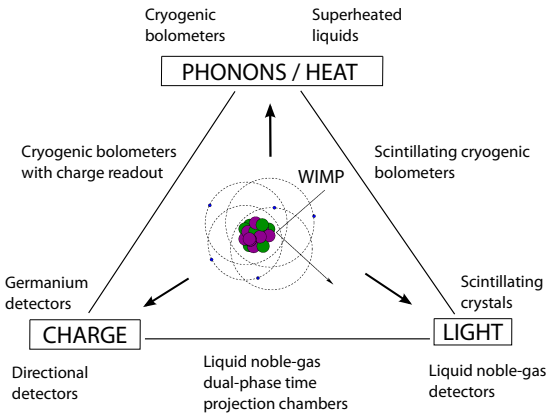


Fig. 16.37 Schematic representation of the different types of WIMP detectors and the respective signals they deliver when a WIMP is elastically scattered off a nucleus (from [698]). The signal origins—ionisation charge, scintillation light and phonons—are displayed at the three corners of the triangle along with the detector types which utilise these signals. The detector types which use two complementary signal types are listed along the sides of the triangles which connect the corners with the corresponding signals.

this text the limits range from about 1 event/kg/year to 1 event/ton/year depending on the WIMP mass. Without respective provisions, energy deposits by radioactivity are much more frequent. Therefore, a detector should be surrounded by materials which contain as little radioactivity as possible and it should additionally be shielded against neutrons, for example by polyethylene or water. Moreover, the experiments have to be operated in deep mines or tunnels in order to be protected against cosmic radiation. Underground laboratories which offer such conditions are listed in section 2.3.3. Experiments which we mention in the following are DAMA (DArk MAtter), CRESST (Cryogenic Rare Event Search with Superconducting Thermometers) and XENON (name stands for the target material xenon), all three of which are placed in the Gran Sasso underground laboratory. We also discuss below the experiments CDMS (*Cryogenic dark matter Search*) in the Soudan Mine in Minnesota and PICO at SNOLAB in a mine near Sudbury in Ontario.

Detector signals. If WIMPs indeed scatter off nuclei they can be detected through the recoil of the nucleus which it is scattered off. The energy transferred to the WIMP scattering will in turn be released to the target medium leading to the following types of recordable signals:

- The hit nucleus ionises the medium which is detectable by collecting the ionisation charge.
- Ionisation and excitation of the surrounding atoms can also lead to the generation of scintillation light detectable by photosensors.
- The nuclear recoil excites lattice vibrations in the crystal (phonons) which are detectable by sensitive temperature measurements.
- In superheated liquids the ionisation can act as condensation seed for bubble development. The formation of bubbles can be detected as acoustic or visual signals.

Figure 16.37 displays schematically different detector types and the WIMP signals they provide. As we will see in the following, detectors using a combination of two different signal types are particularly efficient in rejecting background.

Dual-signal readout. The WIMP detectors which are currently the most sensitive suppress the background from γ and β radiation very effectively by a 'dual-signal

readout' of the energy deposited in the detector. In general, the signals from lattice vibrations, ionisation charge and scintillation light are dependent on whether the energy has been deposited by elastic WIMP scattering or β and γ background. For example, the emission of scintillation light generally depends on the ionisation density, which is lower for β and γ background than for the recoil nucleus. However, the applied measurement methods cannot distinguish between WIMP scattering and scattering of other massive and/or neutral particles, including neutrons and neutrinos. Therefore the detectors have to be particularly well shielded against neutron background (see above).

16.7.3 Selected examples of direct WIMP search

16.7.3.1 Scintillator crystals at room temperature

A WIMP detection method which seems to be least difficult is the measurement of the energy loss of the recoil nucleus through ionisation and excitation, for example using scintillating crystals. However, for the level of WIMP rates reached until now, the signal does not stand out from the background. Therefore an additional observable for discrimination of background has to be found. Such an observable could be a time modulation of the WIMP signal following the above described periodic variations of the WIMP velocities. To this end one can search for annual variations of WIMP rates and/or signal sizes. Because of the low rates it would be too difficult to determine statistically significant daily variations.

Indeed, the experiment DAMA found such modulations with the expected phase by measuring over many years using NaI crystals [183]. However, the evidence from the DAMA experiment seems to be in contradiction to the exclusion limits from many other experiments, as demonstrated by the cross section versus mass plot in fig. 16.36. This discrepancy has strengthened the motivation for new experiments and upgrades of existing experiments with larger target masses to increase the relevant product mass $\times$ time. More details about the current status and future prospects of experimental results can be found, for example, in the DM review in [765] or in [698].

16.7.3.2 Dual-signal readout detectors

In the following we demonstrate for some examples how γ and β background for elastic WIMP scattering can be suppressed by comparing for each event the signals obtained by different measurement methods.

Cryo detectors. The excitation of lattice vibrations in a crystal by the recoiling nucleus leads to a small temperature increase, which in principle can be calorimetrically measured. For calorimetric measurements low temperatures are favourable because the temperature change per deposited energy is largest. The employed crystals, for example silicon, germanium, Al_2O_3 (sapphire) and $CaWO_4$, are cooled down to temperatures around 10 mK. At these low temperatures the heat capacity $C(T)$ of pure crystals is approximately proportional to T^3, which means that the temperature change is the larger the lower the temperature: $\Delta T = \Delta E/C(T) \propto 1/T^3$. With 1 kg target mass of the crystals listed above the temperature change per absorbed energy is of the order of $\Delta T \approx 10^{-7}$ K/keV.

For the measurement of such small temperature changes (fig. 16.38) the thermometers have to be very sensitive. Particularly favourable is the operation of a resistance

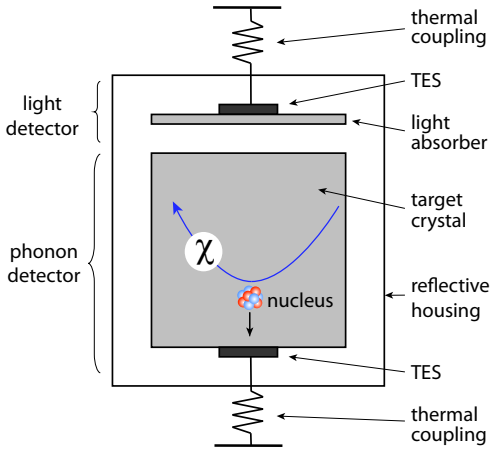


Fig. 16.38 Principle of the measurement of elastic WIMP scattering off a nucleus taking a module of the CRESST-II detector as example [100]. The recoil of the nucleus produces lattice vibrations (phonons) which are measured with a thermometer (TES, see text). At the same time the energy loss due to ionisation and excitation is converted into scintillation light which is detected in a light absorber, in this case as well by temperature measurement. The thermometers are coupled to a heat sink for temperature equalisation.

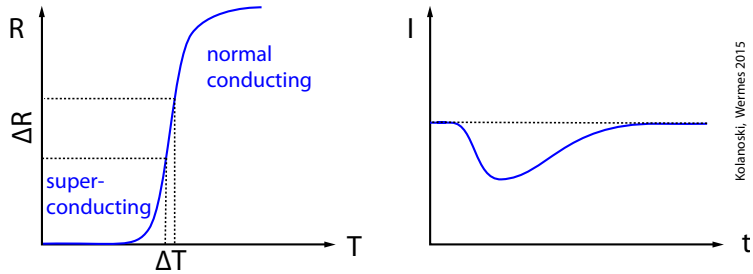
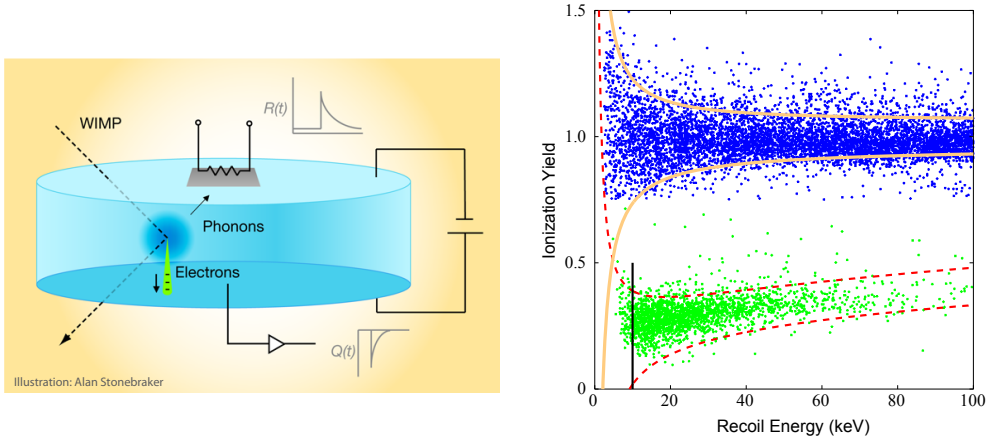


Fig. 16.39 Measurement of a temperature change at the transition from a superconducting to normal conducting state of a metal. The thermometer is a resistor through which a current flows. The strong resistance change at a small temperature change (left) generates a signal in the current (right). By coupling to a heat sink, as depicted in fig. 16.38, the temperature is equalised again and the signal falls off.

thermometer at the transition temperature to superconductivity (SPT = *superconducting phase-transition thermometer* or TES = *transition edge sensor*) [561]. The superconductors are designed as thin films. The operating principle is explained in fig. 16.39. The temperature increase ΔT leads to a resistance increase ΔR such that the current flowing through the resistor decreases. This current change is the signal which can be electronically processed.

The temperature increase per deposited energy in a crystal of given mass is inversely proportional to the sum of the heat capacities of the crystal and the sensor. This limits either the target mass or the energy threshold or both and hence limits the discovery potential. However, by systematically utilising the phonon dynamics one can achieve an effective heat capacity which is close to that of the thermometer and thus typically several orders of magnitude smaller than that of the complete system [561, 642]. The phonons are generated by the recoiling nucleus on a time-scale of nanoseconds with energies in the meV range (corresponding to a frequency range of THz) and thus much higher than the thermal energies of the crystal of $E_{therm} \approx 1\,\mu\text{eV}$. The meV phonons are reflected at the crystal surface or are absorbed by the thermometer film where it contacts the surface. In the superconducting thermometer the



(a) (b)

Fig. 16.40 Detector and measurement principle of the CDMS experiment. (a) Schematic of a single CDMS detector module (from [197], credit: APS/Alan Stonebraker). By the scattering of a WIMP off a germanium nucleus electron–hole pairs and phonons (lattice vibrations) are generated. Under the influence of a small electric field the charge carriers drift to the detector surface and are converted into an electronic signal $Q(t)$. Phonons which reach the surface can be absorbed in a superconducting aluminium layer leading there to a temperature increase. To the aluminium layer a 'transition-edge sensor' (TES) is thermally coupled which records the temperature change as a jump in the resistance $R(t)$ of a current circuit at the transition between normal conductivity and superconductivity (see fig. 16.39). (b) Calibration of the Ge detector with a source for neutrons and gamma radiation (^{252}Cf) (from [68], with kind permission of the American Physical Society). The measured ionisation charge is plotted versus the recoil energy as determined with the TES. Nuclear recoil events due to neutron scattering populate the lower band, the upper band is predominantly made up by Compton electrons generated by the gammas. The vertical line at 10 keV shows the threshold for the WIMP search which extends over the energy range up to 100 keV.

phonons quickly thermalise (within microseconds) due to a strong coupling to the electron system. This leads to a temperature increase on a time-scale of 10 µs with a relaxation time on a scale of 100 µs, given by the coupling to a heat sink. On a much longer time-scale of milliseconds the phonons also thermalise in the crystal, leading also there to a temperature increase which, however, is much smaller because of the higher heat capacity.

The dual measurement method provides a very efficient background discrimination and thus a substantial gain in sensitivity, for example by combining a temperature measurement with a measurement of the ionisation charge generated by the nuclear recoil in the crystal (fig. 16.38). The ionisation can be measured in different ways, for example by charge collection in a semiconductor detector (examples are the Ge or Si detectors of EDELWEISS and CDMS, fig. 16.40) or by measurement of light generated in a scintillating crystal (an example is the $CaWO_4$ detector of the CRESST experiment, fig. 16.38). For the very high ionisation densities (high dE/dx) of heavy recoiling nuclei the scintillation yield is suppressed due to the so-called quenching effect (Birks' formula in section 13.2.3) and also the efficiency of charge collection decreases because of the higher recombination rate. As a result the ratio of the WIMP signals

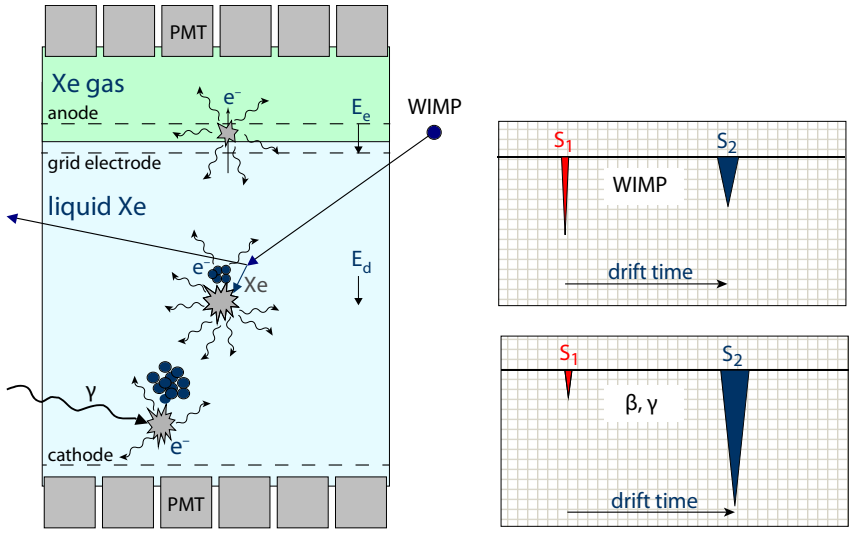


Fig. 16.41 Schematic representation of the WIMP detection method using a TPC filled with xenon in the liquid and gaseous phase (adapted from [735]). Three electrode grids provide the electric fields. While the drift field E_d in the liquid volume between the cathode and the 'grid electrode' is about 1 kV/cm, the field E_e between the grid electrode and the anode for extracting the electrons from the liquid to the gas phase has to be more than 10 kV/cm in order to obtain a high extraction efficiency. In the gas volume the latter field accelerates the electrons such that they can excite the xenon atoms. The scintillation light, which is recorded by PMTs at both ends of the TPC, is composed of the direct light due to the ionisation and excitation energy loss of the recoil nucleus (signal S_1), and the delayed scintillation light from the electrons accelerated between anode and grid (signal S_2). Signals from a WIMP are different from those of β, γ background leading to $(S_2/S_1)_{\mathrm{WIMP}} \ll (S_2/S_1)_{\beta,\gamma}$ (right plot). Further explanations can be found in the text.

from phonons to that from ionisation is smaller than the same ratio for signals of, for example, minimum-ionising particles. As can be seen in fig. 16.40(b) for a germanium detector, the background from β and γ radioactivity can be efficiently rejected. The background from elastic neutron scattering, however, has the same signature as that of the WIMPs and can only be suppressed by a well-designed neutron shielding. Eventually the remaining background has to be statistically subtracted.

WIMP detection in liquid noble gases. Cryo detectors which are sufficiently sensitive to temperature changes should not have too large heat capacities or, correspondingly, not too large masses, practically below about 1 kg. On the other hand the discovery potential grows with the mass. While cryo detectors therefore have to be split into many separated small modules, much larger masses of a single module are achievable with detectors based on liquid noble gases.[12] For WIMP detection the commonly used liquid gases are xenon and argon [115]. Using noble gases two measurement quantities for detection can again be utilised, namely ionisation charge and scintillation light, whose comparison allows for a suppression of β and γ radiation.

[12] The use of liquid noble gases for calorimetry is described in section 15.5.2 (see also table 15.6) and for scintillation detectors in section 13.3.4 (see also table 13.4).

As an example we take the XENON detector [112] in fig. 16.41 to explain the principle. A 'time projection chamber' (TPC, see section 7.10.10) is filled with liquid xenon in the drift space and with gaseous xenon above the drift space ('dual-phase TPC'). The electrons drift in the liquid towards the anode, pass a grid electrode just below the liquid–gas interface and are extracted from the liquid by a strong field between the anode and the grid (more than about 10 kV/cm). In the gas volume below the anode this field is sufficient to accelerate the electrons to energies above the threshold for excitation of xenon atoms with subsequent emission of scintillation light. If the energy gained by the electrons between collisions remains below the ionisation threshold and hence below the threshold for charge amplification, the scintillation light generated by a charge cloud is proportional to the charge ('proportional scintillation' or 'electroluminescence'). Compared to detecting the electron at the anode wire, proportional scintillation generally provides better energy and position resolutions than obtainable via gas amplification in an avalanche (see e.g. [232, 115]). The reason is that the emission of scintillation photons (of the order of 1000 per electron) is uncorrelated which is not the case for electron generation in an avalanche.[13] In principle the proportional scintillation can also be achieved in the liquid phase but because of the much shorter mean free paths between collisions with atoms the field strength necessary for reaching excitation energies is prohibitively high. Exploiting scintillation light also for the electrons, has the advantage that the signals can be recorded using a single technique. The scintillation light, generated both in the liquid and the gas volume, is recorded by photomultipliers (PMTs) which view the xenon volume from above and below.

The elastic scattering of a WIMP is simultaneously detected through the scintillation light and the ionisation charges, both generated by the recoiling nucleus. In the XENON experiment the ionisation charge is also measured via scintillation light, which in this case arises from the electron acceleration in the gas after the drift through the liquid. Hence one observes two light signals (S_1 and S_2) which follow each other with a separation given by the drift time of the electrons in the liquid, as shown in the right panel of fig. 16.41. In the figure it is also depicted that the signal height ratio of the drifting charge to that of the primary scintillation light is much smaller for the nuclear recoil by WIMPs than for the ionisation by β and γ radioactivity. Due to the high mass of the nucleus the ionisation charge is spatially more concentrated than for radioactivity and hence the recombination probability of the electrons with the ions is higher. The photons emitted by the recombination and de-excitation contribute to the (direct) scintillation light. Hence, because of the recombination the drifting charge is smaller and the scintillation yield larger. The background from β and γ radiation is mainly due to minimum-ionising particles, mostly Compton electrons, for which the recombination probability in the applied drift field is lower because of the lower ionisation density.

16.7.3.3 Bubble chambers for WIMP detection

A completely different approach to WIMP detection is the application of the bubble chamber principle (section 6.2) [167, 92, 698]. In a superheated liquid the energy deposited by the recoil of a nucleus induces a bubble nucleation. The bubbles can be visually recorded, usually taking camera pictures from different directions at typical

[13] Different methods for the registration of extracted electrons together with the correspondingly achievable resolutions are discussed, for example, in [232].

sample frequencies of about 100 frames/s such that the bubble formation can be reconstructed in time and three-dimensional space. The phase transition forming bubbles is an explosive process, which can also be acoustically detected employing piezoelectric acoustic transducers. The acoustic signals are often used for triggering the camera readout.

The bubble chamber principle is particularly well suited for WIMP detection since such a detector can be operated nearly without background. This is because the liquid-to-gas phase transition requires a local heat deposition above a certain threshold. The threshold can be adjusted by pressure and temperature settings such that only recoils caused by massive particles initiate bubble nucleation, thus rejecting background from β and γ radiation. Furthermore, an analysis of the frequency spectrum of the acoustic signal allows one to distinguish between neutron/WIMP nuclear recoils and alpha decays [130].

Another speciality of bubble detectors is that they are usually run with liquids containing fluorine (freons), such as C_4F_{10}, CF_3I or C_3F_8. With fluorine as target medium the physics scope of direct WIMP searches is broadened. Since the only stable fluorine isotope ^{19}F is a spin-$^1/_2$ nucleus, WIMP scattering becomes sensitive to spin-dependent forces (for more explanations see the DM article in [765]). In fact, the most sensitive spin-dependent cross section limits come from bubble detector experiments (in fig. 16.36 only spin-independent limits are plotted where bubble detectors are less competitive).

As an example we give here the parameters of the PICO-60 experiment located in the SNOLAB underground laboratory (see section 2.3.3) for a run employing a bubble chamber filled with 52 kg of C_3F_8 [92]. The event data readout is triggered by a fast comparison of consecutive pictures recognising the appearance of a gas bubble. After each trigger the chamber is compressed to above 10 bar within about 100 ms and then expanded to a pressure of about 2 bar at a temperature of about 14 °C, yielding a threshold for WIMP detection of about 3.3 keV. For the run described in [92] the efficiency for being in an expanded (superheated) state was about 77%, resulting in an efficiency-corrected DM exposure of 1167 kg-days. In this data sample no WIMP candidate was found.

A variant of the bubble chamber concept is bubble detectors where the superheated liquid is contained in small droplets which are suspended in a gel matrix (examples for experiments using this technology are PICASSO and SIMPLE, see references in [698]). This detection method has long been used for neutron dosimeters (see e.g. [557]). Since only the droplet where the recoil occurs undergoes a phase transition, a droplet detector can stay active for much longer periods than a conventional bubble chamber. However, with this technology larger sensitive volumes seem to be more difficult to reach. Therefore, and since the running stability of conventional bubble chambers has been continuously improved, the droplet technology has lost importance.

16.7.3.4 Status and future of WIMP searches

In fig. 16.36 sample exclusion limits for the spin-independent cross section of elastically scattered WIMPs are shown as a function of the WIMP mass (status 2017). At the time when the plot was made the best limits for WIMP masses above several GeV had come from liquid xenon detectors. The region down to 1 GeV or less is covered by crystal and semiconductor detectors employing cryo-technologies. The bubble chamber experiments are superior for limits on the spin-dependent cross sections (not shown).

For several experiments it is planned to increase the target masses into the regime of a few tons, which is best achievable with liquid noble gases as detector medium. Another path towards improving the limits is to lower the thresholds for the recoil energies (and thus also for the WIMP masses).

17
Signal processing, readout and noise

17.1 Electronic readout of detectors

Today electronic readout and processing of detector signals is the most common form of signal acquisition in particle physics. One of the few exceptions is photographic capturing and processing of reaction pictures, for example using photographic emulsions (see section 6.3). This technique is still used when data rates are low and spatial resolutions in the range of one micrometre or below are demanded. A particular choice of an electronic readout chain in an experiment depends on the type of detector and on the experiment's goals. While choices can be very different for different applications, some basic concepts exist which are followed for most detector systems. The diversity of the applied electronic concepts reaches deeply into the art and science of electronic circuits, exceeding by far the scope of this book. We refer to the corresponding literature on electronics and electronic systems, in particular Horowitz and Hill [540] (basics of electronics), Oppenheim and Willsky [748] (mathematical descriptions of electronic systems), Radeka [799] and Gatti [448] (electronic noise and readout techniques), as well as Gray-Meyer [472] or Laker-Sansen [639] (analog electronics design). A detailed discussion of concepts employed especially for semiconductor detectors can be found in Spieler [905], for wire chambers in [217] and for pixel detectors in [837]. In this chapter we restrict ourselves to some fundamental concepts that often appear in detectors and detector systems and have a certain general learning aspect. In particular we address the so-called front-end electronics which typically comprises signal amplification, shaping and discrimination as well as digitisation and signal transport on cables. Higher level processing like triggering and data acquisition are treated in chapter 18.

The tasks of the readout electronics in a detector system can be summarised as follows:

– electronic coupling to the detector to receive the detector signal (signal generation is described in chapter 5);

– amplification of the input signal and electronic processing (pulse shaping, discrimination);
– optimisation of the output signal depending on the application regarding
  - signal-to-noise ratio,
  - determination of the primary deposited energy in the detector (energy measurement),
  - determination of a precision mark of the signal arrival time (time measurement),
  - rate tolerance;
– digitisation and storage (buffering) of the output signal.

How these tasks can be solved in detail depends on the detector type and on the specific application requirements. In the large majority of applications the input signals end up in the front end as electric charges on which we therefore concentrate in this chapter. In magnitude and form the signal pulses differ much for different detector types. While for wire chambers (chapter 7) gas amplification inside the detector already provides an amplified signal by factors of 10^4–10^7 resulting in input charges of the order of pC, for semiconductor detectors (chapter 8) signal charges are typically very small: about 3 fC for highly energetic particles in silicon strip or pixel detectors and only 0.25 fC for example for the signal (in Si) of the 6 keV X-ray line of a ^{55}Fe radioactive source. In scintillation detectors (chapter 13) detection is indirect: first the light signal generated by deposited ionisation energy in the detector is transformed into an amplified (e.g. in a PMT) electronic signal which is then fed to the readout electronics. Often the PMT output signal is large enough that no further amplification is needed. At the other end of sensitivity sometimes single-photon signals must be detected, for example in Cherenkov detectors (chapter 11) by sensitive photodetectors (chapter 10).

Another quantity influencing the readout scheme (especially when charge is read out) is the detector capacitance or, more precisely, the total capacitance parallel to the preamplifier input including stray capacitance (see fig. 17.2), which is 'discharged' by the amplifier. Its size affects the noise performance of the readout (see section 17.10.3) and the design of the amplification and pulse shaping electronics needs to be tuned to a given input capacitance.

Detector readout systems, consisting of amplifiers and pulse filters, must be (a) causal, (b) time invariant, and (c) linear at least in the first amplification stage. A system is said to be causal if—at any time—it only depends on the value of its input at this time. It is time invariant if the ratio of output to input, for example $\text{signal}_{out}/\text{signal}_{in}$, is not time dependent. Linearity of the system means that the output pulse (e.g. $v_{out}(t)$) does not depend on the size of the input signal (e.g. $i_{in}(t)$), that is,

$$v_{out}\left(\alpha \times i_{in}(t)\right) = \alpha \times v_{out}\left(i_{in}(t)\right) \,. \tag{17.1}$$

Linearity is an important property for the filtering stage (pulse shaping stage, see section 17.3) following the first stage (amplification) of the readout chain (cf. fig. 17.2), which only then can be realised by a certain electronic circuit if the shape of the input to this filter stage does not change with its amplitude. If this were not guaranteed, any interpolation between output pulses of neighbouring electrodes (often employed for a more precise reconstruction of the position of a particle entering the detector) would require a pulse height dependent correction.

Signal pulse. We use fig. 17.1, which displays an individual example pulse (e.g. the output of a detector or a photomultiplier upon a detected photon or particle) to

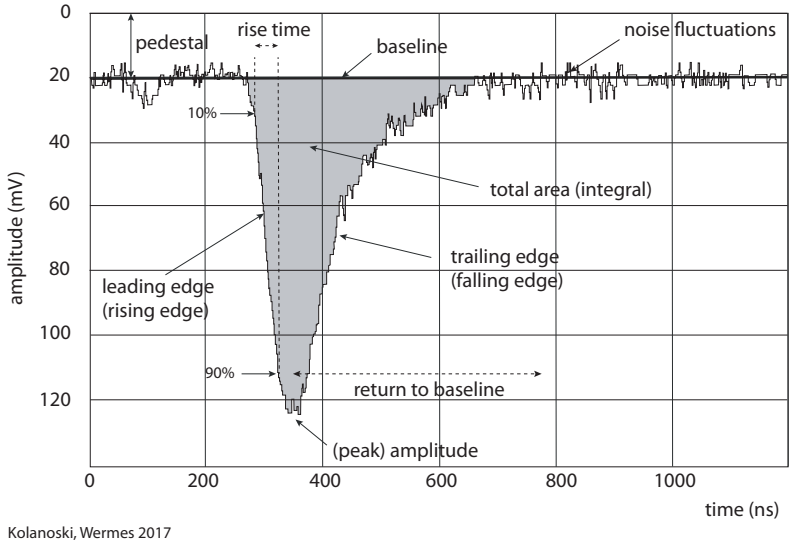


Fig. 17.1 Example of a detector pulse used to define characterising quantities (see text, adapted following [355]).

introduce some quantities for pulse characterisation. Such a pulse could for example be a drift chamber signal or the output of a photomultiplier on an oscilloscope. It could also be an already preamplified analog pulse from a semiconductor detector with subsequent pulse shaping electronics.

Assuming that the signal generation in the detector is short compared to the electronic processing times, which is a typical but not an always given case, an electronic pulse is characterised by the following quantities:

- The *peak amplitude* is the maximum pulse amplitude, also called *pulse height*. In linear systems its value is proportional to the primary deposited energy in the detector.
- The *peaking time* is the position of the peak amplitude in time.
- The *area* or *pulse integral* is the integral over the pulse shape. In linear systems and for short, δ-like signals, its value should be proportional to the primary deposited energy.
- The *pulse width* is the width of the pulse, usually defined as the *full width at half maximum* (FWHM).
- *Leading* and *trailing edges* are the rising and falling slopes of the pulse.
- The *rise time* characterises the steepness of the pulse's rise. It is commonly defined as the time needed for the pulse to rise from 10% to 90% of its peak amplitude, but other definitions also exist.
- The *slew rate* indicates the voltage change per time dV/dt in units V/s.
- The *baseline* or *pedestal* value is the output value upon zero input. It defines the 'zero' level from which to measure the signal height. Although the baseline usually has a fixed value, deviations can occur for some (short) time. They are called *baseline shifts*.

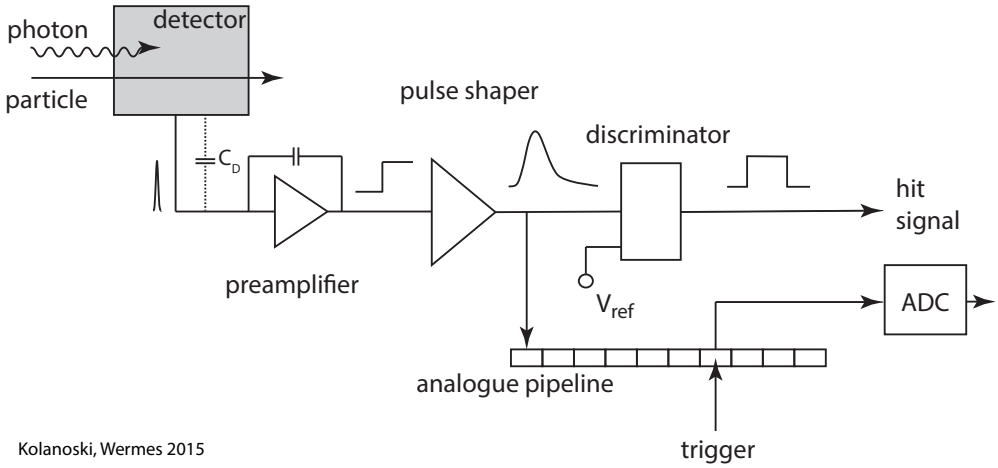


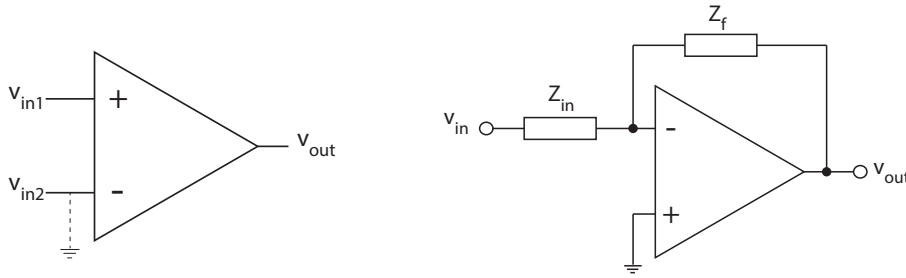
Fig. 17.2 A typical front-end readout scheme, often used (in parts) for the readout of a detector including amplification, pulse shaping, discrimination as well as signal buffering (here analog) and digitisation (represented here by an ADC).

- The *return-to-baseline* time is the time needed for the amplitude to return to the baseline value.
- An *undershoot* one calls the part of a pulse amplitude having opposite sign (with respect to the baseline) compared to the main amplitude.
- A *unipolar signal* is a pulse form where, except for noise fluctuations, the value of the amplitude is above or below the baseline at all times t. Usually included in this definition are signals with small undershoot amplitudes.
- A *bipolar signal* is a pulse form where the part of the pulse occurring later in time has a sign opposite to the part earlier in time (see also fig. 17.13).

Readout scheme. A general front-end readout scheme, which is often used in detector applications and which can thus serve as a base concept for this chapter, is shown in fig. 17.2. The detector reacts to an impinging particle, in general by delivering a short current signal to the input of the so-called preamplifier. This preamplifier usually is the first (and often only) amplification stage of a readout chain. If a capacitive feedback loop is realised the amplifier acts as an integrator of the input current. At its output it generates a voltage step proportional to the signal charge.[1]

In order to generate a pulse that returns to the baseline after a finite time, the voltage step is filtered by a *shaper*. Filtering means a bandwidth limitation of the signal, which also reduces the electronic noise (see section 17.10.3.3). The output signal is proportional to the energy deposited in the detector (i.e. proportional to the generated charge) due to (17.1). This is true for the integral over the signal pulse as well as for the pulse height provided that the pulse shape is unchanged, which usually is the case under the assumptions made. In the *discriminator* the signal voltage is compared to a reference voltage V_{ref}; whenever the signal voltage exceeds V_{ref} a digital signal of a given width is generated. Often a trigger, that is, a decision whether to keep or reject the signal information, is employed for further processing (see chapter 18). The trigger pulse either involves the readout detector or can be formed by other

[1] The necessary discharge of the feedback capacitor is treated in section 17.2.3 on page 726 ff.



(a) OpAmp wiring symbol. (b) OpAmp with feedback.

Fig. 17.3 Operational amplifier: (a) wiring symbol; (b) inverting OpAmp with external feedback.

(sub)detectors of the experiment and usually is available only after some time (trigger delay, typically some μs). In this case the signal must be delayed and stored, for example in sequential buffers (*pipelines*), as indicated in fig. 17.2, or—in less complex detector systems—simply by means of delay cables. Depending on the requirements and on further processing, the full analog signal or only a hit signal must be buffered. Upon arrival of the trigger an analog signal is digitised in an *analog-to-digital* converter (ADC), or its arrival time in a *time-to-digital* converter (TDC), and then is passed on to the *back-end* electronics which collects the information from several detectors for final storage.

17.2 Signal amplification

17.2.1 Amplifiers

Detector signals can be very small, in the nanoampere, femtocoulomb or microvolt range, and must be (pre)amplified before the data can be read out for further processing. Preamplifiers as in fig. 17.2 are usually operational amplifiers to the extent that the external circuit configuration specifies the operation characteristics: voltage or current amplifier, charge-sensitive amplifier, etc.

Operational amplifiers (OpAmps) are amplifiers with high internal gain a_0 whose behaviour is determined to first order by external circuit elements, in particular by impedances fed back to the input (fig. 17.3(b), see e.g. [540]). For this reason the amplifier itself can be treated as a generic circuit element represented by a triangle. In the ideal case the (internal) 'open loop gain' a_0 of the OpAmp is infinite. In practice, a_0 is smaller than ∞ (typically $\gtrsim 10^5$) and frequency dependent (see fig. 17.8 on page 726).

An OpAmp has two voltage inputs, whose difference is amplified by a_0, and one output voltage, all taken with respect to some reference potential, for example 'ground' (fig. 17.3(a)). The functional principle of an OpAmp can be described in terms of two 'golden rules' which already lead quite far in understanding [540]: (1) the output attempts to do whatever is necessary to make the voltage difference of the inputs vanish,[2] and (2) the inputs draw (almost) no current. If one input is grounded the other is said to be on *virtual ground* due to golden rule 1 (see fig. 17.3). Using field

[2]The input voltage difference is not actually changed to zero, but rather the output voltage is swung up such that the feedback counteracts the voltage difference at the input.

effect input transistors rule number 2 is approximated at the pA level; with bipolar transistors input currents of some 100 nA are possible.

With these rules the OpAmp's operation characteristic can be easily derived for a given feedback. For example, for the inverting amplifier shown in fig. 17.3(b) the (closed loop) gain is entirely determined by the external impedances Z_{in} and Z_f:

$$\frac{v_{out}}{v_{in}} = -\frac{Z_f}{Z_{in}}\,, \tag{17.2}$$

where $v_{in} = v_- - v_+$. If the feedback is realised by a capacitor C_f the output voltage v_{out} is proportional to the integral of the input current (i.e. the charge), which must completely flow through C_f since no current flows into the input (golden rule 2). This principle is used for the charge-sensitive amplifier (CSA) described in section 17.2.3.

Front-end electronics in modern readout systems is usually realised by employing integrated circuit designs including one or two amplification stages (called preamplifier and amplifier) which typically contain only few transistors realising the OpAmp functionality. Figure 17.4 shows two types of simple amplifier realisations which are found as first (and often only) stage in most amplifier circuits used in detector readout chips: (a) a CMOS inverter stage[3] (used for example in the CMS pixel readout chip [589]), (b)–(d) a simple transistor amplifier with common variants that improve its performance properties.

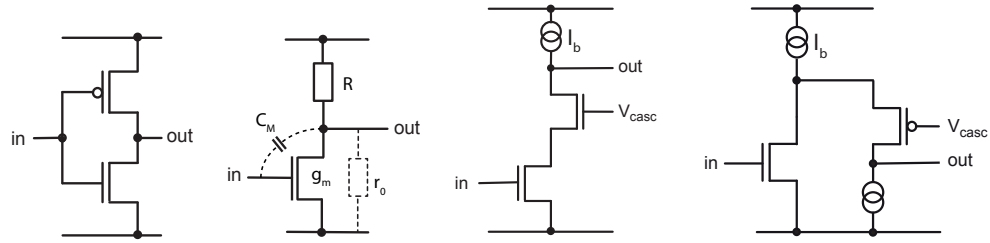
The preamplifier should feature low noise as well as short signal response and processing times (typical rise times are between nanoseconds and tens of nanoseconds). Low noise and fast signal response are, however, in conflict with each other (for a given power budget; see also section 17.10) and cannot be both optimised arbitrarily. A compromise between noise and response time properties must always be found, optimised for a given application. A comprehensive discussion of this topic, can be found for example in [905] with a focus on semiconductor detectors and in [217] with a focus on wire (drift) chambers.

Because the *signal-to-noise ratio* (SNR) deteriorates with the length of the subsequent connection, the preamplifier is usually implemented on or very near the detector. Hence not only the cable and strip line lengths between detector and preamplifier should be minimal, but also special care is needed to avoid the already small detector signals receiving distortions from noise or electromagnetic interference (EMI or pick-up noise). Behind the preamplifier the output signals are already amplified such that they are less sensitive to subsequent noise sources. They can more easily be transported via cables for further processing.

Usually the input signals from the detector to the preamplifier are transformed into voltage (V) or current (I) signals. Depending on the types of input and output signal we distinguish:

- voltage amplifier: $V \to V$,
- current amplifier: $I \to I$,
- transconductance amplifier: $V \to I$,
- transimpedance amplifier: $I \to V$,
- charge amplifier: $Q \to V$ (or I).

[3] While the name 'inverter' refers to a digital signal operation, the same transistor arrangement can be an amplifier stage.



(a) CMOS inverter. (b) Simple transistor amplifier. (c) Cascoded transistor amplifier. (d) Folded-cascode stage.

Fig. 17.4 Simple realisations of amplifier types (without external feedback). (a) The CMOS inverter (see footnote 3 on page 720), realised by a PMOS on top of an NMOS transistor between the supply lines, delivers an input-dependent, inverted output signal which can also be used for an analog functionality, where the steep (step-function like) output characteristics provides the gain dv_{out}/dv_{in}; (b) single transistor amplifier delivering an amplified inverted voltage signal at the output. The transistor's open loop gain is given by $|a_0| = g_m R||r_0$, where r_0 is the dynamic output resistance acting here parallel to R and $g_m = di_D/dv_{gs}$, $i_D =$ drain current, $v_{gs} =$ gate-source voltage, is the transconductance. Via the parasitic (Miller) capacitance $C_M = (a_0 + 1)C_{gd}$, (dashed, $C_{gd} =$ gate-drain capacitance) the output couples to the input in an undesired way. This problem is removed in (c), in the so-called 'cascoded' variant of (b). The output is moved up above the drain of the *cascode* transistor which is held at constant gain potential V_{casc}. Hence also the potential at the position of the former output point in (b) is now held constant at $V_{\text{casc}} - V_{GS}$, where V_{GS} is the (constant) gate-source voltage of the cascode transistor, while the amplified voltage signal swings at the new output (the current source I_b can be thought of acting like a resistor as in (b)). As a result the Miller effect is suppressed and gain and bandwidth of the stage as well as the output resistance dv_{out}/di_{out} are increased. In (d), finally, the *folded cascode* is shown, where amplifier and cascode transistor have separated bias currents. Cascoded transistor stages are often used as input stages for (pre)amplification of detector signals. The transistor potentials are set between an 'upper' and a 'lower' supply rail, indicated by the horizontal lines.

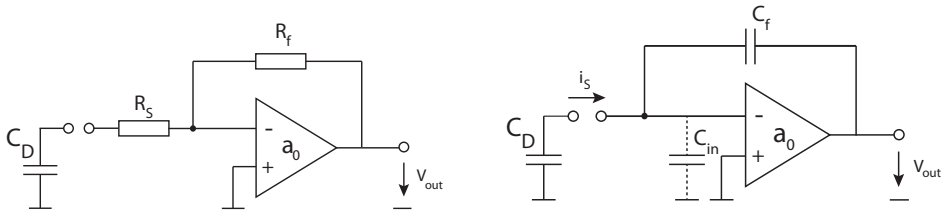
17.2.2 Current- and voltage-sensitive amplifiers

A voltage amplifier amplifies a voltage signal at its input. It has a high input and a low output impedance like a voltage source. In general it is realised by an operational amplifier with resistive feedback (fig. 17.5(a)). The detector, represented as a capacitance to be discharged, delivers the signal current i_S through the resistor R_S to (virtual) ground for a time Δt.

If the discharge time is large compared to the signal duration time ($\tau = R_S C_D \gg \Delta t$), then, in a certain sense, the detector integrates the signal current on the detector capacitance ($V_D = Q_S/C_D$), and at the amplifier input we have the voltage[4] $v_{in}(t) = V_D \exp(-t/R_S C_D)$. The output voltage is proportional to v_{in} according to (17.2) and the system operates as a *voltage amplifier*:

$$v_{out}(t) = -\frac{R_f}{R_S}\, v_{in}(t) = -\frac{R_f V_D}{R_S} \exp(-t/R_S C_D)\,. \tag{17.3}$$

[4]We denote as in chapter 5 operation voltages and currents by V, I and signals by $v(t)$ and $i(t)$, respectively.



(a) Resistive feedback. (b) Capacitive feedback.

Fig. 17.5 Typical operational amplifier circuits for detector readout. The detectors capacitance C_D is discharged by the amplifier. (a) resistive feedback: voltage- or current-sensitive amplifier (see text), (b) capacitive feedback: charge-sensitive amplifier. C_{in} is the dynamic input capacitance (see text).

Voltage-sensitive preamplifiers are often used when voltage input signals are sufficiently large and fast rise times are aimed for. In some detectors, for example in semiconductor detectors, the input voltages are, however, small, typically $v_{in} = Q/C_D \approx 3\,\text{fC}/10\,\text{pF} = 0.3\,\text{mV}$. For such detectors a voltage amplifier is usually less suitable as it generally yields a comparatively bad SNR and develops some sensitivity to pick-up signals and cross talk. Depending on the output resistance a voltage amplifier can have a (low ohmic) voltage output ($V \to V$ configuration) or a (high ohmic) current output $V \to I$ configuration. In the latter case the amplification di/dv then is a *transconductance* in units A/V and the amplifier is called *transconductance amplifier*.

If, however, the discharge time of the detector capacitance is small compared to the signal duration time ($\tau = R_S C_D \ll \Delta t$), this means the detector immediately delivers the signal current to the amplifier, we have:

$$v_{out}(t) = -\frac{R_f}{R_S} v_{in}(t) = -R_f i_S(t)\,, \tag{17.4}$$

and the systems operates as a *current amplifier*. In $I \to I$ configuration the amplifier has a small input impedance and a large output impedance.

If the system is configured with a small output resistance (e.g. by using a 'source follower' [540] at the output), then the amplifier operates in $I \to V$ configuration as *transimpedance amplifier* (or *transresistance amplifier*). The amplification dv/di is called *transimpedance* or *transresistance* in units V/A.

Transimpedance amplifiers are often found in fast applications, for example if one is interested in the time-dependent shape of the signal pulse, which is influenced by the electric field and the space charge situation in a detector (*transient current technique* (TCT), see also section 8.4).

17.2.3 Charge-sensitive amplifier

The charge amplifier ($Q \to V$) or *charge-sensitive amplifier* (CSA) shown in fig. 17.5(b) is often used for detector readout, in particular when the signal-to-noise ratio for a detector is small. This amplifier type can, for example, be realised by an (inverting) OpAmp with capacitive feedback (integrator). The signal current is integrated on the feedback capacitor C_f:

$$v_{out}(t) = -a_0\, v_{in}(t) = -\frac{1}{C_f}\int_0^t i_S\, dt' = -\frac{Q_S(t)}{C_f}\,. \tag{17.5}$$

Over the feedback capacitor we have the voltage difference

$$v_f = v_{in} - v_{out} = v_{in}\,(a_0 + 1) = \frac{Q_f}{C_f}\,. \tag{17.6}$$

The charge Q_f on C_f is equal to the signal charge Q_S since no current flows into the OpAmp (golden rule 2, page 719): $Q_S = Q_f = C_f(a_0+1)\,v_{in}$. The amplifier hence has a 'dynamic' input capacitance

$$C_{in} = \frac{Q_S}{v_{in}} = C_f\,(a_0 + 1)\,. \tag{17.7}$$

For large internal (open loop) gain a_0, we usually have $C_{in} \gg C_D$ and the charge amplification $A_Q = v_{out}/Q_S$ only depends on the capacitance C_f of the feedback capacitor and not on the detector capacitance C_D (see below):

$$A_Q = \left|\frac{v_{out}}{Q_S}\right| = \frac{a_0 v_{in}}{(C_D + C_{in})\,v_{in}} \approx \frac{a_0}{C_{in}} = \frac{a_0}{a_0+1}\,\frac{1}{C_f} \approx \frac{1}{C_f}\,. \tag{17.8}$$

The smaller the feedback capacitance C_f the larger is A_Q.

Residual charge. The charge Q_S created in the detector is distributed on the existing capacitances, that is, on the detector capacitance C_D, here representing the sum of all capacitances at the input—also stray capacitances—, and on the feedback capacitance C_f:

$$Q_S = v_{in}\,(C_D + C_{in}) = Q_D + Q_f = C_D\,v_{in} + C_f\,(v_{in} - v_{out})\,, \tag{17.9}$$

where (17.6) has been used. Since the dynamic input capacitance C_{in} in (17.7) is not infinitely large, a charge residual remains on the detector capacitance:

$$Q_{res} = v_{in}\,C_D = Q_S\,\frac{C_D}{C_D + C_{in}}\,. \tag{17.10}$$

For example, for $a_0 = 1000$ and $C_f = 1\,\mathrm{pF}$ the fraction of the signal charge appearing at the amplifier is:

$$Q_{in} = Q_S - Q_{res} = \begin{cases} 99\,\%\;Q_S & \text{at} \quad C_D = 10\,\mathrm{pF}\,, \\ 67\,\%\;Q_S & \text{at} \quad C_D = 500\,\mathrm{pF}\,. \end{cases}$$

A (large) residual charge Q_{res} can cause unwanted effects in between segmented electrodes (pixel or strips), for example capacitive cross talk. In the ideal case there is no residual charge Q_{res}. This is effectively achieved if the dynamic input capacitance is large compared to the detector capacitance: $C_{in} = (a_0 + 1)\,C_f \gg C_D$.

Cross talk. One speaks about *cross talk* between electrodes if an input signal, apart from generating an output signal at an electrode, also causes output signals on neighbouring electrodes. This situation is sketched in fig. 17.6. The cross talk charge is given by

$$Q_{cross} = v_{in}\,C_{inter} = Q_{res}\,\frac{C_{inter}}{C_D} = Q_S\,\frac{C_{inter}}{C_D + C_{in}}\,, \tag{17.11}$$

where, as defined above, C_D includes all (stray) capacitances at the input and hence also C_{inter}, the capacitance between electrodes; v_{in} is the voltage signal at the amplifier input according to (17.10). Minimal cross talk is hence achieved when capacitances

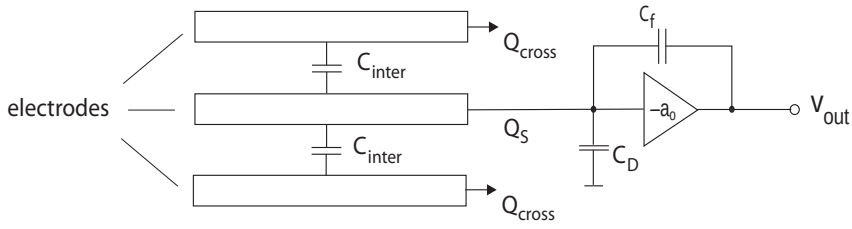


Fig. 17.6 (Strong) capacitive coupling between electrodes causes cross talk at the output.

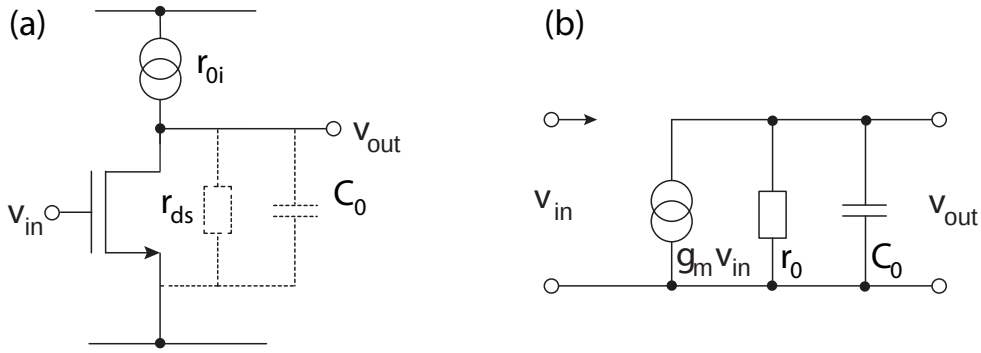


Fig. 17.7 Single-stage amplifier (a) and small signal equivalent circuit (b). The current is represented by $i_{in} = g_m v_{in}$, g_m = transconductance. The resistor r_0 consists of resistor r_{0i} and *drain-source* resistor r_{ds}, acting in parallel (with respect to the current supply). The capacitance of the amplifier stage is denoted C_0. The voltage signal v_{in} at the input (*gate*) is DC-wise not connected to the output, indicated by the arrow symbol in (b).

between electrodes are small, a condition that cannot always be realised, or by maximising the denominator of (17.10), which means in particular $C_{in} \approx a_0 C_f$, since enlarging the detector capacitance C_D would also increase the noise (see section 17.10.3) and influence the timing properties among other effects. Therefore a small residual charge and minimal cross talk demand large open loop gain a_0 and large C_f, in conflict with the desire for a large charge gain A_Q of a stage, which requires a small C_f in the feedback loop. Depending on the application a compromise must be found.

Frequency behaviour. Only an ideal integrator delivers a step function at the output which corresponds to an infinitely large bandwidth, which means that all frequencies are amplified equally. Real amplifiers have limited bandwidths, resulting in finite output rise times. In fig. 17.7 a single-stage amplifier is shown together with its equivalent circuit diagram. Its time response can be described by a single time constant τ. The response in the frequency domain is shown in fig. 17.8(a) and in the time domain in fig. 17.8(b). In a double-logarithmic representation (Bode diagram [221]) the frequency dependence of the amplification, the *transfer function*,[5] can approximately be represented as pieces of straight lines. Using $\omega = 2\pi f$, $f =$ frequency,[6] the dependence is given as

[5] The transfer function specifies the output of a device as a function of its input.

[6] In electronics the frequency is usually denoted by the letter f instead of ν, usually used in physics.

$$a(\omega) = -\frac{v_{out}}{v_{in}} = -\frac{i_{out}}{v_{in}}\,(r_0 \| C_0) = -g_m\,\frac{r_0}{1 + i\omega r_0 C_0} = \frac{a_0}{1 + i\dfrac{\omega}{\omega_c}}\,, \tag{17.12}$$

where r_0 combines the two resistors r_{ds} (*drain-source* resistor of the transistor) and r_{0i} (resistor of the current source) (see fig. 17.7(a)), while $g_m = i_{out}/v_{in}$ is the transconductance of the amplification transistor; $r_0 \| C_0$ denotes a parallel configuration of the corresponding impedances. The last equal sign uses the fact that in the DC-limit ($\omega \to 0$) the open loop gain reads

$$a_0 = -\frac{v_{out}}{v_{in}} = -\frac{i_{out}}{v_{in}}\,r_0 = -g_m\,r_0 \tag{17.13}$$

and introduces the *cut-off frequency*

$$\omega_c = \frac{1}{r_0 C_0} \tag{17.14}$$

corresponding to a time constant $\tau_0 = r_0 C_0$ which describes the rising edge of the output pulse in the time domain:

$$v_{out} = v_0\left(1 - \mathrm{e}^{-t/\tau_0}\right)\,. \tag{17.15}$$

The gain is commonly expressed in decibels (dB) defined as

$$a_{dB} = 20\log_{10}|a_v| \qquad \text{for amplitudes}\,, \tag{17.16}$$

$$a_{dB} = 10\log_{10}(P_{out}/P_{in}) \qquad \text{for power}\,. \tag{17.17}$$

The amplification is about constant (see fig. 17.8) up to the cut-off frequency ω_c, where $|a(\omega_c)| = \frac{1}{\sqrt{2}}\,a_0$ (-3 dB), and rolls off thereafter inversely proportional to ω (by 20 dB per decade[7]) until amplification ceases completely ($a(\omega) = 1$). The amplifier acts as a low-pass filter. At low frequencies r_0 dominates the frequency behaviour and the phase shift of the (inverting) amplifier is close to 180° as can be seen from (17.12). At high frequencies the capacitance C_0 dominates, introducing a −90° phase shift, such that the phase difference between input and output becomes +90°.
The *gain–bandwidth product* (GBP)

$$\mathrm{GBP} = |a(\omega) \times \omega| \approx g_m r_0\,\frac{1}{r_0 C_0} = \frac{g_m}{C_0} \tag{17.18}$$

is constant[8] in the region of the falling slope and equal to its (ideal) value at the cut-off corner ($= |a(\omega_c) \times \omega_c|\,\sqrt{2} = a_0\,\omega_c$, solid corner point in fig. 17.8(a)).

The (complex) input impedance of a charge-sensitive amplifier (fig. 17.5(b)) with the feedback capacitance C_f is given by:

$$Z_{in}(f) = \frac{1}{i\omega C_{in}} = \frac{1}{(a(\omega) + 1)\;i\omega C_f} \approx \frac{1}{ia(\omega)\,\omega C_f}\,, \tag{17.19}$$

[7] 20 dB is a gain change by a factor of 10; −3 dB is a decrease by a factor $\sqrt{2}$.

[8] Strictly constant GBP applies to amplifiers with a 'single dominant pole', that is, a single falling slope of $a(\omega)$.

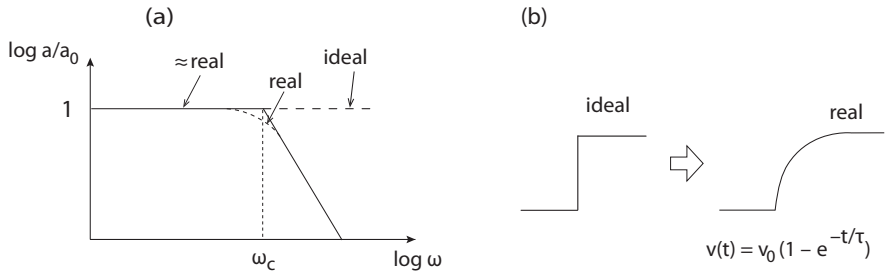


Fig. 17.8 Frequency (a) and time (b) response of a single-stage amplifier. In (a) the ideal (dashed line), an almost real (solid line) and the real (dotted line) frequency response is shown (Bode diagram). In the time domain (b) the ideal step function response turns into a real step function signal with finite rise time.

where $C_{in} = C_f\,(a(\omega)+1)$ according to (17.7) has been used with $a_0 \to a(\omega)$ and on the right-hand side $|a(\omega)| \gg 1$ has been assumed. For the transfer function $a(\omega)$ we find at low frequencies $a(\omega) \to a_0$ and at high frequencies that $a(\omega) \to -ia_0\frac{\omega_c}{\omega}$ according to (17.12). Hence at low frequencies the input impedance is purely capacitive:

$$Z_{in}(\omega) \to \frac{1}{ia_0\omega C_f}\,, \tag{17.20}$$

whereas for high frequencies—as long as the assumption $|a(\omega)| \gg 1$ is still valid—the amplifier behaves like a resistor:

$$Z_{in}(\omega) \to \frac{1}{\left(-ia_0\frac{\omega_c}{\omega}\right)\, i\omega C_f} = \frac{1}{a_0\omega_c C_f} = \frac{1}{\mathrm{GBP}\cdot C_f} = \frac{C_0}{C_f}\,\frac{1}{g_m} = R_{in}\,, \tag{17.21}$$

since the transconductance $g_m = \partial i/\partial v$ is an inverse resistance.

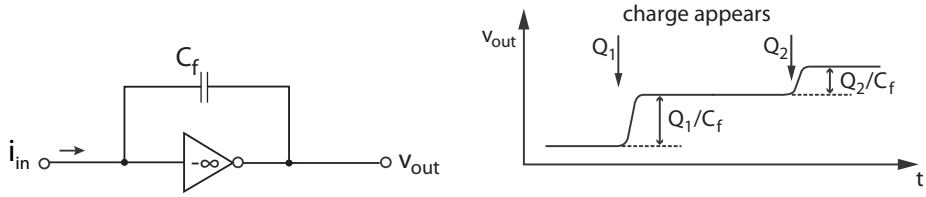
The time constant of a charge-sensitive amplifier arises from the product of the input capacitance ($\approx$ detector capacitance C_D) and the input resistance R_{in} with (17.18):

$$\tau_{CSA} = C_D\,R_{in} = \frac{C_D}{C_f}\,\frac{1}{a_0\omega_c} = \frac{C_D}{C_f}\,\frac{C_0}{g_m}\,. \tag{17.22}$$

In practice amplifiers with several stages are often used, resulting in a description with more than one slope in the Bode plot and several cut-off frequencies. However, most often one stage dominates the bandwidth and is responsible for the amplifier's frequency dependence and noise behaviour (see section 17.10.3).

Discharging C_f. In the ideal case the charge seen at the input of the preamplifier is integrated on C_f and the corresponding voltage value v_{out} is seen at the output. For every newly arriving signal at the preamplifier input additional charge is added. The output voltage jumps accordingly by a value Δv_{out}. For subsequent pulses corresponding voltage steps appear at the output, as shown in fig. 17.9(b) and sketched in fig. 17.11. At a certain amount of charge the limit of the dynamic range of the CSA is reached and the amplifier output saturates. The feedback capacitor, therefore, must be discharged in order that the system remains operative. Several *reset* techniques are in use:

– discharging by a (transistor) switch (fig. 17.10(a)),



(a) Circuit diagram. (b) Output voltage.

Fig. 17.9 Response of an ideal charge amplifier (without *reset*) on two successive input signals.

– discharging with a resistor (fig. 17.10(c)),
– discharging by a constant current source (fig. 17.10(e)).

Circuit and output signal for discharging by means of a switch are shown in figs. 17.10(a) and (b), respectively. By closing the switch the capacitor is discharged and the preamplifier is reset, ready for new signals. This method becomes problematic for high signal rates when high reset frequencies generate sharp transient pulses (*switching noise*).

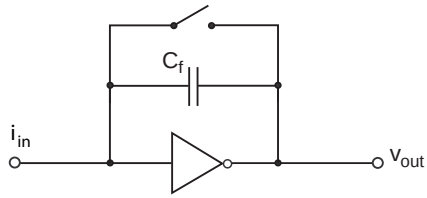
No switching is needed when the feedback capacitor C_f is discharged by a resistor, as shown in fig. 17.10(c). As soon as a signal pulse arrives the capacitor is discharged through the resistor with an exponential characteristic with (discharge) time constant $\tau_d = C_f R_f$ (fig 17.10(d)). No external control is needed. In particular at high signal rates a decay time as short as possible is desired to avoid superposition of signals at the output of the amplifier. Short time constants, on the other hand, increase the thermal noise contribution of the system, as shown in section 17.10.3. The time constants for discharging C_f through a resistor can be as large as 10–100 µs, for example in detector applications for spectroscopic measurements which require very low noise.

With values of $C_f = 4\,\mathrm{fF}$ (typical for readout chips of semiconductor pixel detectors with small detector capacitance C_D) one needs for $\tau = 10\,\mu\mathrm{s}$ a resistance of

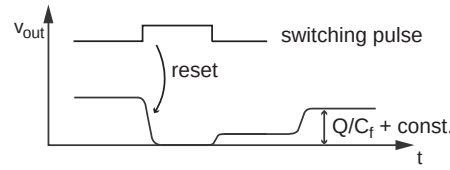
$$R_f = \frac{\tau_d}{C_f} = \frac{10\,\mu\mathrm{s}}{4\,\mathrm{fF}} = 2.5 \times 10^9\,\Omega = 2.5\,\mathrm{G}\Omega\,.$$

The fabrication of ohmic resistors in integrated circuits (chips) is very demanding. Resistors in the MΩ range can still be realised by MOS transistors (see section 8.3.5 on page 297) operating in the linear region of the transistor characteristic. Other methods are described in section 8.5.4.

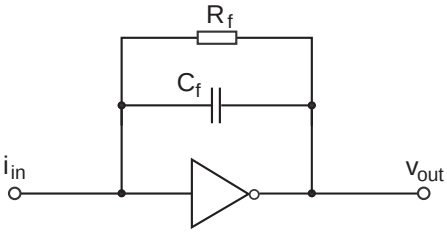
Another possibility to discharge C_f is by means of a current source in the feedback circuit. Circuit diagram and output signal are shown in figs. 17.10(e) and (f), respectively. In contrast to resistor discharging, the capacitor is here discharged by a constant current. This also results in a linear decrease with constant slope of the pulse after having reached the peak amplitude. The output pulse has a triangular shape (fig. 17.10(f)). Its width in time Δt is proportional to the peak height as well as to the total area of the pulse. A measurement of the time during which the output pulse is larger than a given threshold value hence is a measurement of the corresponding total signal, which itself is proportional to the energy deposited in the detector. This is a convenient and simple method to measure the pulse integral, called the *time-over-threshold* (ToT) method.



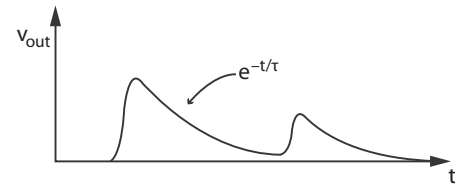
(a) Switch reset: circuit.



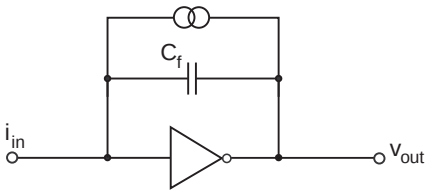
(b) Switch reset: output voltage.



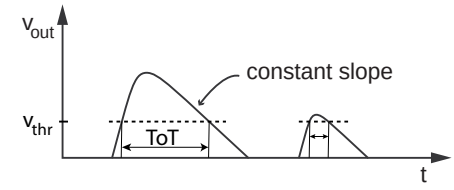
(c) Reset via resistor: circuit.



(d) Reset via resistor: output voltage.



(e) Reset via current source: circuit diagram.



(f) Reset via current source: output voltage.

Fig. 17.10 Preamplifier reset methods: (left) circuit diagrams, (right) output pulse evolution. (a,b) Discharging by means of a switch. The renewed rise with the trailing edge of the reset pulse is caused by so-called charge injection from the switching. (c,d) Discharging the feedback capacitor by means of a resistor. (e,f) Discharging the feedback capacitor by a constant current source. In (f), additionally, the principle of the ToT method is indicated (see text).

17.3 Pulse shaping

Further pulse processing is necessary (a) to avoid pulse superposition (*pile-up*), and (b) to reduce noise by frequency filtering.

As described in the previous section, the feedback capacitor in the preamplifier must be discharged to avoid spurious superposition of output signals (fig. 17.11 top). Typical decay times of electronic pulses of a preamplifier are in the range of some 10 ns to about 10 µs. Shorter and longer times are also possible depending on the application. If further signals arrive during the decay time, pile-up of pulses at the output still occurs despite discharging, as sketched in fig. 17.11 (bottom left). The superimposed signals can be separated by an RC high-pass filter (see also fig. 5.9 in section 5.3). A combination with an additional low-pass filter or even several high- and low-pass stages leads to more or less Gaussian shapes of the output pulses for every signal pulse at the input (fig. 17.11, top and bottom right). This electronic circuitry is hence called *pulse former* or *shaper*.

In practice the preamplifier output signals are superposed by noise signals which—with the exception of $1/f$-noise (see section 17.10.3.1)—contributes at all frequencies (white noise). To increase the SNR one can filter parts of the frequency spectrum.

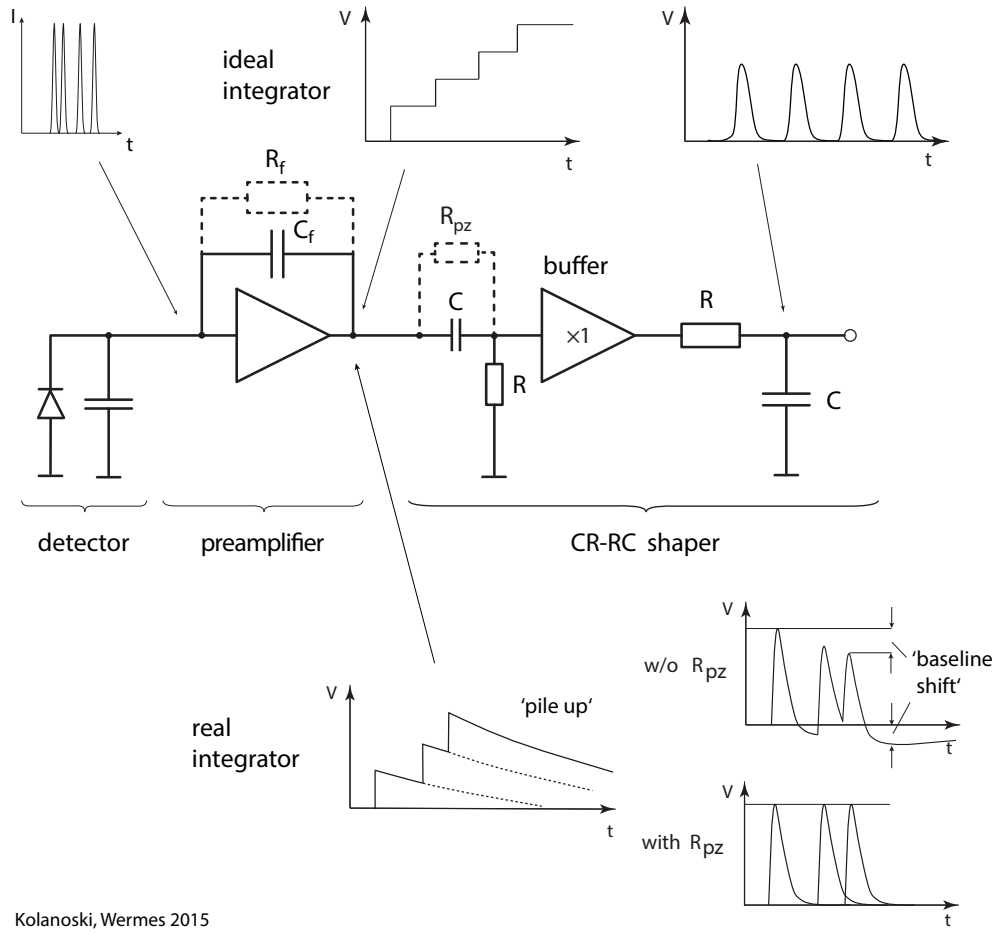


Fig. 17.11 Readout system consisting of detector, preamplifier and shaper. The pulse shapes at input and output of the preamplifier and at the output of the shaper are drawn at the top part of the figure for an ideal case (step functions without discharging) and at the bottom part for a real case (discharging of C_f, here by a resistor R_f, dashed). In the latter case an 'undershoot' develops at the shaper output (see text). By the resistor R_{pz} parallel to the capacitor of the high-pass filter (dashed) the undershoot is compensated (*pole-zero cancellation*, see text). The (×1)-amplifier in between the band-pass parts is a buffer preventing the low pass from supplying a real load to the high pass.

This is achieved by the shaper composed of high- and low-pass filters that limit the bandwidth. In addition, high-frequency transient signals are suppressed by the shaper. Low-pass filtering, however, also leads to larger signal rise times, which sometimes is undesirable.

17.3.1 Unipolar pulse shaping

In the simplest case the shaper consists of a high pass (CR filter) and a low pass (RC filter) with the same time constant $\tau = RC$, hence called a CR–RC shaper (fig. 17.11). The normalised output signal of this shaper upon a steep (step-function like) input

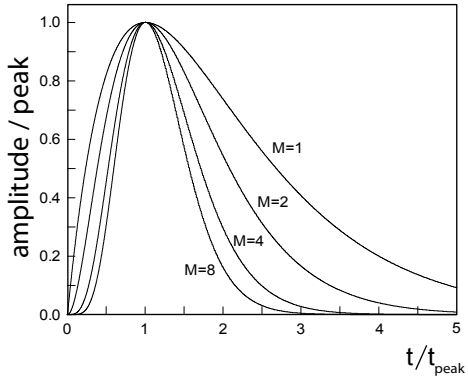


Fig. 17.12 Step response output of CR–$(\mathrm{RC})^M$ shaper stages of different order M. Amplitudes and times are normalised to their respective peak amplitude.

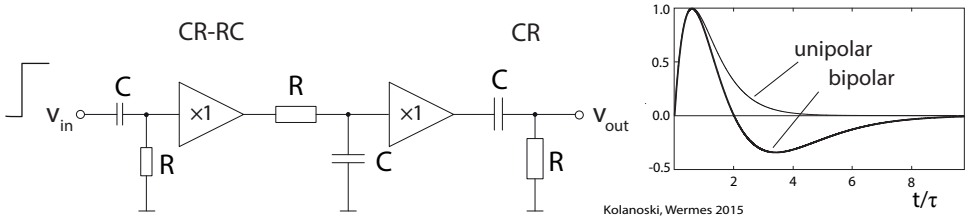


Fig. 17.13 Bipolar pulse shaping through diffentiation of a CR–RC pulse (circuit and output pulse) as well as a unipolar pulse (dotted line) for comparison. The pulses are normalised to equal peak amplitude and peaking time.

has the form (see also appendix H, eq. (H.13)):

$$A(t) = \frac{t}{\tau}\, \mathrm{e}^{-\frac{t}{\tau}}\,, \qquad t \geq 0\,. \tag{17.23}$$

Here $\tau = 1/\omega_c$ is the filter time corresponding to the inverse of the cut-off frequency ω_c of the filter, also called *peaking time* or *shaping time*. In higher order shaping circuits N high-pass and M low-pass filters are used ($(\mathrm{CR})^N$–$(\mathrm{RC})^M$ shaper). Usually only $N = 1$ is realised in most practical applications. In fig. 17.12 shaper output signals of different low-pass order M for $N = 1$ are displayed, normalised in amplitude and time to their respective peak amplitude values. The CR–RC pulse form (17.23) is modified to:

$$A(t) = \frac{1}{M!}\left(\frac{t}{\tau}\right)^M \mathrm{e}^{-\frac{t}{\tau}}\,, \qquad t \geq 0\,, \tag{17.24}$$

with the peaking time

$$t_{peak} = M\tau\,.$$

With higher order M the shape of the pulse becomes more Gaussian and, if one chooses a shorter peaking time, also narrower in absolute terms. The double-pulse resolution hence improves with the number M of low-pass elements. However, the electronic effort for higher order shapers is substantially larger than for CR–RC shapers.

17.3.2 Bipolar pulse shaping

By adding a further high-pass filter to any unipolar shaper a bipolar pulse shape is obtained with equal area in both wings. This can be done by adding a CR high-pass

filter after M low-pass filters (see e.g. [217]) or simply by adding a high pass to a CR–RC filter [905], as shown in fig. 17.13. For equal time constants τ in such a CR–RC–CR filter the obtained normalised pulse has the form

$$A(t) = \left[\frac{t}{\tau} - \frac{1}{2}\left(\frac{t}{\tau}\right)^2\right] \mathrm{e}^{-\frac{t}{\tau}}\,, \qquad t \geq 0\,, \tag{17.25}$$

which should be compared to the unipolar pulse of (17.23). The total pulse length of the bipolar pulse is larger than for unipolar shaping.

Historically, bipolar filtering has been introduced to cope with high counting rates in gamma-ray spectroscopy with scintillators [391]. Only by applying pole-zero networks (see below) and other electronic corrections, did unipolar shaping again became popular.

Bipolar pulse shaping is useful when at high count rates the circuit must be robust against (rate-dependent) baseline fluctuations and undershoots and when simplicity of an electronics implementation takes priority over a somewhat worse SNR, somewhat more power consumption and larger usage in chip area [391, 217]. For bipolar shaping the level at the output is independent of baseline fluctuations at the preamplifier input. Bipolar shaping is often found in time-critical applications, where the *zero-crossing* time, when the pulse crosses the baseline, can be exploited for precise timing. If an electrode is anyway AC coupled, for example to decouple an anode wire high voltage from the amplifier, unipolar shaping is not realisable with a linear electronics circuit.

17.3.3 Undershoots

Without further corrections unipolar shaping also features bipolar pulse tails in realistic applications. Due to the (slow) discharge of the feedback capacitor C_f of the preamplifier the shaper output pulse has an *undershoot*. This effect is sketched in fig. 17.11 (bottom right) and can be explained as follows. From its voltage input the shaper produces a pulse shaped according to (17.23) at its output only under the (ideal) condition that the input voltage remains constant (fig. 17.11(top)). In real applications this is not the case because C_f needs to be discharged, for example by a resistor R_f parallel to C_f (drawn dashed in fig. 17.11). The voltage at the shaper input hence slowly decreases and thus is not constant during the time of pulse shaping (17.23). The high pass of the shaper (differentiation) causes the shaper output to fall below zero (fig. 17.11(bottom)). The amplitude of a superimposed subsequent pulse will hence be distorted. Since the signal amplitude is proportional to the deposited energy charge in the detector, the energy resolution can also be affected, if the undershoot causes a noticeable baseline shift.

Pole-zero cancellation. The undershoot behaviour of our preamplifier–shaper system in fig. 17.11 can be compensated by adding a suitably dimensioned resistor parallel to the capacitor of the first high pass of the shaper. For an explanation we employ the mathematical treatment by means of transfer functions $H(s)$ in Laplace space (see Appendix H and references therein) which yield the (frequency) output of an electronic circuitry for each possible input:

$$v_{out}(s) = H(s)\, v_{in}(s)\,. \tag{17.26}$$

Here s is the complex Laplace variable which is related to the frequency ω as $s = \sigma + i\omega$, where σ is a constant generating convergence for integrals over s.

The transfer function of a preamplifier with capacitive feedback is in Laplace space:

$$H_{CSA}(s) = \frac{R_f}{1 + s\tau_f} \tag{17.27}$$

with $\tau_f = R_f C_f$, which in the time domain yields the shape given in (17.23). Since the CSA converts an input current into a voltage the corresponding transfer function has resistor units, different to the examples given in appendix H with dimensionless transfer functions. The frequency dependence, however, is the same for both.

For frequency analysis *zeros* and *poles* in the absolute value of the transfer function determine the frequency characteristic as well as phase and asymptotic behaviour of a circuit element. $H_{CSA}(s)$ in (17.27) has a pole on the real axis for $s = \pm 1/\tau_f$ which is responsible for the frequency dependence and the position of the cut-off frequency ω_c (see fig. 17.8).

The CSA, without discharge through R_f, responds to a δ-pulse input with a step function (see fig. 17.11), which in Laplace space is proportional to $1/s$ (eq. (H.11) in appendix H). In the appendix it is shown that when applying a CR–RC shaper transfer function to this step function, one obtains as output a unipolar exponential function ((H.13) and (17.24) for $M = 1$). Note that in this case the zero for $s = 0$ in the numerator of the (high-pass part of the) shaper transfer function (H.7) cancels the $1/s$ factor of the step function (H.11). If, however, the CSA feedback capacitor is discharged by the resistor R_f, the step function's $1/s$ shape turns into a shape as in (17.27) which is no longer cancelled by the s in the numerator of the shaper function (H.9). The zero of the numerator at $s = 0$, together with the poles of CSA and the high pass of the shaper at real values of s, cause a zero crossing of the pulse in the time domain and thus an undershoot (or overshoot, depending on the primary polarity).

By adding the resistor R_{pz} parallel to the capacitor of the shaper's high pass (drawn dashed in fig. 17.11) a further zero is introduced which can be dimensioned such that it cancels the pole in $H_{CSA}(s)$, that is, choosing $\tau_{pz} = R_{pz}C = R_f C_f = \tau_f$ (*pole-zero cancellation*).

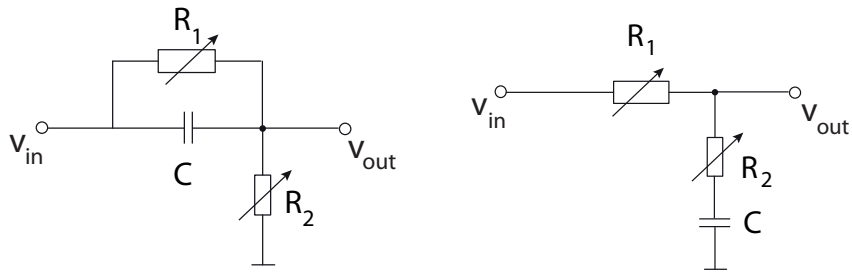
With $H_{pz}(s)$ being the high-pass part of the CR–RC shaper we have:

$$\begin{aligned} H(s) = H_{CSA}(s) \cdot H_{pz}(s) &= \frac{R_f}{1 + s\tau_f} \frac{1 + s\tau_{pz}}{1 + \frac{R_{pz}}{R} + s\tau_{pz}} \\ &\underset{\tau_{pz}=\tau_f}{=} \frac{R_f}{1 + \frac{R_{pz}}{R} + s\tau_f} \underset{R_{pz} \ll R}{\longrightarrow} \frac{R_f}{1 + s\tau_f}, \end{aligned} \tag{17.28}$$

where R_{pz} is chosen such that $\tau_{pz} = \tau_f$ with not too large R_{pz} and labels as in fig. 17.11. The undershoot is cancelled if R_{pz} is dimensioned as described, such that the form of $H_{CSA}(s)$ in (17.27) is now recovered for a system of CSA plus high pass (of the shaper).

Almost all spectroscopy amplifiers contain a *pole-zero compensation* network of the kind, as shown in fig. 17.14(a). The transfer function is

$$H_{pz}(s) = \frac{s + 1/\tau_1}{s + 1/\tau_2} = \frac{\tau_2}{\tau_1} \frac{1 + s\tau_1}{1 + s\tau_2} \tag{17.29}$$



(a) *Pole-zero* network. (b) *Zero-pole* network.

Fig. 17.14 (a) *Pole-zero* and (b) *zero-pole* networks for formation and shortening of signals.

with $\tau_1 = R_1 C > \tau_2 = \frac{R_1 R_2}{R_1+R_2} C$. One can easily verify that (17.28) corresponds to (17.29) by substituting $R_{pz} \to R_1$, $R \to R_2$ and using the definitions of τ_1 and τ_2. The network transforms an exponential signal proportional to $\exp(-t/\tau_1)$ in a signal of the form $\exp(-t/\tau_2)$ (see also [228, 217]). The circuit weights low-frequency parts down by the factor $k = R_2/(R_1 + R_2)$.

In a corresponding *zero-pole filter* (fig. 17.14(b)) with the transfer function

$$H_{zp}(s) = \frac{1 + s\tau_1}{1 + s\tau_2} = \frac{\tau_1}{\tau_2} \frac{s + 1/\tau_1}{s + 1/\tau_2}, \tag{17.30}$$

where $\tau_1 = R_2 C < \tau_2 = (R_1 + R_2)C$, high frequencies are weighted down by the factor k. Both filter types serve to form and shorten signal pulses (see also the following sections).

17.3.4 Ballistic deficit

Another problem appears if the pulse evolution (with time) at the shaper input lasts much longer than the high-pass time constant of the shaper. The former can be caused by a large charge collection time, for example when the drift velocity of charge carriers is slow, by a large input capacitance or also by an intrinsically slow preamplifier. In such a case the output amplitude of the shaper is trimmed by the slow rise of the preamplifier pulse. The falling edge of the shaper output already sets in before the preamplifier output has reached its maximum value. This problem is commonly called *shaping loss* or *ballistic deficit* (see also fig. 5.9 and text on page 146).

In fig. 17.15 the effect is sketched for the case without and with shaping loss. To avoid the ballistic deficit, the charge collection time and the preamplifier rise time must be sufficiently short compared to the peaking time of the shaper.

17.3.5 Tail cancellation

The shape of a signal pulse at the anode wires of gas-filled multiwire chambers (chapter 7) is characterised by a very short rise time in the ns-range, caused by the sudden start of the gas amplification near the wire, and a very long pulse tail, caused by the slow drift of ions from the anode to the cathode with time constants in the 100 µs- to ms-range (see section 5.3.2). The ion signal, being the dominant contribution for wire chamber signals, has been derived in (5.64) as

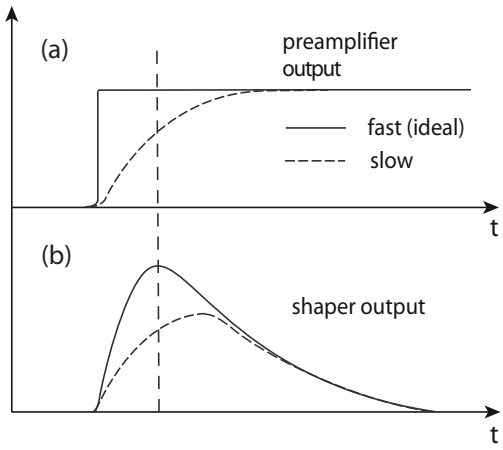


Fig. 17.15 Shaping loss: (a) step function pulse (without discharge) at the preamplifier output with a fast (solid line) and a slow rise (dashed line). (b) Resulting shaper output. If the shaper peaking time is short in comparison to the rise time of the preamplifier an amplitude loss occurs (*shaping loss* or *ballistic deficit*).

$$i(t) = \frac{Ne}{2 \ln b/a} \frac{1}{t + t_0^+} \tag{17.31}$$

with the signal charge Ne, anode and cathode radii a and b, respectively (see fig. 5.8), and the characteristic time t_0^+ for the rise of the ion drift signal given by (5.62). The pulse shape is shown in fig. 5.9(a) on page 146.

For the drift tube example given on page 145 with ratio $b/a = 1000$ and a typical characteristic time $t_0^+ = 1\,\text{ns}$ the charge signal (integrated current) reaches its maximum only after a time of $T^+ = 10^6\, t_0^+ = 1\,\text{ms}$. However, after about 10 ns, that is, after $10^{-5}\, T^+$, almost 20% of the charge amplitude is already reached.

Therefore, for high rate applications, short shaping times are needed, which, however, are usually accompanied by a loss in (charge) signal amplitude (shaping loss, see section 17.3.4). In addition—to avoid *pile-up* of signals—the long tail of the signal pulse must be eliminated electronically (*tail cancellation*).

A simple possibility [217] to achieve both is using bipolar pulse shaping (see section 17.3.2) by a sequence of high and low passes. The occurring negative part of the pulse (undershoot) and related disadvantages of bipolar pulse shaping (somewhat worse SNR than for unipolar shaping, larger total pulse duration, possible cross talk to neighbouring channels) are unwanted, but are to some extent compensated by the simple realisation of the bipolar circuit in practice.

If a shorted unipolar output pulse is used, a dedicated cancellation circuit is necessary, which compensates the accompanying effects (undershoots) of pulse shortening by high-pass filters by several subsequent pole-zero filters. The functionality of these filters has been described in section 17.3.3 and is effectively a transformation into an exponential signal with a different (shorter) time constant [217]. The RC times of CSA and shaper/pole-zero filters must be tuned such that the desired pulse length without undershoot results. Figure 17.16 shows a wire chamber signal behind a shaper with four low-pass stages ($M = 4$) [217] which shapes the input signal (17.31) and shortens it, but still shows tails towards times larger than 100 ns (dashed line, $t_0^+ = 1\,\text{ns}$). Using two appropriately chosen pole-zero filters a pulse is generated with negligible undershoot, already returning to the baseline after about 40 ns.

The filter time constants must be precisely tuned with respect to each other and to the characteristic signal pulse time t_0^+ in order to achieve short, unipolar pulses

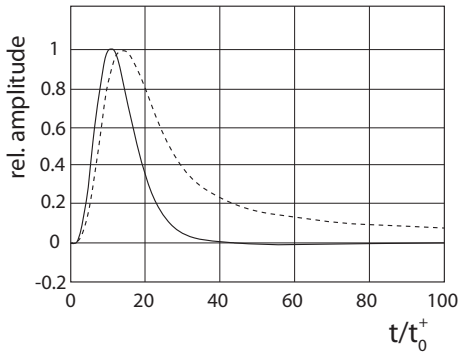


Fig. 17.16 Signal of a wire chamber formed by a (four-stage) unipolar shaper with $10\,t_0^+$ peaking time (dashed line) and reduced to a total length of about $40\,t_0^+$ by two subsequent pole-zero filters (solid line, adapted from [217] with kind permission of Springer Science+Business Media).

without undershoots. Variations in the fabrication of electronic components or chips lead to channel-to-channel variations which can amount to up to 10–20%. Furthermore, for filter-based pulse-tuning the entire electronics chain must be DC-coupled [217], a requirement not realisable for example for HV-biased wire readout in a wire chamber, as already mentioned in section 17.3.2. Further examples and calculations for pulse shaping in wire chambers can be found in [217].

17.4 Sample-and-hold technique

In many experiments the temporal shape of the input signal and also its arrival time, for example relative to a trigger signal, are known. In this case it is possible to *sample* the voltages of input signals at a fixed moment in time, *hold* them for some time, and further process them at a later time. This technique is called *sample-and-hold* (see e.g. [540]). Figure 17.17 shows an application. The switches S_A are assumed to be closed, S_B open. A sample pulse, which is derived from a trigger pulse and relatively delayed such that it is coincident with the peak of the signal pulse, opens the switches S_A for a sampling time interval. The voltages lying at the capacitors C_i during this time charge them and are thus memorised. For readout the switches S_B are successively closed (here shown for channel 2), hence sequentially transferring the stored voltages serially onto the analog-to-digital converter (ADC) for digitisation. The first buffer amplifier (with amplification 1) acts as impedance converter generating a copy of the input signal with small impedance at the input to the capacitor. The output buffer has a high input impedance, ensuring a small load to the capacitor and an unbiased transfer of the voltage from the capacitor to the output.

By sampling the pulse in successive time intervals and storing the voltages in a series of sample-hold cells, called analog memory, one can memorise the whole pulse shape (*wave form sampling*).

17.5 Discrimination

For a measurement one wants to distinguish real signals, generated for example by particles, from noise hits caused by (electronic) fluctuations and to suppress the latter. A discriminator decides if an input pulse surpasses a set (voltage) threshold and, if so, generates a rectangular output pulse, usually in a logic standard (fig. 17.18(a)). This way noise hits with amplitudes below the threshold are filtered out. At the same time digital output signals are created from analog inputs that can be used in further processing logic. Discriminators are particularly used also for gating and triggering

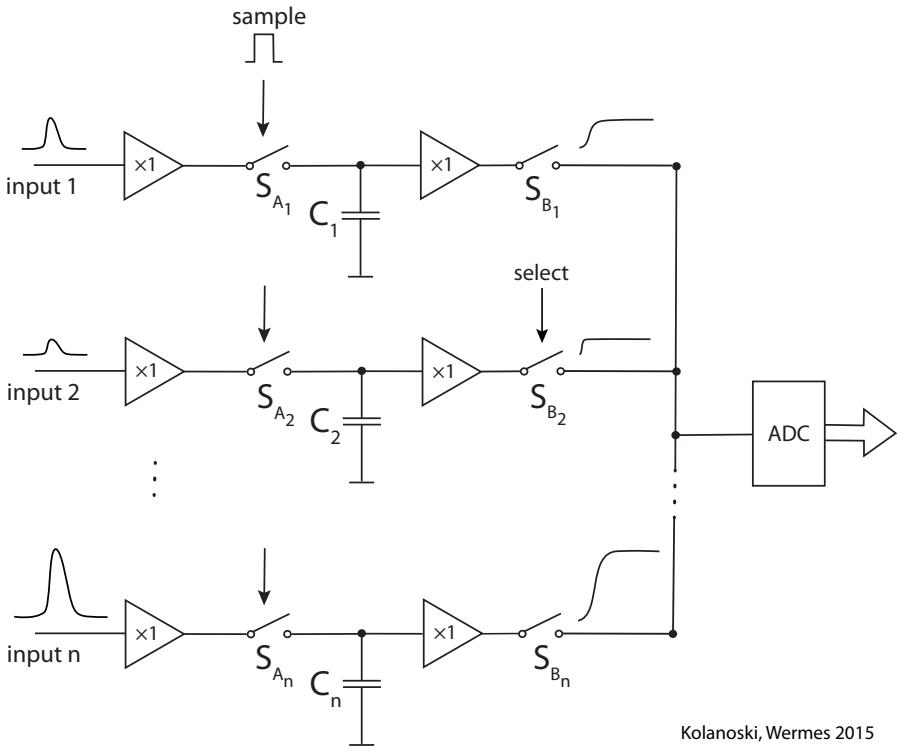


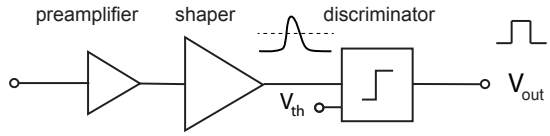
Fig. 17.17 Sample-and-hold technique applied in an example with parallel input and serial, digitised output (see text).

of other parts of the electronics chain. Discriminator signals can for example control data recording of an oscilloscope (discriminator pulse on the trigger input of the oscilloscope) or of a multi-channel analyser (MCA, discriminator signal on gate input of the MCA) or they can be used as logic signals in subsequent electronic logic gates.

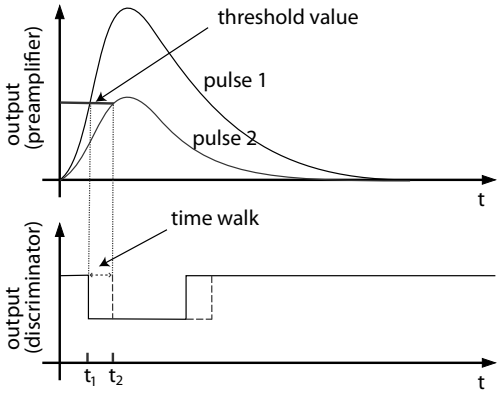
17.5.1 Time walk

In time-critical applications the problem of *time walk* is encountered for input pulses of different height. For different peak amplitudes at the discriminator input the output pulses are issued at different times depending on when the input pulse crosses the threshold. Figures 17.18(b) and 17.18(c) show how the response time depends on the pulse height of the input signals. The precision of time marking by *leading edge discrimination* typically lies in the range 400–500 ps at best [652].

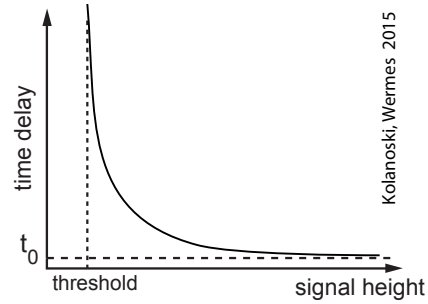
In applications demanding high timing precision, such as time-of-flight measurements with fast scintillators and resolution demands in the order of 100 ps (see section 14.2.1), time fluctuations must be minimised. They can arise for example from amplitude variations (time walk) as well as from fluctuations of the pulse shape (*time jitter*) or other electronic fluctuations. If the shape of pulses does not vary with time, time walk caused by different amplitude heights can be avoided by triggering on the zero-crossing point of a bipolar pulse (*zero-crossing triggering*). Such circuits are for example given in [652] or [905].



(a) Readout system with amplifier, shaper, and discriminator.



(b) Time walk.



(c) Dependence of time walk on the signal pulse height (schematic).

Fig. 17.18 (a) Readout chain with discriminator. (b) Time walk behaviour in systems with discriminator. (c) Time delay dependence of the discriminator output on the pulse height above discriminator threshold; t_0 quantifies the minimal delay caused by the discriminator electronics only.

Every unipolar pulse can be turned into a bipolar one, for example by a subsequent RC filter. For the same pulse shapes, the zero-crossing time of a bipolar pulse is independent of the pulse height. This method, however, requires a very stable pulse shape and reacts in a very sensitive way with time jitter to shape fluctuations. A very efficient and stable timing method is described in the next section.

17.5.2 Constant-fraction discrimination

It was empirically found that for leading-edge discrimination the best timing resolutions are obtained if the threshold is set at about 10–20% of the pulse peak amplitude. This has led to an extended method that generates a time mark by issuing an output pulse when reaching a predefined fraction k (*constant fraction*) of the peak amplitude. This point in time is independent of the absolute pulse height as long as the rising edge of the pulse is unchanged in shape (fig. 17.19(a)).

The electronic realisation of constant-fraction discrimination is shown in fig. 17.19(b). The output pulse is split and in one branch inverted and attenuated by a fraction $k < 1$. In the other branch the output pulse is delayed by a time t_d, which must be smaller than $(1-k)t_{peak}$, where t_{peak} is the peaking time of the pulse. The pulses of both branches are then superimposed:

$$v_{sum}(t) = v_{in}(t - t_d) - kv_{in}(t)\,. \tag{17.32}$$

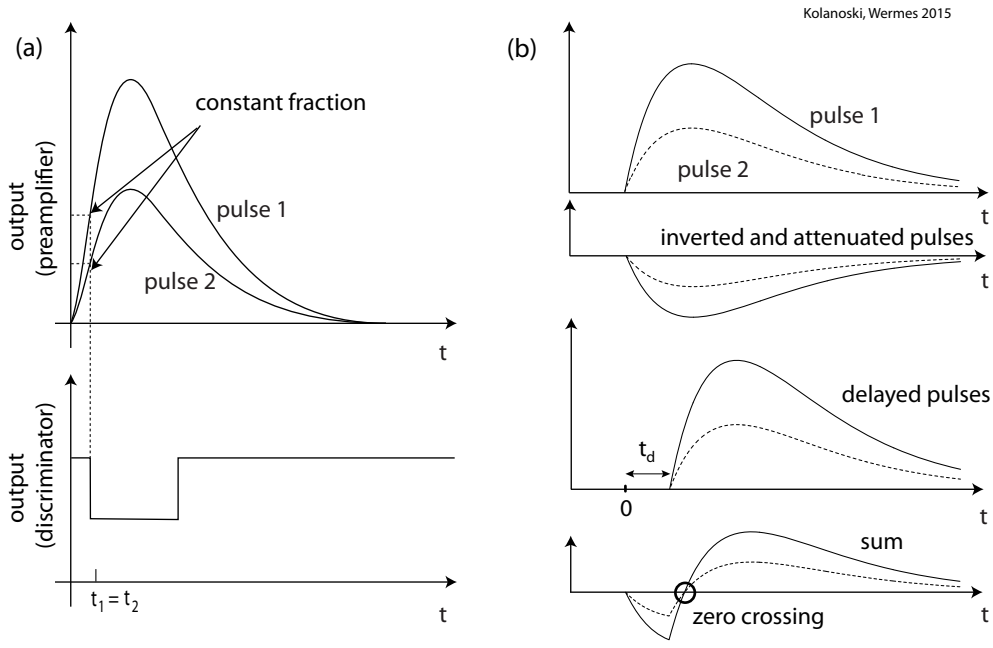


Fig. 17.19 *Constant-fraction triggering* features small timing fluctuations for pulses with constant shape. (a) Two pulses of different height are discriminated at a fixed peak-amplitude fraction; top: pulse shape, bottom: output of the discriminator, constant in time. (b) Electronic realisation: an inverted and attenuated pulse is superimposed on the delayed pulse. From top to bottom: two output pulses of different size; inverted and attenuated pulses; delayed pulses; addition of inverted and delayed pulses yielding with constant zero-crossing point.

The zero-crossing point depends on the constant fraction k by which the original pulse is attenuated. The method exhibits comparatively small timing fluctuations (typically 10–30 ps) over a large range of amplitudes.

17.6 Integrated circuits for detector readout

In modern detectors the first stages of the readout are often realised by custom-designed integrated circuits (ASICs), which contain in one densely integrated VLSI chip besides amplification also pulse shaping and discrimination, often also digitisation and a complete readout architecture. Depending on the experimental demands the readout of a detector can be asynchronous or can be started synchronously by a *trigger*. The variety of applications is particularly large for micropattern detectors (semiconductors, also gas-filled detectors) with strip or pixel patterning (sections 8.6, 8.7, and 7.9). We have therefore here chosen these types of detectors as representative examples for a variety of detectors.

17.6.1 ASIC chips: concept examples

Experimental conditions often demand very different readout variants even if the type of detector is the same. For semiconductor detectors in the form of microstrip or pixel detectors the signal current is usually integrated for every readout channel (strip

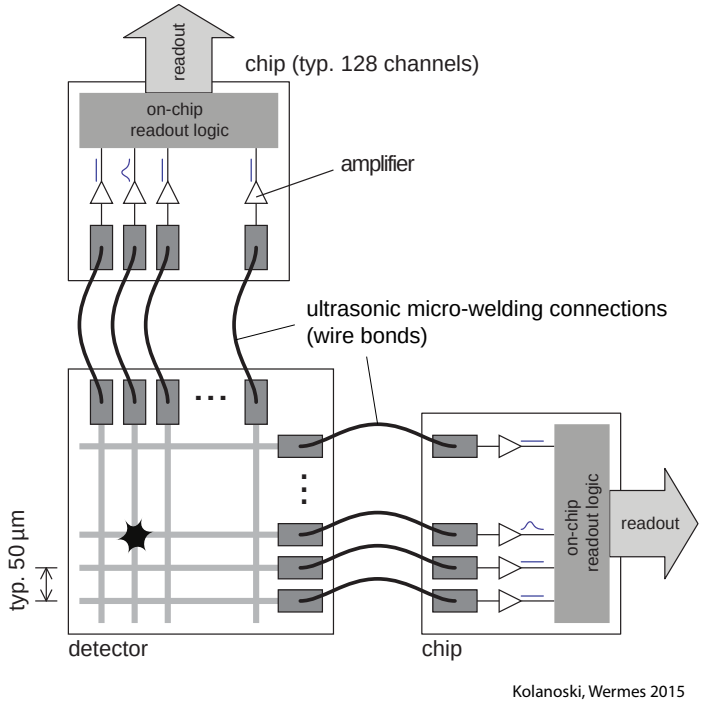


Fig. 17.20 Typical microstrip detector with chip electronics.

or pixel) and the charge is stored as an analog value. The stored information can be supplemented by further characteristic data, for example information about the selection (trigger) or the time of signal arrival relative to the trigger time. In imaging applications, especially biomedical imaging, trigger or signal arrival-time information is usually not of central importance. Often only the number of hits in every readout cell is counted for a certain exposure time and only the counter status is read out after exposure. Counting architectures are also interesting for high-precision measurements of spatial profiles, for example particle beam profiles to determine beam polarisation (see e.g. [359]). The readout of individual events is not needed for these applications and often also not possible due to the high hit rate.

In what follows we describe a few principle concepts. To distinguish one- and two-dimensional space information the integrated circuits for detector readout will simply be called strip or pixel chips, respectively, even if the usage of such readout circuits is also suitable for other detector types, for instance gas-filled detectors.

Figure 17.20 shows a typical layout of a microstrip detector with readout of both (orthogonally oriented) sides by amplifier chips. Fabrication and functional principle of double-sided semiconductor detectors are described in section 8.6.

By means of ultrasonic wire bonding the strip detector is connected to the readout chips. Typically one chip reads out 128 strip electrodes. A channel of a typical readout chip contains a charge-sensitive preamplifier, a pulse shaper, and a discriminator which allows setting a threshold to recognise a real signal above a noise background. The signal is stored either in analog or digital fashion. Sometimes a dedicated logic selects only channels with signal hits above threshold to be read out (*zero suppression*). If external trigger signals are used to initiate transmission of the locally stored hit information, events without a trigger are discarded, thus reducing the amount of data

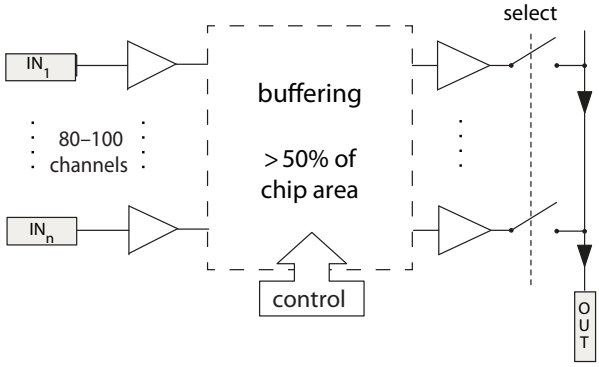


Fig. 17.21 Microstrip chip architecture for LHC experiments

to be processed. In high energy physics experiments trigger signals for strip detectors are often derived from other subdetectors of the experiments that feature fewer readout channels for which fast trigger generation can more easily be accomplished. Also the clock signal for event recording (e.g. the beam crossing frequency) can be used as a trigger.

Microstrip detectors are used in fixed-target and in collider experiments for precision measurement of charged particle tracks and for secondary vertex identification of long-lived particles (see chapter 8 on page 256). The design of the chips depends on the requirements and experimental conditions of a given experiment. In the following we describe a few typical examples.

17.6.2 Microstrip readout chips

17.6.2.1 Chips for collider experiments with low particle rates

In comparatively low-rate experiments as for example at e^+e^- storage rings (compared to hadron collider experiments) the following operation conditions are important:

- As for all collider experiments, the time of the collision of the beams is precisely known to $\mathcal{O}(100\,\text{ps})$.
- The interactions per crossing of e^+ and e^- bunches that generate signals in the detector are rather rare (typically $\mathcal{O}(10^{-6})$), in contrast to hadron colliders where several to several hundred interactions per bunch crossing take place within one bunch crossing.
- A trigger selects the collision events of interest.
- A readout taking a relatively long time (some ms) can be afforded.

Under such conditions the following scheme can be realised:

(a) A chip reset before a new signal arrives is possible and is needed.
(b) The charge can be locally stored on a capacitance (sample-and-hold, see section 17.4) at a fixed time.
(c) Sequential readout of all channels (analog or digital) is possible after a trigger.

A typical readout scheme for these experimental conditions employs the sample-and-hold technique discussed in section 17.4 (fig. 17.17).

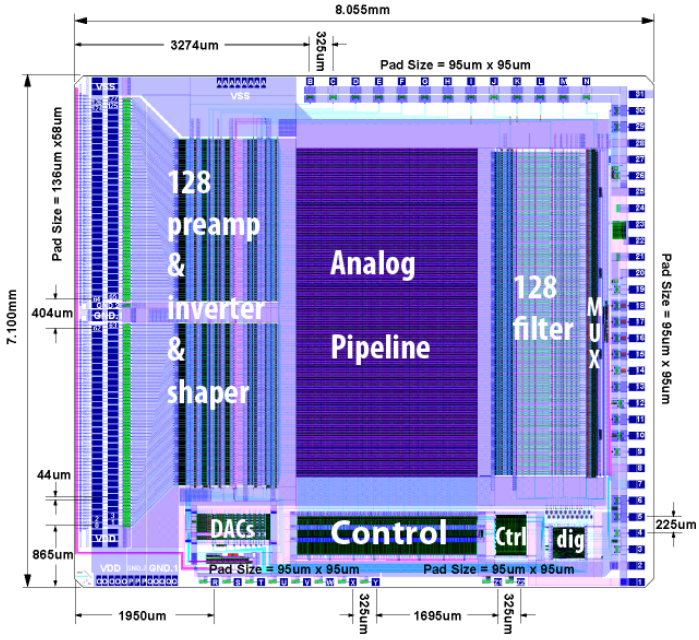


Fig. 17.22 Floor-plan of the APV chip [810] (8.055 mm × 7.1 mm) of the CMS experiment at the LHC [754]: bottom: digital steering, top: analog part with amplifier, shaper, pipeline and a switched-capacitor filter, as well as the output multiplexer.

17.6.2.2 Chips for collider experiments with high particle rates

At accelerators with very high event rates (e.g. LHC, HERA, TEVATRON) further additional requirements must be added (numerical examples for LHC):

- The collision rate is very high (up to 40 MHz, collisions every 25 ns).
- On average more than one collision occurs per bunch crossing, for example about 25 at the design luminosity of the LHC of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and about 200 at the HL-LHC [285].
- Per bunch crossing a large charge particle multiplicity is created (at the LHC about 1200 per collision, about 10^{11}/s).
- The collision point is precisely known in time ($\ll$1 ns).
- The trigger signal is issued only after a comparatively long time (>2.5 µs at the LHC) after the collision, during which hundreds of further bunch crossings occur.
- The readout time lasts at least several microseconds (see chapter 18).

These additional conditions mean that a local storage of all hit information is necessary until the trigger arrives. Electronically, this is particularly extensive if signals are stored on-chip in an analog rather than digital way, which can lead to usage of more than 50% of the chip area. Figure 17.21 shows the architecture of an LHC microstrip readout chip in a very simplified way. Data buffering is done in so-called *data pipelines*. Figure 17.22 shows a microphotograph of the APV chip used in the CMS experiment. The areas used by the analog part (preamplifier and shaper) and pipeline as well as those used by filters and digital control blocks are typical examples for comparison.

Zero suppression and hit discrimination. In experiments with large latency of the trigger or without trigger at all, it is often not possible to provide sufficient on-chip storage and all data must be read out periodically. In order to reduce the readout bandwidth, channels without hits are not read out (zero suppression). Hit recognition

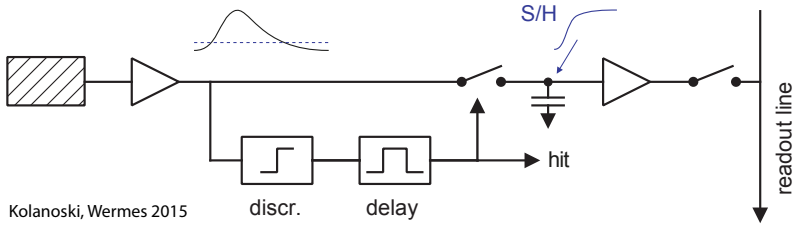


Fig. 17.23 Block diagram for X-ray detection with self-triggering and sample-and-hold (S/H).

for every event or image is performed by means of an on-chip discriminator. For hits found above a given threshold the corresponding hit coordinates (address, time stamp) are read out.

17.6.2.3 Chips with self-triggering

Detection of radiation in imaging applications for example with X-rays or in β autoradiography often faces very different requirements than the detection of particles from collision events, in particular:

- The time when X-rays or other radiation impinges the detector is random.
- No external trigger signal is available.
- Data rates typically can be high ($\gtrsim 1$ MHz per channel).
- Medium to high spatial resolution (10–100 µm) is usually aimed for.
- Signals are often smaller than those of minimum-ionising particles such that low threshold and low noise operation is mandatory.

Chips for such applications also require a discriminator in every channel that allows separation of signal hits from noise and can generate (self) trigger signals. Only hit strips and sometimes nearest neighbours are read out.

Figure 17.23 shows how the signal hit itself can generate a fast trigger (*self triggering*) which after some appropriately chosen time delay can initiate sampling of the signal peak amplitude on a sample-and-hold capacitor (section 17.4). The readout is usually also initiated by the trigger.

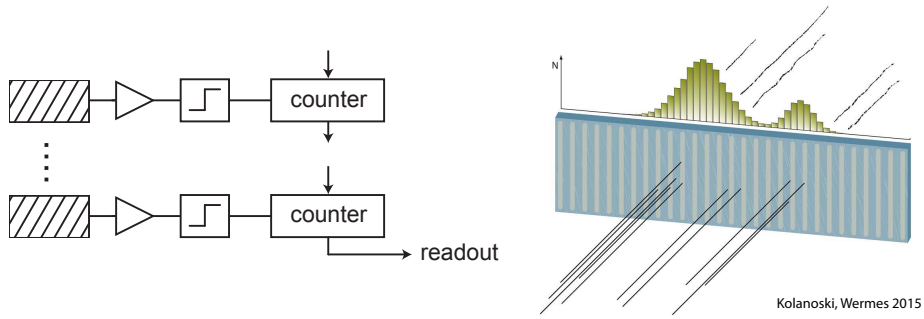
17.6.2.4 Profile measurements

In situations with very high counting rates, for example near the primary beam of an accelerator or in applications measuring the polarisation of an accelerator beam by laser backscattering, often one is only interested in the distribution of hits (the profile). An application example of such a profile measurement using strip detectors is shown in fig. 17.24. The hits are not individually read out but rather the number of hits on a strip electrode are counted if they issue a discriminator output above a certain threshold. A 'counting' architecture counts all hits per channel. Counter readout and reset is necessary before overflow occurs.

17.6.3 Pixel chips

17.6.3.1 Readout chips for LHC pixel detectors

For 2D-structured pixel detectors (see section 8.7) it is no longer possible to place the readout chip at the sides of the sensor as done with strip detectors. The number of



(a) Counting microstrip chip. (b) Beam profile measurement.

Fig. 17.24 Readout chips for microstrip detectors with counting circuitry: (a) principle layout of a cell, (b) application example.

channels increases quadratically with the linear dimensions of the detector. Contact access can no longer be obtained by wire bonds but a bonding technique is employed using bump bonds whereby both parts, chip and sensor, are micro-mechanically mated using flip-chip technology (so-called hybrid assembly described in section 8.7). The pixel cells have identical dimensions for readout chip and sensor (see fig. 8.53 on page 315). Larged hybrid pixel detector assemblies are used for example in the LHC experiments.

The pixel-chip readout architecture is shown in fig. 17.25. In the ATLAS FE-I3 pixel chip [771] a charge-sensitive amplifier is discharged by a constant current source (see fig. 17.10(e)). The linearly rising and falling flanks create a triangular shape of the amplified pulse which via the discriminator output generate pixel address (row and column number) and time stamp as well as *time-over-threshold* (ToT) information (see also section 17.5). The ToT provides a coarse measure (about 7 bit) of the deposited charge in the detector. Address and time stamps are first stored in a RAM located in the pixel cell. By means of a fast[9] scan process over all pixels in a double-column the hits are moved into buffers in the passive zone at the bottom of the chip where they are kept waiting for a potential trigger coincidence.

17.6.3.2 Pixel chips for X-ray imaging

In imaging applications with X-rays individual readout of detected photons is not mandatory, but image recording only based on the hit count in every cell has many advantages (fig. 17.26(a)). The principle [414, 271] is very similar to that of counting chips of microstrip detectors described in the previous subsection. However, interesting for X-ray imaging are mainly counting pixel detectors with two-dimensional spatial resolution and additional spectral information. For the latter an estimate of the energy of the detected X-ray quanta is needed. Since the photon energies are not individually measured at least a classification into different spectral X-ray 'hardness' is useful. This can for example be achieved on-chip by comparing the detected photon signals with thresholds of several different discriminators, each with a subsequent counter counting the signals above threshold. After an exposure time the counters are read out. Counting X-ray detection with up to eight discriminators with different thresholds is realised in

[9] The maximum bit transfer rate from the column to the end-of-column buffers is 20 MHz in the FE-I3 chip which is half the beam crossing frequency.

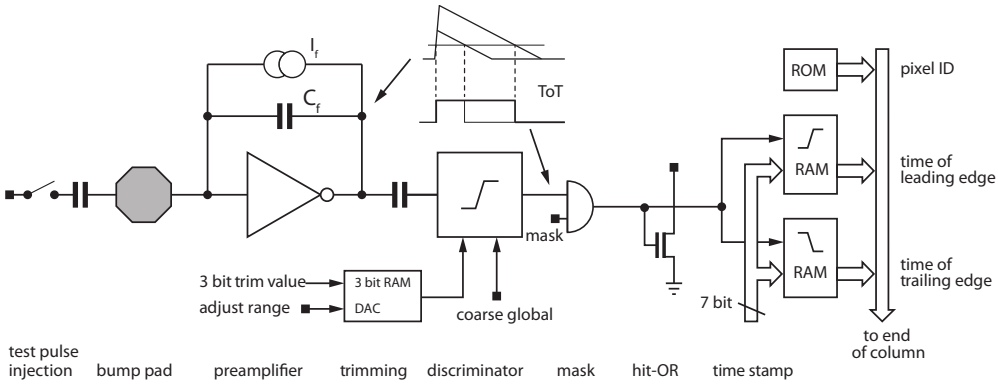


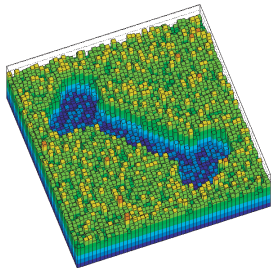
Fig. 17.25 Readout scheme for the ATLAS pixel chip (FE-I3) [771]. The input is connected to the pixel cell of the sensor (not shown) via the octagonal bump contact (far left) as in fig. 8.53. For test purposes charge can be injected to the preamplifier through an injection capacitor. The charge-sensitive preamplifier is discharged by a constant current, the value of which determines the pulse duration. Upon threshold crossing at the rising and falling edges the discriminator issues an output voltage step (rising or falling, respectively) which triggers the storing of time marks for each pixel. These are issued with a precision of 25 ns from a distributed 40 MHz (bunch crossing) clock, and are stored in two RAM cells with their difference providing a time measure (*time-over-threshold*, ToT) which is proportional to the pulse integral, that is, the deposited charge (see fig. 17.10(f) on page 728). The ToT time (7 bit) and the pixel address (row, column) are written into the output buffers whose contents are read out upon a trigger signal which selects a particular time stamp associated to each hit to be 'interesting'. By 'mask' defect cells can be deactivated; 'hit-OR' is an 'OR' output of many (selectable) pixels or even the entire chip.

the Medipix3 chip [145]. The principle is shown schematically in fig. 17.26(b). Pulses of every readout channel are given on two discriminators with different thresholds, the outputs of which are counted. The counters are read out after a predefined exposure time (for more information see the Medipix Collaboration webpage [270]).

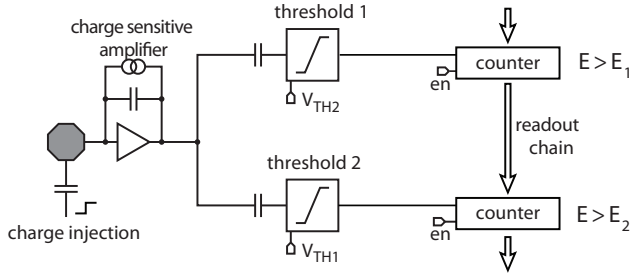
In particular, at low count rates the counting principle is superior to the charge integrating detection principle of film or foil systems: the response is linear over a large range and under- or overexposure effects are substantially reduced compared to many other X-ray detection systems. Saturation due to dead time effects (see section 17.9), however, limits the dynamic range[10] of counting pixel chips. This typically occurs around count rates above about 5 MHz per pixel corresponding to mean time differences between hits of 200 ns. At higher photon rates the integration technique therefore has advantages.

A combination of counting and integrating X-ray detection in every pixel combines the advantages of both principles, thus improving the imaging performance. This has been shown in [408] and is explained in fig. 17.26(c)). The input charge is simultaneously stored on the feedback capacitances of both branches, counting C_f and integrating C_{int}. While C_f in the counting branch is continuously discharged by a constant current, as explained in the previous section for the ATLAS pixel chip, in the

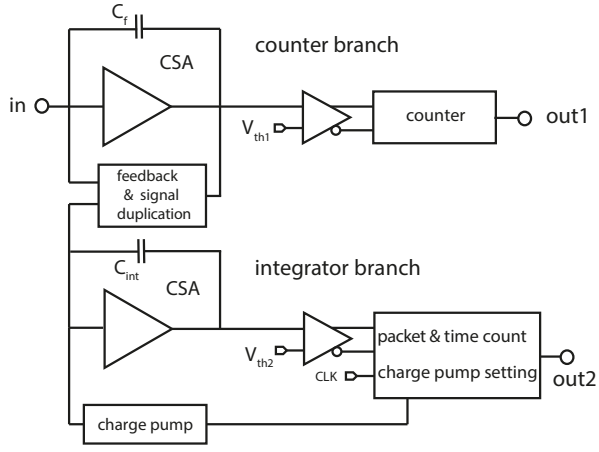
[10]The *dynamic range* here specifies the range between minimum and maximum count rate per channel without losses.



(a) Imaging by counting (illustration).



(b) Block diagram of the counting circuitry.



(c) Simultaneously counting and integrating readout chip (block diagram).

Fig. 17.26 Principle of X-ray imaging by counting of individual X-ray quanta. (a) The concept: after counting the quanta the raw image is given already in digital form by the counts of the counters in every cell; (b) circuit principle as realised for example in [145] and [668]. The analog part is similar to the one in fig. 17.25. However, no address or time information is issued, but the hits are individually counted if their energies are above threshold E_1, respectively E_2; (c) simultaneously counting and integrating architecture (from [408], description see text).

integrating branch C_{int} is discharged by a pulsed charge pump, pulsed with frequency f_{clk}. An adjustable amount of charge Q_p is successively subtracted (pumped off) for a time $1/f_{clk}$ from the charge stored on C_{int}. As long as a signal current I_{sig} is present at the input, a saw-tooth type charging and discharging of C_{int} occurs. This allows determining I_{sig} by the amount of subtraction cycles N_f:

$$I_{sig} = \frac{N_p\ Q_p}{\Delta t} = \frac{N_p\ Q_p\ f_{clk}}{N_f}\ . \tag{17.33}$$

Here Δt is the (measured) time between the first and the last pump cycle within the measurement interval.

In the low-rate region counting readout is superior to integrating in terms of linearity and contrast. At high rates $\gtrsim 10\,\mathrm{MHz}$ integrating is superior. Within these limiting

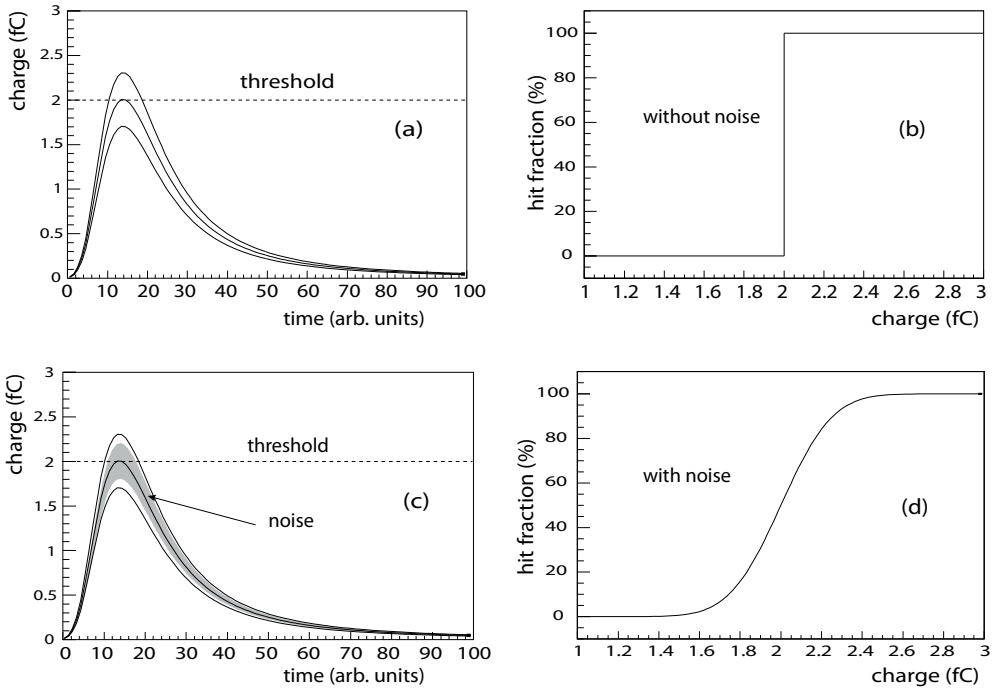


Fig. 17.27 Principle of threshold scans: (a,b) ideal case without noise, (c,d) real case including noise. Left plots: discriminator input pulses of different heights close to threshold; right plots: discriminator response in percent as a function of the injected charge (corresponding to different pulse heights). In (c) the influence of noise is displayed by the grey shaded area superimposed on an input pulse just touching the threshold. Noise causes an S-shaped smearing of the ideal step function shown in (d) (adapted from [484] with kind permission).

regions one obtains for mid size rates a twofold information: the photon count by the counting branch and the energy, that is, the total charge deposited in a pixel, by the integrating branch. The ratio yields the average energy per photon, an information which can be used for example to visualise the spectral hardening of an X-ray beam in imaging applications [637].

17.6.4 Chip characterisation: the threshold scan

For systems with many channels, as in strip or pixel detectors, the exact knowledge of threshold values (voltages) of every readout channel and their dispersion is important to characterise the response of the detector. By implementing fine tuning for every channel, threshold inhomogeneities can be minimised.

An often used method to determine the threshold of a channel is to inject a variable amount of charge (a voltage pulse injected over a capacitance) at the input of the charge-sensitive amplifier, as shown for example at the inputs in figs. 17.25 and 17.26(b). A scan is conducted by successively injecting pulses of pulse height V_{inj} on the injection capacitance C. Behind the capacitance the pulses correspond to a fixed amount of charge $Q_{inj} = C\,V_{inj}$, similar to the signal charge of a detected particle. The (voltage) output of the CSA is compared to the threshold voltage set at the dis-

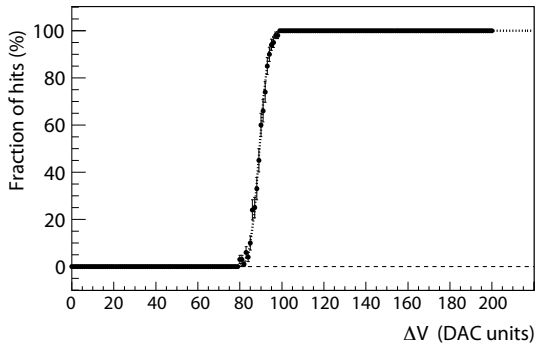


Fig. 17.28 Measured discriminator response (S-curve) in the ATLAS pixel detector [484] [3]. Plotted is the discriminator response in percent as a function of the pulse height of the injection pulse as in fig. 17.27 in units of the digital-to-analog converter (*DAC units*).

criminator. If the pulse height is higher than the threshold voltage the discriminator generates a digital output signal, else the output signal is suppressed. This is repeated N times for every injection voltage step with V_{inj} being steadily increased. Comparing the number of output pulses and injection pulses at each step one obtains a response characteristic, as given in fig. 17.27. Alternatively the threshold voltage can be varied for constant injection charge. For this measurement the size of the input capacitance must be known or a separate calibration measurement by using radioactive sources of known energy is needed.

In fig. 17.27(a) and (c), signal pulses of different height are shown together with the set threshold, in (a) without and in (c) including noise. Figure 17.27(b) and (d) show the scan curves obtained from the discriminator response when stepwise incrementing the injection voltage V_{inj}.

In the ideal case without noise the number of output pulses is zero, as long as the input pulses are below threshold, jumping abruptly to 100% as soon as they are above threshold. Due to noise fluctuations of the preamplifier signal at the input of the discriminator the response characteristic changes from the step function form to an S-shaped curve, as shown in fig. 17.27(d).

Figure 17.28 shows the measurement of an S-curve for one readout channel of the ATLAS pixel detector [484]. From the inflexion point of the S-curve the threshold value $Q_{thr} = C_{inj}\, V_{thr}$ is determined. From the width one obtains the noise σ_{noise} of the preamplifier (assuming that the noise distribution is Gaussian) by the following functional fit:

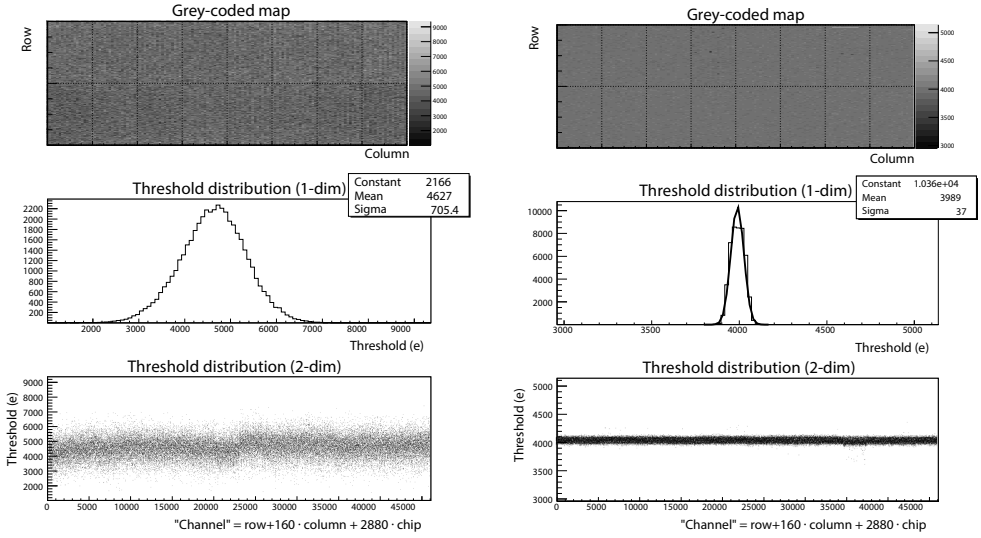
$$P_{hit}(Q) = \Theta(Q - Q_{thr}) \otimes \exp\left(\frac{-Q^2}{2\sigma_{noise}^2}\right) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{Q_{thr} - Q}{\sqrt{2}\sigma_{noise}}\right). \tag{17.34}$$

The $\otimes$ sign denotes a convolution, here of the Θ step function with the Gaussian. The function

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp\left(-\tau^2\right) d\tau = 1 - \mathrm{erf}(x) \tag{17.35}$$

is the complementary error function, derived from the error function erf(x).

In fig 17.29 the distributions of the threshold values measured by the S-curve method for the 2880 pixels of one chip of a 16-chip ATLAS pixel detector module are shown. For this detector with more than 80 million channels a narrow threshold distribution is an indispensable requirement for reliable operation at the LHC. There-



(a) Before fine-tuning. (b) After fine-tuning.

Fig. 17.29 Measured threshold distributions for 46 080 pixels of an ATLAS pixel module using a radioactive gamma source (^{241}Am, 60 keV γ) [484] (a) before and (b) after fine-tuning of discriminator thresholds: (top) grey-coded maps that allow for spotting deviations from the average value by which outliers can be detected, (middle) one-dimensional distributions and (bottom) two-dimensional scatter plots. In the case of the shown modules the grey-coded maps show little structure, indicating that almost all pixels are efficient.

fore, after a coarse pre-tuning, the thresholds are fine-tuned by means of a dedicated calibration circuitry.

17.7 Digitisation

The measured signals in detectors are first processed as analog electronic quantities. At the end of the readout chain usually there is a computer processing the digital information. The digitisation of the analog input therefore is part of the readout chain. Where in the chain, early or late, digitisation is done depends on whether measurement information is obtained from analog quantities (e.g. zero suppression as described in section 17.6.2.2) or the combination of several channels to a 'signal sum') need to be formed or not. Often the processing of analog information can be done faster and with less electronic overhead than digital processing. In such cases digitisation at the end of the readout chain is preferred.

Often digital values have to be converted to analog values, for example to set the voltage level of a discriminator threshold. In order to supply them, digital values received from a computer need to be converted into analog voltages.

In this section we discuss some basic concepts and properties of analog-to-digital (A/D) and digital-to-analog (D/A) conversion. For a detailed description we refer to the manifold literature, for example [768, 599].

17.7.1 Analog-to-digital conversion (A/D)

Digitisation of an (analog) voltage pulse is performed in a so-called *analog-to-digital converter* (ADC). A voltage proportional to the area or the peak voltage of a pulse is encoded in a binary number, the *code*. Sometimes the entire shape of an input pulse is stepwise sampled and digitised. In many ADCs the conversions are carried out at a fixed clock rate. A clock rate of 100 MHz for example then means for some ADCs one digitisation every 10 ns, for others it means 10 ns per bit are needed.

A binary number with n bits can take 2^n values. During the digitisation the voltage range from V_{min} (usually $V_{min} = 0$) to V_{max} is subdivided into $2^n - 1$ intervals, each of which is assigned a binary value. If the conversion is linear, the range is divided into intervals of equal size. The finest voltage step corresponds to the *least significant bit* (LSB) positioned most to the right in a binary number. In an n-bit ADC one LSB corresponds to a voltage step of

$$1\,\mathrm{LSB} \mathrel{\widehat{=}} \frac{V_{max} - V_{min}}{2^n - 1}\,, \tag{17.36}$$

taking V_{min} and V_{max} as the centres of the first and last voltage bins, respectively. In a 12-bit ADC, for example, and an input voltage range of 2.2 V the LSB corresponds to a voltage step of 0.537 mV.

A number of methods exists to digitise a signal (usually a voltage signal). The term *charge-integrating* ADC means that the integrated input current signal is first converted into a voltage signal proportional to the total charge which is then digitised. In the previous sections of this chapter we have described the different processing steps separately, combined in such a unit (pre-amplification, shaping, ADC). In a *peak-sensing* ADC the peak voltage of a pulse is sensed (e.g. by methods employed in the *sample-and-hold* technique, section 17.4), held for some time, and then digitised.

Further down we discuss some digitisation techniques often used in detector readout chains. Before, however, some characteristics by which performance quality of ADCs can be assessed should addressed:

- resolution: precision of the code;
- integral linearity: proportionality of output to input;
- differential (non-)linearity: homogeneity of digitisation steps;
- conversion speed: time of converting an analog signal into a digital code;
- rate capability: how fast successive signals can occur and still can be correctly digitised;
- stability: sensitivity of conversion quality with respect to time, temperature change, and other external parameters.

Besides conversion speed and time stability, mainly conversion precision (ADC resolution) and (deviations from) linearity are important. The overview given in this section and the definitions are mainly applicable for the digitisation of a voltage being constant over the considered time interval. More advanced treatments and definitions can be found for example in [599].

17.7.1.1 ADC resolution

An 8-bit ADC subdivides the input voltage range to be digitised into $2^8 = 256$ possible codes, a 12-bit ADC into 4096 codes. The number of effective bits used in the conversion is called 'ADC resolution', specified by the number of bits used to represent

the analog value, in principle giving $2n$ signal levels for an n-bit signal. Which resolution is needed and meaningful in a given application depends on the signal-to-noise ratio (SNR) of the voltage to be digitised. If the noise of an input voltage is larger than 1 LSB an ADC with higher bit resolution would merely provide a more precise digitisation of noise contributions. The signal, however, will not be determined any better.

The quantisation error δ generated by the *width* of the code can be expressed as the standard deviation of a uniform distribution, computed as described in appendix E (eq. (E.4)):

$$\delta = \frac{\text{LSB}}{\sqrt{12}}\,, \tag{17.37}$$

where the code width is determined by the LSB, the smallest digital voltage step of the ADC[11]. For example, the quantisation error of a voltage in an 8-bit ADC given an input voltage range of 2.2. V is 2.5 mV, in a 12-bit ADC it is 0.16 mV.

The SNR in ADCs is defined as the ratio of the full input range to the quantisation error assuming the latter to be the only noise contribution. SNR is given in decibels (dB):

$$\text{SNR}_{\text{ADC}} = 20\log_{10}\,(2^n\,\sqrt{12}) \;\approx\; (6.02\,n \,+\, 10.8)\,\text{dB}\,, \tag{17.38}$$

here given for the digitisation of a DC voltage level with n bits. For signals varying in time the reference value for the comparison with DC levels is the effective voltage $V_{max}/\sqrt{2}$ of a sine wave (see [599]).

There are always fluctuations in the conversion process in every ADC. If they are large enough, the effective number of bits (ENOB) can be less than the actual number of bits. A definition of the ENOB can be obtained if the rms of these conversion fluctuations (rms_{conv}) is determined by precisely scanning the range of the input voltage and histogramming the resulting digital set-actual variances. The ENOB of the ADC reduces n to a value $m < n$ depending on the ratio of the measured rms_{conv} to the ideal bit resolution $(1/\sqrt{12})$ for the same SNR_{ADC} specified in (17.38):

$$m \;=\; \text{ENOB} = n - \log_2\left(\frac{\text{rms}_{\text{conv}}}{1/\sqrt{12}}\right)\,. \tag{17.39}$$

Often the desired voltage range to be digitised is at odds with the effort needed to cover this range with homogeneous ADC resolution. In most cases, however, the resolution does not need to be equally precise in all regions. For example, for the digitisation of signals of a wire chamber for a dE/dx measurement (see section 7.10.10 on page 246 and section 14.2.2) one can afford to measure dE/dx values, which are much larger than the maximum of the Landau distribution, with lower resolution than values near the maximum. It therefore would make sense to increase the dynamic range of an n-bit ADC to an effectively larger bit range $m > n$ by using a nonlinear characteristic, however coming at the price of a resolution loss at large input voltages [494]. This has been done for example with flash ADC applications (described in section 17.7.1.3) for pulse form digitisation of wire chamber pulses [394, 220]. The

[11] In more general definitions the code width can differ from the voltage value of the LSB [599], for instance when missing codes, that is, codes that do not appear in the output combination, shall also be accounted for in the definition.

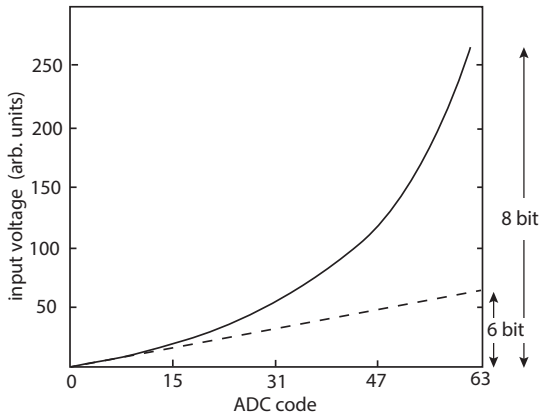


Fig. 17.30 Enlargement of the input range of an ADC by a nonlinear characteristic [220] obtained by successive increase of the input step size (see text).

characteristic was extended from originally 6 bit to effectively 8 bit by adding a fraction of the input voltage V_{in} to the reference voltage V_{ref}

$$D = D_{max} \frac{V_{in}}{V_{ref} + aV_{in}}, \tag{17.40}$$

where D is the digital code and D_{max} is its range (here 64 corresponding to 6 bit), thus increasing the voltage step size for every code step. To extend the dynamic range from 6 to 8 bit the fraction to be added must be $a = 0.75$. Electronically this can elegantly be realised by applying weighted inputs by means of a resistor chain [494]. Figure 17.30 shows the increased input voltage range as a function of the output code

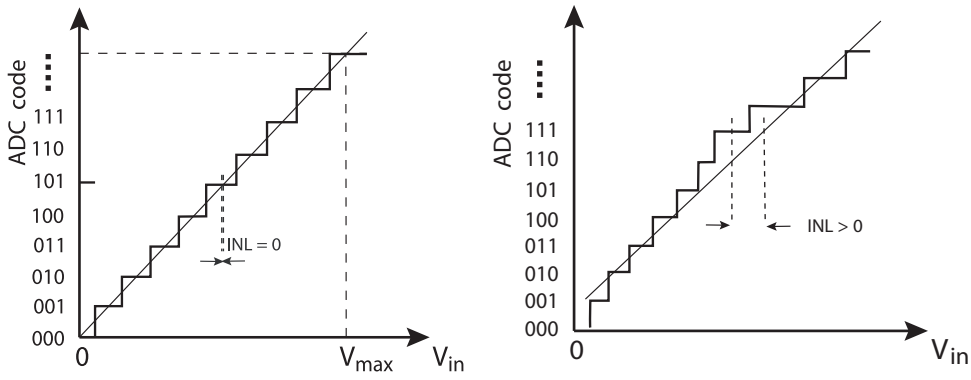
$$V_{in} = \frac{DV_{ref}}{64 - aD} \tag{17.41}$$

with the (arbitrary) choice of $V_{ref} = 64$ for illustration.

17.7.1.2 Linearity

The digitisation transfer function, which maps voltage values onto binary codes, has steps, as shown in fig. 17.31(a) for an assumed ideal ADC characteristic. For every step, each ranging from $\langle V_{in}\rangle - \frac{1}{2}$ LSB to $\langle V_{in}\rangle + \frac{1}{2}$ LSB, where $\langle V_{in}\rangle$ is the input signal averaged over the bin, the ideal characteristic passes through the steps' midpoint values. Perfect linearity also means that the midpoints form a straight line through the origin. Deviations from ideal linearity between voltage input and digital output are listed as *ADC errors*. Besides the *quantisation error* accounting for the deviation introduced by the digitisation steps, there are *zero-point errors* quantifying the deviation of the offset from zero (this is usually included in the quantisation error), and also the *slope error* $(s - s_{\text{ideal}})$, usually given as fractions of s. In addition there are *linearity errors*.

Deviations from the linear characteristic caused by ADC errors can be taken into account by proper calibration. Local errors can then be quantified as maximum deviations from the straight-line characteristic. Deviations from linearity are often characterised by two defined and measurable quantities, the integral nonlinearity (INL) and the differential nonlinearity (DNL) with the former (INL) describing the total deviation from the expected linear behaviour, while the latter (DNL) measures the deviation of



(a) Perfectly linear ADC. (b) Integral non-linearity.

Fig. 17.31 Integral non-linearity: (a) Linear ADC with an ideal characteristic calibrated such that the first code transition from 000 to 001 starts at $\frac{1}{2}$ LSB; (b) definition of the integral nonlinearity (INL), see also explanation in the text. Adapted from [725] with kind permission from the author A.S. Nastase, PhD.

local ADC codes from their expected values. Note, that (slightly) different definitions for DNL and INL exist in the literature.

Integral nonlinearity. The INL is a measure for the deviation of the ADC characteristic from a straight line. This line is often a best-fit line among the points in the plot but can also be a line that connects the highest and lowest data points, or the endpoints. The INL is defined as the maximal deviation of the measured midpoints of the quantisation steps from this line, expressed either as a fraction of the input voltage range (V_{max}) or in units of the voltage step which corresponds to the lowest-valued bit (LSB). The INL value is measured from a step's midpoint to the point where the horizontal extension of the step crosses the straight line[12] (fig. 17.31(b)). An ideal ADC has INL = 0 in this definition.[13]

Differential nonlinearity. The DNL is a measure of the homogeneity of subsequent digitisations (hence 'differential'): the midpoints of the voltages of subsequent codes should ideally be separated by exactly 1 LSB. The DNL describes local relative deviations of ADC codings from the (ideal) step width (1 LSB) and is defined as

$$\mathrm{DNL_i} = \frac{V_{i+1} - V_i}{\mathrm{LSB}} - 1 \quad \text{for} \quad i = 0,\, 1,\, 2,\, \ldots,\, 2^n - 2\,, \tag{17.42}$$

where V_{i+1} and V_i are consecutive voltage steps. The last of the $2^n - 1$ values is exempted from the definition, since usually it is used as overflow for input signal values lying outside of the definition range. Figure 17.32(a) shows how DNL errors can occur. The DNL is specified as a fraction or a multiple of ± 1 LSB. DNL errors larger than -1 LSB are required to guarantee no 'missing codes'.

Figures 17.32(c) and (d) show two examples for DNL errors of $+1$ and -1 LSB, respectively, while fractional DNLs are illustrated in fig. 17.32(a). The DNL errors of the entire voltage range can be obtained by a measurement as shown by fig. 17.32(b).

[12] At this point the code considered here ought to correspond to the correct input voltage (assuming a linear characteristic).

[13] Other INL definitions exist for which an ideal ADC has INL = 1/2 LSB (see e.g. [686]).

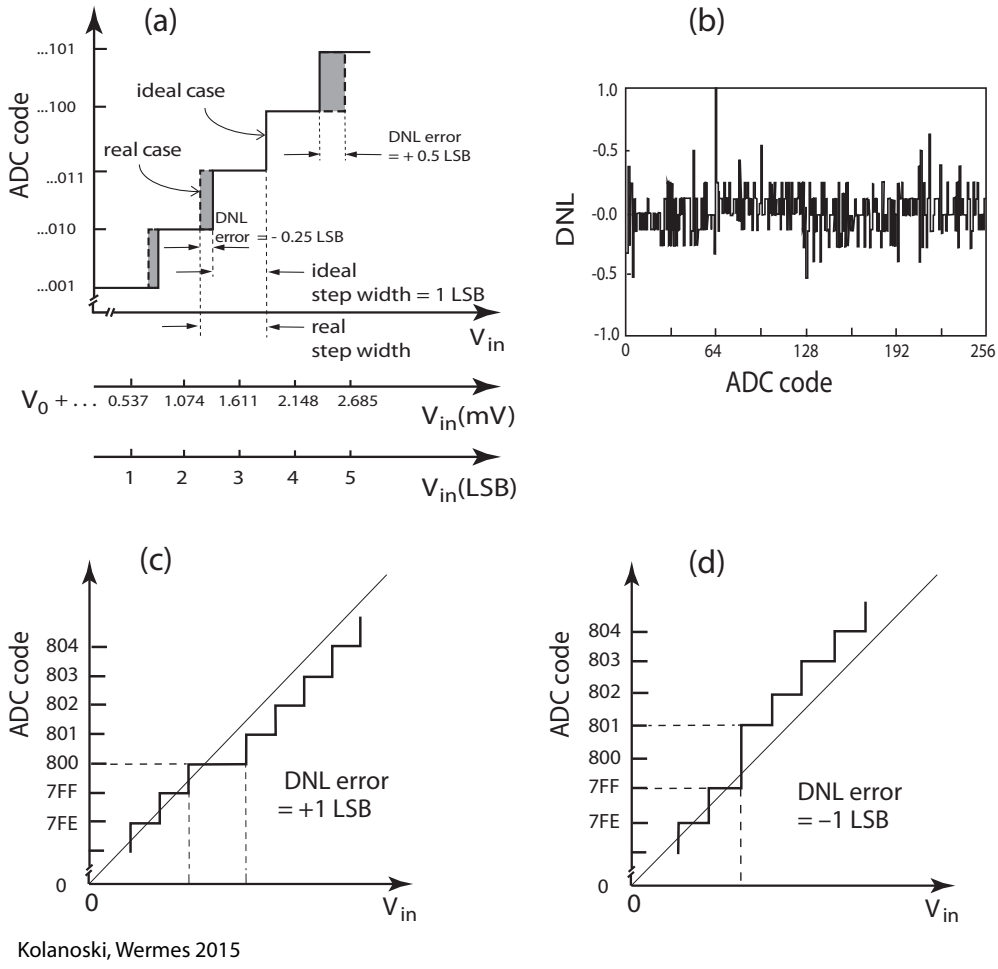


Fig. 17.32 Differential nonlinearity (DNL). (a) On the definition of DNL: solid and dashed lines represent ideal and realistic example responses of an ADC, respectively. The x-axis is given in mV and in LSB units. (b) Plot of 256 measured DNL errors for an 8-bit flash ADC (adapted from [328]); (c,d) show ADC code suppressions (compare eq. (17.42)), manifesting themselves as (+1 LSB) DNL error (code change only after 2 LSB) or as (−1 LSB) DNL error (code missing). The y-axis is represented in hexadecimal coding. Adapted from [725] with kind permission from the author A.S. Nastase, PhD.

The input voltage is increased linearly and the DNL_i are plotted as a function of the ADC code. In order to specify an overall DNL value characterising the ADC, either the maximum absolute value or the rms of all DNL_i is quoted:

$$\mathrm{DNL} := \max_i(|\mathrm{DNL}_i|) \qquad \text{or} \qquad \mathrm{DNL}_{rms} := \left(\frac{1}{2^n - 2} \sum_{i=0}^{2^N-2} (DNL_i)^2 \right)^{1/2} .$$

For the inverse conversion of a binary number into a voltage by a digital-to-analog converter (DAC), corresponding definitions of INL and DNL hold (see section 17.7.2).

17.7.1.3 ADC concepts

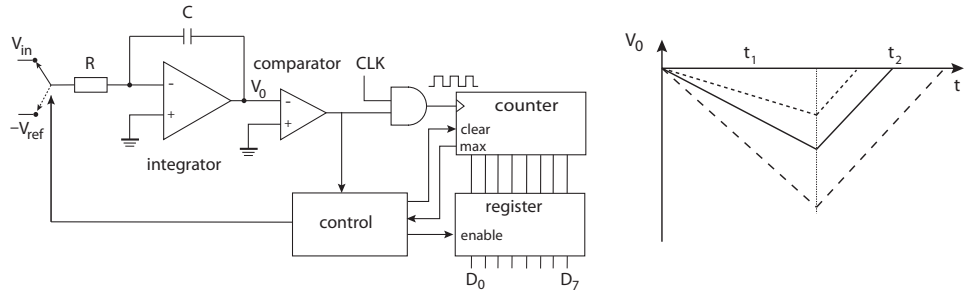
Electronic conversion of a voltage into a number can be realised in different ways. Choosing one or another depends on the requirements regarding conversion speed, linearity, resolution required, and the respective necessary circuit complexity, in particular since nowadays ADCs are usually realised as integrated circuits. We present here three conceptually different realisations which find frequent use in detector readout systems. Hereby we restrict our description mainly on the digitisation of an input voltage which is constant during the conversion time. To digitise a complete pulse shape the voltages of consecutive time intervals are successively made available at the ADC input, for example by employing a sample-and-hold technique (section 17.4), and are digitised one after another. For a detailed discussion of A/D conversion we refer to the comprehensive literature, for example in [768, 701, 599, 540].

Successive-approximation ADC. A digitisation method often employed is the *successive approximation register* or SAR. The input voltage (e.g. 0.6 V) is successively approximated by comparing to digitally selectable reference voltages. Starting with the midpoint voltage $V_{1/2}$ of a reference range (e.g. 0–1 V) the most significant bit (MSB) is set to 1 or 0 depending on whether the input voltage is larger (1) or smaller (0) than $V_{1/2}$ (here 0.5 V). To determine the value of the next to most significant bit (MSB-1) the reference voltage in the next step is chosen as $V_{1/2} \pm V_{1/4}$ with $V_{1/4} = \frac{1}{2}V_{1/2}$ where $+V_{1/4}$ is selected for MSB = 1 and $-V_{1/4}$ for MSB = 0 (in the example this is $V_{1/2} + V_{1/4} = 0.75\,\mathrm{V}$) with $V_{1/4} = \frac{1}{2}V_{1/2}$. In the next step the reference voltage is increased or decreased by $\pm V_{1/8}$ (here to 0.675 V). In the example so far we then have a bit sequence of 100. The process is continued until the chosen digital resolution is achieved. The needed selectable reference voltages are supplied by means of a DAC (see section 17.7.2). The SAR ADC represents a good compromise between speed and circuit effort. Compared with other ADC realisations the achievable precision with a SAR lies in mid range. The needed chip area grows linearly with the required resolution.

Wilkinson ADC. The class of so-called integrating or *Wilkinson ADCs* is based on comparison of the input voltage with the voltage produced by a capacitor which is being charged. The capacitor is allowed to charge until its voltage is equal to the amplitude of the input pulse and is then linearly discharged again. The duration of the discharge time is hence directly proportional to the input amplitude, measurable by clock pulses.

We describe the functional principle in fig. 17.33 using a variant with a linear ramp for charging and discharging (*dual-slope* ADC), a follow-up development of the original Wilkinson concept [993]. ADCs of the Wilkinson type can be realised with high resolution and medium circuit effort. The chip area needed does not increase much with increasing resolution. The advantage of the dual-slope method lies in its independence on fabrication variations of electronic components and in its comparatively good noise immunity. They are, however, slow in comparison to other ADCs.

Flash ADC. A conceptually very simple technique with high conversion rate is *flash* conversion. The voltage to be digitised is applied at the parallel inputs of comparators with different threshold voltages between the minimum and the maximum of the input voltage range. In the simplest case a linear division by a voltage divider is chosen. The principle is illustrated in fig. 17.34(a). An n-bit flash



(a) Block diagram of the principle. (b) Time development of V_0.

Fig. 17.33 Dual-slope ADC. (a) Block diagram of the function principle, (b) time development of the voltage V_0 at the output of the integrator. The circuit controls a counter by a linear charging and discharging of the feedback capacitor of an operational amplifier integrating the DC input voltage. The capacitor is charged for a fixed time t_1, defined for example by the time until the counter reaches its maximum value (*max*), after which the counter is reset and the input is switched to $-V_{\mathrm{ref}}$. This is the voltage which determines the discharging part in slope and length, while the slope of the charging part depends on V_{in}. Discharging proceeds with constant slope (set by V_{ref}) but with variable duration t_2 until V_0 reaches zero. The duration of t_2 depends on the height of V_0 reached after the end of the charging process. Hence the counter value at t_2 encodes V_{in} and is transferred into the register.

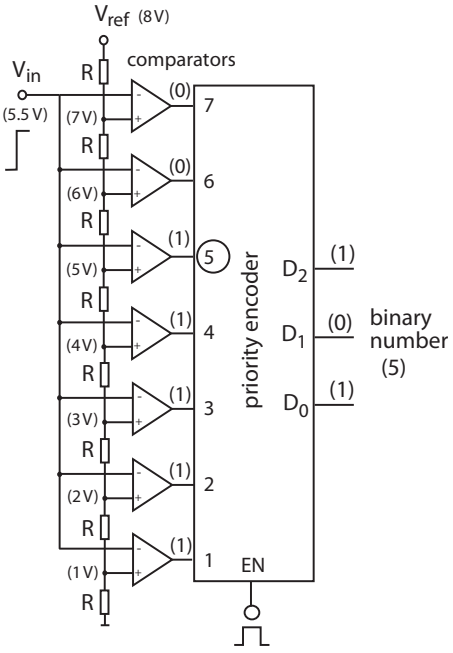
ADC needs $2^n - 1$ comparators. The parallelisation of many input stages is the reason for high conversion speed. Principle disadvantages are high circuit complexity and comparatively large power consumption. This is the reason that only with the advancement of fast integrated electronics have flash ADCs become a competitive digitisation technique. The resolution of current flash ADCs is limited to between 10 and 12 bit. The required chip area increases exponentially with the resolution. Modern realisations reach Gsps (giga samples per second) conversion rates at resolutions of 12 bit. Fast flash ADCs allow time-wise sampling of the pulse shape of an analog detector pulse with a typical duration of several hundred nanoseconds. The pulse is digitised in time windows, as shown in fig. 17.34(b).

Pipeline ADC. An ADC architecture often used in detector readout electronics is the *pipeline architecture* operating in several steps, which are usually realised as flash ADCs. In every stage a coarse (e.g. 3 bit) fast conversion is performed, the result of which is converted back (via a DAC) into an analog signal and subtracted from the temporarily buffered input signal. The difference is amplified to the full ADC range and then converted finer in the next stage. The results are combined in a last step.

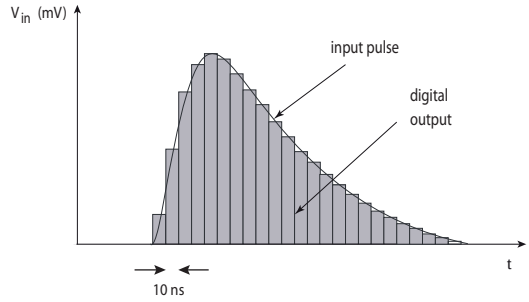
Table 17.1 lists the characteristics of the mentioned ADC types which are often important for a choice in a given application. Besides conversion speed and resolution often power consumption and, for usage in a chip, the needed chip area are the most important criteria.

17.7.2 Digital-to-analog conversion (D/A)

The reverse of A/D conversion is D/A conversion performed in corresponding digital-to-analog converters (DACs). A DAC converts a digital code at its input into an analog



(a) Flash ADC block diagram.



(b) Digitisation of a pulse.

Fig. 17.34 Flash ADC: (a) schematic diagram explaining the function principle (here 3 bit FADC). The threshold voltages of the comparators are increased from bottom to top. All comparators having a threshold smaller than the input voltage deliver a logic 1 at the output, those with a higher threshold than the input a logic 0. A concrete example is illustrated by the numbers in parenthesis. A *priority encoder* initialised by the enable input EN finds the highest comparator (comparator no. 5 in the example, encircled) and delivers a binary number at the data outputs D_0, D_1, D_2. The input voltage of 5.5 V of the example is encoded as a '5'. (b) Completely digitised pulse shape by means of a fast (100 MHz) flash ADC.

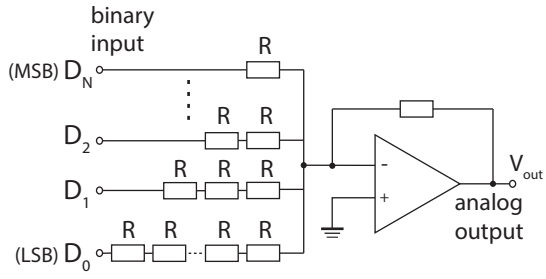
Table 17.1 Characteristics of some ADCs often used in detector readout (various sources, 2018).

ADC type	Rel. speed (samples per s)	Resolution (bits)	Chip area	Power consumption
SAR	slow–medium (< 2 Ms/s)	8–16	small	low
Dual-slope	slow (< 100 ks/s)	12–20	medium	low
Flash	very fast (< 5 Gs/s)	4–12	large	high
Pipeline	fast (< 500 Ms/s)	8–16	medium	medium

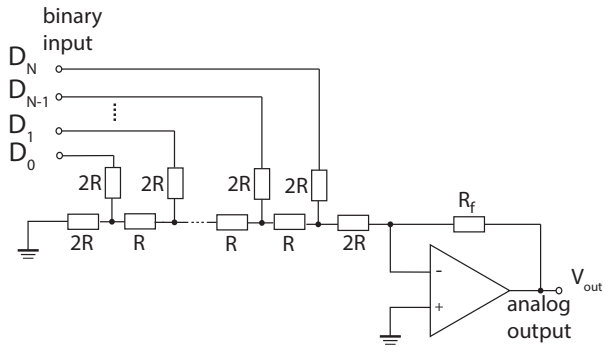
quantity, usually a voltage. DACs are essential for the communication direction from computer to electronics. An example is setting the reference voltage of a discriminator to which signal amplitudes are compared (see e.g. figs. 17.25 and 17.26).

DACs are characterised similarly to ADCs by resolution, linearity, and conversion speed. Integral (INL) and differential (DNL) nonlinearity are also defined as for ADCs (section 17.7.1.2, see also [599]).

A simple DAC realisation is a resistive network (fig. 17.35(a)). The output signal



(a) Resistive ladder.



(b) R-2R ladder.

Fig. 17.35 D/A conversion by resistive networks. (a) Generation of the output voltage by (binary) weighted current contributions; (b) R-2R conductor network using only two different resistor values. $V_{out} = 0$ if all inputs D_i are '0'(grounded). If one input is '1' (on voltage V_{ref}) the resistance to ground seen by the current at the corresponding node in the resistor chain (black dot) always is 2R to the left and 2R to the right (virtual ground of the OpAmp), such that the current is halved at this node and is halved again at the next node and so on. If more than one D_i is '1' the corresponding currents are superimposed. The input to the OpAmp and correspondingly the output voltage is determined by the total current receiving appropriately 'weighted' contributions from every $D_i =$ '1'.

is created by as many resistors as there are binary digits; every resistor chain has a weight corresponding to its assigned digit. The currents generated by the network are converted into a voltage by a current-to-voltage converter. The disadvantage of this construction is the need for many different resistors. Simpler in realisation is the equivalent R-2R network (fig. 17.35(b)) based on halving of the current at every node in a resistive divider chain (see caption of fig. 17.35).

17.7.3 Time-to-digital conversion (TDC)

Measurement and digitisation of a time is also frequently required in experiments. A very direct application example is the time-of-flight (TOF) measurement of particles, as shown in fig. 14.3 on page 548. The (digital) time measurement features a fast clock in combination with a counter counting the number of clock pulses between a start and a stop signal (fig. 17.36). The achievable time resolution is determined by the speeds of clock and counter, respectively, typically in the order of 10 GHz or more, allowing time resolutions in the order of less than 100 ps.

High resolutions can also be obtained by *time-to-amplitude conversion* (TAC) with a subsequent ADC for digitisation, as shown in fig. 17.37. A capacitor of capacitance C is charged by a constant current I for a time $\Delta t = t_{\mathrm{start}} - t_{\mathrm{stop}}$ such that a voltage

$$V = \frac{Q}{C} = (t_{\mathrm{start}} - t_{\mathrm{stop}})\,\frac{I}{C}$$

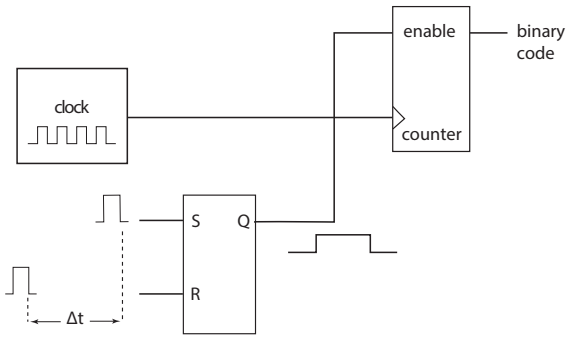


Fig. 17.36 Realisation of a time-to-digital converter (TDC) by a clock and a counter which is started and stopped by external signals. The SR flip-flop generates a gate pulse with length Δt (start minus stop).

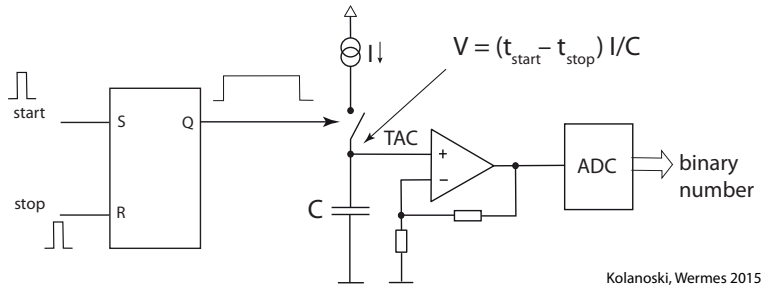


Fig. 17.37 Realisation of a *time-to-digital* converter (TDC) by a *time-to-amplitude* converter (TAC) with subsequent voltage amplifier and ADC. At the input an SR flip-flop generates a pulse of length Δt (start minus stop).

is applied at the input of an ADC for digitisation. This arrangement allows time resolutions in the picosecond range to be obtained.

17.7.4 Logic level schemes

For the interpretation of digital electrical signals, for example during transport to the final readout or for conversion at electrical or electro-optical transmission interfaces (see section 17.8), logic levels must be defined. When is a voltage level interpreted as a logic '1' and when as a '0'? The definition of logic standards has existed since digital information was first used during and after World War II. With the arrival of semiconductor-based electronics, bipolar junction and field effect transistors and integrated circuits in the early 1960s, logic levels were first designed using circuits containing transistors and resistors (RTL logic), replaced by DTL (diode–transistor logic) and TTL (transistor–transistor logic) a few years later. With the arrival of MOS transistors and the CMOS technology (patented 1963 [985], see also section 8.3.5 on page 297), integrated circuit logic became overwhelmingly dominated by CMOS logic.

17.7.4.1 Logic circuit standards

TTL. Transistor–transistor logic without passive elements consumed much less power then RTL and DTL and remained the basis for bipolar ICs.[14] TTL logic is

[14]The famous 74xx IC series by Texas Instruments is a particular example for the application of TTL logic.

defined by voltage with levels of 0–0.8 V being interpreted as 'logic 0', levels of 2–5 V as 'logic 1' (see table 17.2). Typical TTL switching times are 40 ns, in special high-speed versions 5–10 ns [540].

ECL. Faster still, but very power-hungry (30–50 mW/gate) is ECL (emitter-coupled logic, current steered) with switching times in the nanosecond range and even below[15]. For reasons of its speed ECL has been popular for high performance computing and also for fast logic modules and systems in high energy physics like NIM or CAMAC (see section 18.2.1).

CMOS. With the advancement of CMOS integrated circuits in the 1980s, using PMOS and NMOS transistors on the same silicon substrate (see section 8.3.5 on page 297) and offering high circuit density and low power per gate, CMOS logic [540] has become the general standard for most integrated circuits today. The logic levels are 3.5–5 V ('logic 1') and 0–1.5 V ('logic 0'), respectively, for 5 V supply voltage, and typically about 0.9 V_{DD} and about 0.1 V_{DD} (V_{DD} = upper operation voltage) for modern low-voltage (LVCMOS) technologies operating at much lower supply voltage (e.g. 1.0–1.2 V in technologies with 90 nm feature size or lower[16]). In static state, CMOS gates do not consume power (see also section 8.3.5 on page 297); typical quiescent currents are below 1 µA. For modern LVCMOS technologies, this statement is not entirely true, due to sub-threshold leakage which starts to play a role. Special design techniques need to be invoked to mitigate excessive static power consumption in multi-million logic gate devices like CPUs or gate arrays (e.g. FPGAs). Dynamically, CMOS logic power is proportional to the switching frequency and can also become large at very high frequencies.

17.7.4.2 Signal transmission standards

For fast signal transmission and signal communication between different electronic modules additional standards are in use which should be distinguished from circuit standards. The transmission standards are usually defined through currents that lead to voltage levels across a termination resistor.

NIM. Already in the 1960s nuclear and particle physics experiments needed to define a standard for electronic data processing and triggering, including a crate system with plugged-in modules for various electronic tasks, as well as cables and connectors. The NIM standard was updated in 1990 [737]. A description is given in sections 18.2.1 and 18.4.2 of chapter 18. The original NIM cable standard is a coaxial cable with a characteristic wave impedance of 50 Ω; meanwhile other cable types are also permitted, in particular those that allow for a higher packing density (see also section 17.8.1). The electronic levels (NIM levels: nominal '0' = 0 V, '1' = −0.8 V) are defined as currents leading to (negative) voltages over a 50 Ω termination resistance, as shown in table 17.2. Besides NIM levels, other standardised logical levels may also be accepted in a NIM system (e.g. TTL and ECL).

[15] Submicron technology has brought ECL power consumption substantially down, close to CMOS levels [1024].

[16] The feature size of a semiconductor technology is defined as the minimum length of the MOS transistor channel between drain and source. The feature size has been shrinking year by year. From the early 2000s it has shrunk from 180 nm to 7 nm designs today (2018).

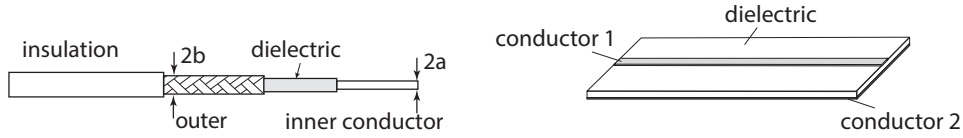
Table 17.2 Logic levels for some electronic standards (circuits and signal transmission) commonly used in detector readout architectures. The logic definition can be defined either by voltage ranges (column 'V') or by a defining current (column 'I'). The values shown are approximate. Positive or negative voltages can be assigned to logic '1'. In general, the allowed ranges for what 'the output must deliver' and what 'the input must accept' differ. The second column gives typical supply voltages, P_g is the power consumed per logical gate and ν_{max} corresponds to the fastest switching time achievable. For detailed specifications the referenced literature or others should be consulted (e.g. [540]). The data for the 65 nm CMOS node are given as a currently typical example for VLSI chips.

Stand-ard	Supply (V)	Logic levels	I (mA)	V (V)	ν_{max} (MHz)	P_g [(1)] (mW)	Ref.
electronic circuit standards							
TTL	5	'1' '0'		2–5 0–0.8	25–100	2–8	[540]
ECL	−5.2	'1' '0'		−0.9 (nom.) −1.8 (nom.)	100–500	25–60	[540]
CMOS	5	'1' '0'		3.5–5 0–1.5	5–100	0.5–1	[540]
CMOS 65 nm	1.2	'1' '0'		1–1.2 0–0.12	250	0.004	[984, 688]
signal transmission standards							
NIM[(2)]	6	'1' '0'	−12 to −36 −4 to +20	−0.6 to −1.8[(3)] −0.2 to +1	300		[737]
LVDS	2.5–3.3	'1' '0'	±3.5[(4)]	±0.350[(5)]	> GHz		[596, 549] [390]

(1) At 1 MHz, depends on (is approximately proportional to) switching frequency.
(2) For fast NIM-ECL logic.
(3) Across 50 Ω (values that inputs must accept); nominal: '0' = 0 V, '1' = −0.8 V.
(4) Definition by current direction causing a voltage swing of 350 mV across ≈ 100 Ω.
(5) A constant current of 3.5 mA across ≈ 100 Ω.

LVDS. With the need for high bandwidth, high-srpeed data communication (> 100 Mbit/s) differential transmission using signals with the same height but opposite polarities on (twisted) cable pairs, the LVDS (low voltage differential signalling) standard was defined in 1994. It does not depend on a specific supply voltage. The LVDS transmitter injects a constant current of 3.5 mA into the wires, which have a typical characteristic (differential[17]) impedance of between 100 Ω and 120 Ω. When passing through the termination resistor and returning in the other wire the current causes a voltage difference of about 350 mV across the resistor with polarity depending on the current direction. The polarity of this voltage is sensed by the receiver and determines the logic level. Differential LVDS transmission is fairly immune to noise (see also sec-

[17] Between the two lines.



(a) Coaxial cable transmission line. (b) Microstrip line.

Fig. 17.38 Signal transmission lines: (a) Coaxial cable with inner and outer conductor, (b) microstrip line, version with full-faced conductor sheet on backside.

tion 17.8) and can be used at data rates of several Gbit/s, consuming comparatively little power: at 1.2 V the power to drive 3.5 mA is only 4.2 mW.

17.8 Signal transport through transmission lines

17.8.1 Signal cables in HEP experiments

Analog or digital detector readout signals are usually transported over (longer) distances by means of transmission lines as wave guides, often realised as coaxial cables or (micro)-strip lines. The signal propagates in between two conduction line elements separated by a dielectric. Figure 17.38 shows wave guides realised in coaxial or strip-line configuration, respectively.

The geometry of the wave guide determines the capacitive and inductive (also resistive) conduction elements. In the case of a coaxial cable these are (per unit length):

$$L' \approx \frac{\mu\mu_0}{2\pi} \ln\frac{b}{a} \qquad C' \approx \frac{2\pi\epsilon\epsilon_0}{\ln(b/a)}\,, \tag{17.43}$$

where a and b are the radii of the inner and outer conductor, respectively, and $\mu\mu_0$ and $\epsilon\epsilon_0$ are relative and absolute permeability and permittivity of the dielectric, respectively.

In a real cable there are losses to be taken into account, described by resistance elements. Typical values for inductance and capacitance of a coaxial cable (e.g. RG-58/U)[18] with a characteristic impedance of 50 Ω (definition in eq. (17.55)) are about 250 nH/m and 100 pF/m. RG-58/U is used frequently in nuclear and particle physics for the connection of electronics modules. A (push/pull) connector standard is the so-called 'Lemo^{TM} connector' with associated cable of about 5 mm diameter. The characteristic impedance of connector and cable is 50 Ω. They have largely replaced the former BNC standard (connector and cable) which has larger dimensions (a typical cable diameter is 8 mm) and comes with characteristic impedances of 50 Ω or 75 Ω.

A similar electrical specification is obtained from a microstrip transmission line (single line as in fig. 17.38(b)) with 200 µm width and 20 µm thickness on Kapton[19] (thickness 100 µm, $\epsilon = 3$) with metallised backside [693].

In today's experiments in many cases *twisted-pair cables* are also used, which have their predominant usage in data or computer networks. Twisted-pair (TWP) cables are twisted conductor leads functioning as a twin conductor. They are a cheap and space-saving alternative to coaxial cables. Signals are transmitted with equal height

[18] RG denotes a *radio guide* standard corresponding to the military standard MIL-C-17. U means *universal* and the number is the fabrication number.

[19] Kapton is a trade name for polyimide of the company DuPont.

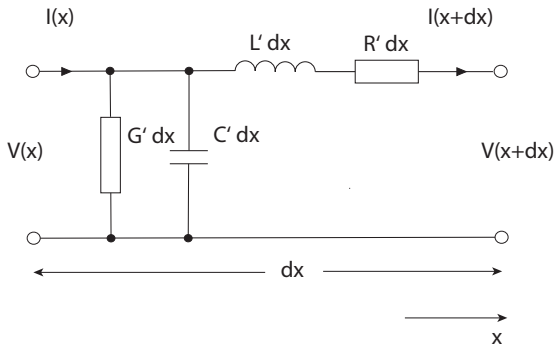


Fig. 17.39 Equivalent network circuit for an infinitesimal element dx of a wave guide.

but opposite polarity on either lead and are read out by a differential amplifier. By differential readout the opposite sign signals add, whereas external electromagnetic pickup, having the same polarity on both leads, is suppressed. Twisting of the leads additionally suppresses all kinds of electromagnetic interference which otherwise (untwisted) would occur due to the different distances between the wires and the source of the disturbance. The individual pairs and/or several pairs merged in a cable are often additionally shielded by foil wrap. Also very frequently several readout channels each having a TWP cable are arranged together in a plane as a ribbon cable. The characteristic impedance of TWP cables is typically 100 Ω.

17.8.2 Signal transmission equation

Transport of signals through transmission lines follows the telegraph equation (see e.g. [634]) which can be deduced from Kirchhoff's current and voltage laws

$$\frac{\partial V}{\partial x} = -L'\frac{\partial I}{\partial t} - R'I\,, \tag{17.44}$$

$$\frac{\partial I}{\partial x} = -C'\frac{\partial V}{\partial t} - G'V \tag{17.45}$$

applied to the transmission line equivalent circuit in fig. 17.39. Here L', C', R' and G' are the inductance, capacitance, resistance and conductance (reciprocal resistance) per unit length, respectively. Applying $\partial/\partial x$ to (17.44) and $\partial/\partial t$ to (17.45) yields the one-dimensional wave equations (telegraph equations) for voltage and current propagation on a wave guide:

$$\frac{\partial^2 V}{\partial x^2} - L'C'\frac{\partial^2 V}{\partial t^2} = (L'G' + R'C')\frac{\partial V}{\partial t} + R'G'V\,, \tag{17.46}$$

$$\frac{\partial^2 I}{\partial x^2} - L'C'\frac{\partial^2 I}{\partial t^2} = (L'G' + R'C')\frac{\partial I}{\partial t} + R'G'I\,, \tag{17.47}$$

with the solution for (17.46)

$$V(x,t) = V_0\,\mathrm{e}^{i\omega t - \gamma x} = V_0\,\mathrm{e}^{\,i(\omega t \mp \beta x)}\,\mathrm{e}^{\mp\alpha x}\,, \tag{17.48}$$

with

$$\gamma = \pm\sqrt{(R' + i\omega L')(G' + i\omega C')} = \pm(\alpha + i\beta)\,. \tag{17.49}$$

The real part $Re\,\gamma = \alpha$ describes attenuation during signal propagation.

For the case without losses ($R' = G' = 0$) we obtain the homogeneous wave equations

$$\frac{\partial^2 V}{\partial x^2} - \frac{1}{c_{ph}^2}\frac{\partial^2 V}{\partial t^2} = 0\,, \tag{17.50}$$

$$\frac{\partial^2 I}{\partial x^2} - \frac{1}{c_{ph}^2}\frac{\partial^2 I}{\partial t^2} = 0\,, \tag{17.51}$$

with phase propagation velocity

$$c_{ph} = \frac{1}{\sqrt{L'C'}} = \frac{1}{\sqrt{\epsilon\epsilon_0\mu\mu_0}} \tag{17.52}$$

and $\gamma = \pm i\beta = \pm\sqrt{-\omega^2 L'C'}$. The general solution of both (17.50) and (17.51) is a superposition of a forward and a backward propagating wave:

$$\begin{aligned} V(x,t) &= V_{01}\,\mathrm{e}^{i(\omega t-\beta x)} + V_{02}\,\mathrm{e}^{\,i(\omega t+\beta x)} &&= V_{forw} + V_{back} \\ I(x,t) &= I_{01}\,\mathrm{e}^{i(\omega t-\beta x)} - I_{02}\,\mathrm{e}^{\,i(\omega t+\beta x)} &&= I_{forw} + I_{back}\,. \end{aligned} \tag{17.53}$$

The current amplitudes can be expressed as

$$I_{01} = \frac{V_{01}}{Z_0} \qquad\qquad I_{02} = \frac{V_{02}}{Z_0} \tag{17.54}$$

where

$$Z_0 = \frac{V}{I} = \sqrt{\frac{L'}{C'}} \tag{17.55}$$

defines the characteristic wave impedance of the line (unit = Ω).

17.8.3 Signal reflections

After processing by the so-called 'front-end' electronics (typically containing preamplifier, shaper, discriminator etc.) further signal transport takes place using transmission lines to the 'back-end' electronics, where for example logic processing of signals is performed. In signal transmission, disturbances and in particular reflections of signals at the ends of the transmission line must be avoided. The ratio of voltage and current of the signal defines the characteristic wave impedance Z_0. If a resistance discontinuity occurs at the line end, for example if the termination resistance is different from Z_0, another voltage-to-current ratio is adjusted such that a backward moving signal wave is caused which appears as a reflection. Therefore the termination resistance R_A at the line end (fig. 17.40(a)) determines how forward and backward propagating waves are superimposed.

With (17.53) we obtain for $t = 0$ assuming for simplicity that the line end is at $x = 0$:

$$V(x=0) = V_{01} + V_{02} =: V_A\,, \tag{17.56}$$

$$I(x=0) = \frac{V_{01}}{Z_0} - \frac{V_{02}}{Z_0} = \frac{V_A}{Z_A} = I_A \tag{17.57}$$

where $Z_A = V_A/I_A$ is the termination impedance at the line end. By adding, respectively subtracting, both equations one obtains

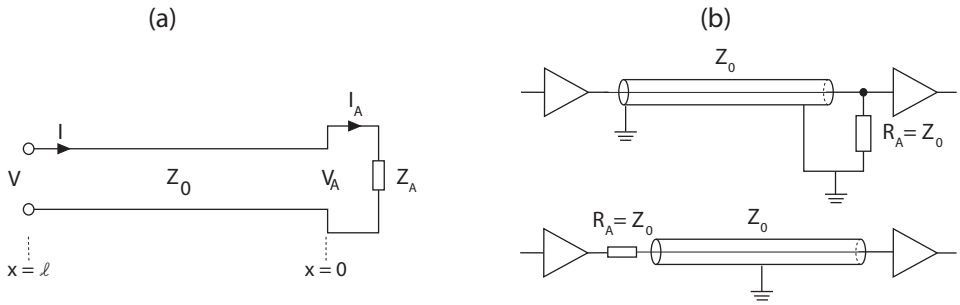


Fig. 17.40 Termination of a signal line: (a) principle; (b) examples with termination by a resistor $R_A = Z_0$ at the receiver side and at the sender side, respectively (see text, adapted from [905], with kind permission by Springer Science+Business Media).

$$V_{01} = \frac{V_A}{2}\left(1 + \frac{Z_0}{Z_A}\right), \qquad V_{02} = \frac{V_A}{2}\left(1 - \frac{Z_0}{Z_A}\right) \tag{17.58}$$

and the relation between backward to forward running wave defines the reflection factor:

$$r = \frac{V_{02}}{V_{01}} = \frac{Z_A - Z_0}{Z_A + Z_0}\,. \tag{17.59}$$

In the case where the line end is open ($Z_A = \infty$), that is, $r = 1$, the forward propagating signal is reflected unchanged. If the line end is shorted ($Z_A = 0$), that is, $r = -1$, the forward propagating signal is reflected with inverted amplitude. In between these extremes partial reflection occurs. Only for a suitable termination $Z_A = Z_0$ are there no reflections ($r = 0$). In experiments one must therefore pay attention to use correct line termination.

Figure 17.40(b) shows termination methods at the receiver end (top part of figure) by an ohmic impedance $Z_A = R_A$ to ground, parallel to the receiver, and at the sender end (bottom part of figure) with R_A in series. Since in most applications driver and receiver amplifiers (drawn as triangles in fig. 17.40(b)) are voltage amplifiers having high input and low output impedances, these configurations (serial at the sender or parallel at the receiver) each yield about R_A as total impedance. At the sender end, R_A and the cable impedance Z_0 form a voltage divider such that for $R_A = Z_0$ the amplitude of a pulse leaving the sender is halved. At an unterminated receiver end, however, this pulse will be reflected with equal polarity such that the superposition of both pulses yields the original amplitude. No further reflections occur since the reflected pulse sees R_A at the sender end.

Typical characteristic wave impedances from 50 Ω to 75 Ω for coaxial cables and 100 Ω for ribbon cables. The transmission velocity

$$c_{ph} = \frac{1}{\sqrt{L'C'}} \approx \frac{c}{\sqrt{\epsilon}}$$

where $\mu \approx 1$ usually holds, corresponds to a typical signal velocity c_{ph} of about 5 ns/m for coaxial cables, about 4 ns/m for ribbon cables.

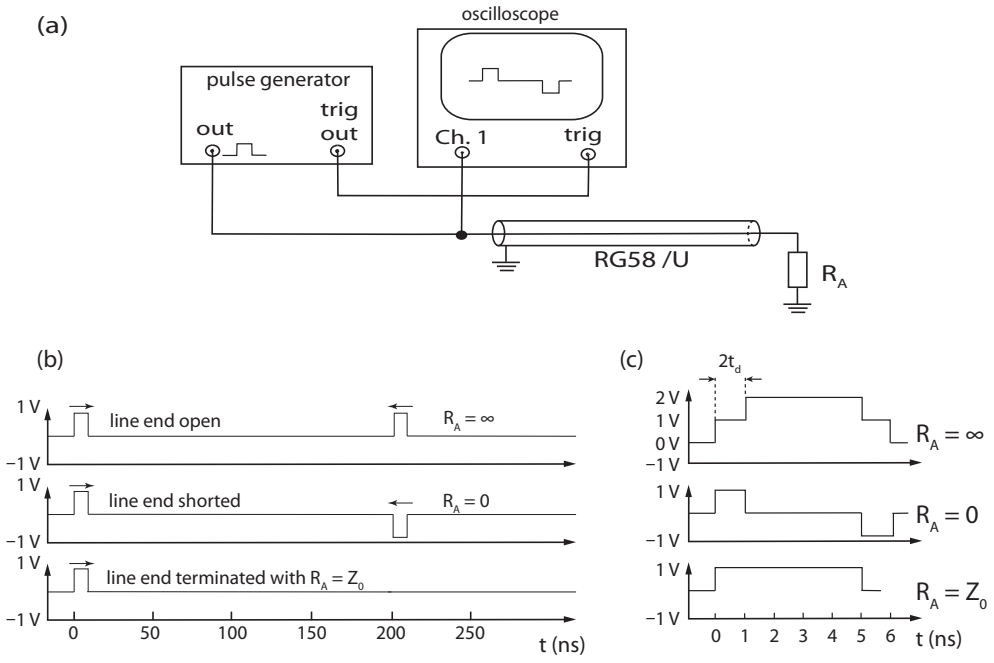


Fig. 17.41 Reflections at the cable line end of a 5 ns long signal pulse. (a) Measurement set-up. (b) Reflections in a 20 m long coaxial cable causing a time delay of $t_d = 100$ ns; (c) reflections occurring in a 10 cm long cable with $t_d = 0.5$ ns, where the reflected pulse, being broader than $2t_d$, is delayed with respect to the original pulse by the time of forward plus backward propagation through the cable (1 ns) and is superimposed on the input pulse. The measurement can for example be done with an oscilloscope at the beginning of the transmission line. The reflected pulse hence appears time shifted by $2t_d$ to the right of the input pulse.

Figure 17.41 shows effects that occur for two different relations of pulse length and cable length. Depending on the termination, pulse reflection can occur at the cable end, which leads to a superposition of original and reflected pulse. Since the output of the sending (voltage) amplifier usually has a low output impedance ($R \approx 0$), a pulse, possibly reflected at the cable end, is reflected again (inverted) and so on. Proper readout is therefore disturbed since signal pulses will be recorded multiple times. Using correct cable termination therefore is an important requirement of detector readout. Most frequently termination at the receiver end by means of a parallel resistor (parallel to the receiver input) is used. Since termination is never perfect, in critical applications one sometimes finds additional serial termination at the sender end (fig. 17.41(b)). This, however, causes a loss in pulse height by 50% which would no longer be compensated, as for the single-end terminated case discussed above (only if $Z_A = \infty$ at the receiver end).

17.8.4 Cable damping

During transmission of signal pulses over longer distances signal attenuation can no longer be neglected. The damping term $\exp(\mp\alpha x)$ in (17.48) takes account of attenu-

ation effects, where the minus sign is assigned to the forward propagating wave, that is, the direction of signal propagation. Typical values for α lie in the order of a few dB/100 m at 100 MHz and 10–100 dB/100 m at 1 GHz, such that attenuation starts to become a problem at cable lengths of typically about 10 m and more and is of course cable type dependent as α depends on frequency.

The skin effect changes R' in a frequency-dependent way and also G' is changed by frequency dependent dielectric losses. The latter can be neglected at frequencies below about 1 GHz compared to the influence of the skin effect, whereas above 1 GHz they become dominant. At high frequencies, when $R/\omega L \ll 1$ and $G/\omega C \ll 1$ is valid, one can express α by [652]

$$\alpha(f) = a\sqrt{f} + b\,f \qquad (f = \omega/2\pi)\,, \tag{17.60}$$

where the terms proportional to $\sqrt{f}$ and f describe skin effect and dielectric losses, respectively, and a and b are parameters depending on the cable type.

The frequency dependence of the attenuation leads to a change in the signal shape (dispersion). A step function pulse jumping from 0 to v_0 at the input of a transmission line of length l, is deformed as (see [404])

$$v(t) = v_0 \,\mathrm{erfc}\left(\frac{1}{2\sqrt{\tau/\tau_0}}\right)\,. \tag{17.61}$$

The complementary error function erfc has already been defined in (17.35). The attenuated signal depends on $\tau = t - l/c_{ph}$ and (17.61) holds for $t \geq l/c_{ph}$. The attenuation time constant $\tau_0 = (l\,\alpha)^2/(\pi f)$ is a characteristic, frequency-dependent rise time function of the attenuated pulse. In the frequency region in which the skin effect is dominant, τ_0 is almost independent of frequency and roughly corresponds to the time needed to reach half of the pulse height amplitude. The total attenuated signal of a rectangular pulse of length Δt is $v(t) - v(t - \Delta t)$, composed of the rising edge and the Δt-delayed falling edge. Figure 17.42 shows rectangular pulses deformed by dispersion [404, 819] for different ratios of pulse length and attenuation time constant.

Eye diagram and damping correction. In high-speed transmission ($\gtrsim$ GHz) of digital signals it should be ensured that the logic values 0 or 1 do not change by the transportation of the signal through the cable, since otherwise the information encoded in the bit stream is distorted. A simple visual inspection of the transmission quality is obtained by the so-called *eye diagram*, obtained when many bit streams are superimposed on an oscilloscope. The eye diagram provides the possibility to assess the signal quality in high-speed data transmission and the quality of a detector readout system at the end of the readout chain. Figure 17.43 shows a simple measurement set-up (fig. 17.43(a)) as well as the relation between a sequence of logic pulses (bit stream) and the eye diagram for the ideal case (fig. 17.43(b)) and for the case that the transmission causes the pulses of the bit sequence to be washed out (fig. 17.43(c)). They can still be interpreted correctly, however, provided the eye is sufficiently open, as defined by the black area.

An electronic damping correction called *pre-emphasis* improves signal transmission when required transmission rate and cable length question a clean transmission of measured data. As an example of a damping correction we take the Belle II experiment

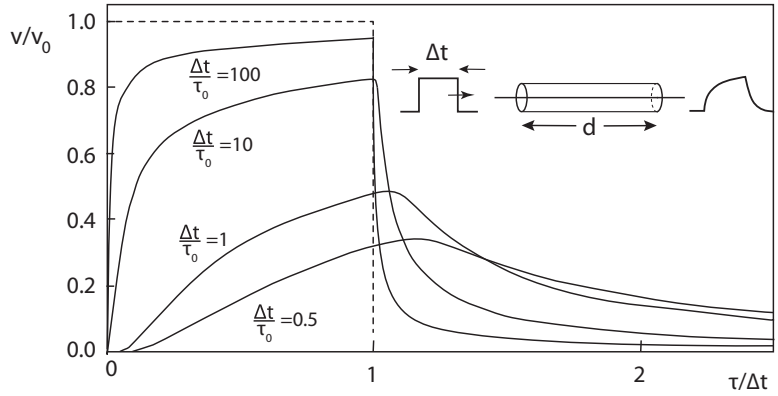
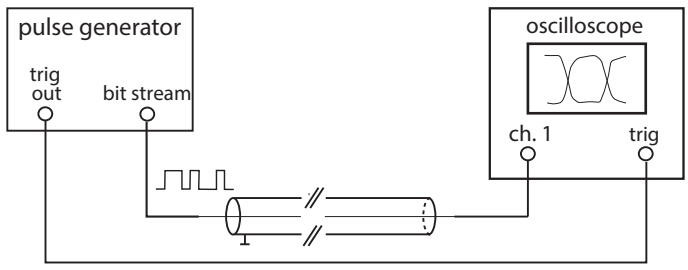
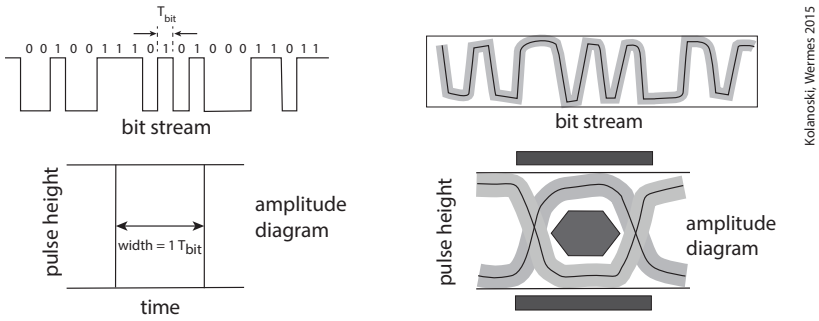


Fig. 17.42 Deformation of a rectangular pulse after propagation through a long cable according to (17.61) [404] as a function of the ratio of the (propagation time subtracted) time $\tau = t - d/c_{ph}$ and the pulse length Δt and for different ratios of $\Delta t \,/\, \tau_0$.

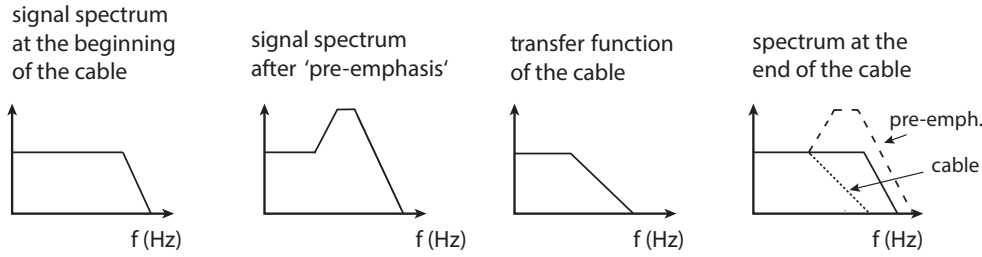


(a) Measurement set-up.

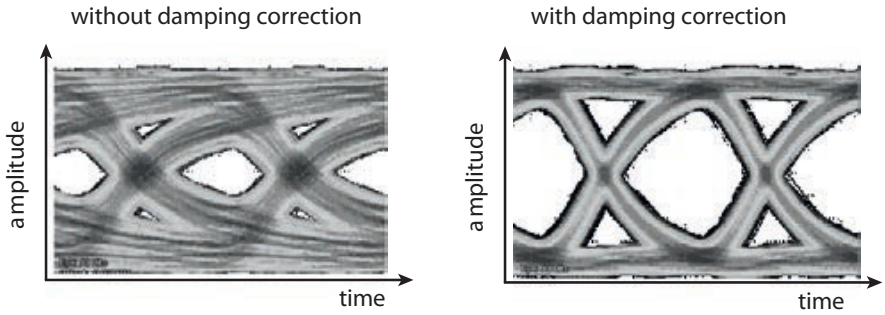


(b) Ideal transmission. (c) Real transmission.

Fig. 17.43 Transmission characteristic expressed by means of the eye diagram. (a) Measurement set-up: the oscilloscope picture shows the progression of a random bit stream (0's and 1's) after a certain propagation distance. A trigger signal synchronises the streams. (b) Ideal bit stream (T_{bit} = width in time of one bit) and eye diagram; (c) real bitstream and eye diagram. The pulse sequence is irregular and varies in amplitude and shape, indicated by the grey areas. The dark areas in (c) are specification masks that must not be touched in order to warrant loss-free transmission of the data streams.



(a) Damping correction by means of *pre-emphasis*-filter (double logarithmic representation).



(b) Eye diagram without (left) and with (right) correction.

Fig. 17.44 Improvement of data transmission quality by means of a damping correction (*pre-emphasis*). (a) Enhancement of the high-frequency part of the signal by a pre-emphasis filter. (b) Comparison of eye diagrams without (left) and with (right) damping correction applied in the Belle II experiment [607].

at the SuperKEKB accelerator at which the digitised data of the DEPFET-Pixel detector [860] of the experiment (see also section 8.10.2) are electrically transmitted over a length of about 15 m [607]. For this transmission, the high-frequency parts of the signal, which are attenuated most by the damping, are amplified at the driver end (pre-emphasis) by a high-pass filter on the driver chip [650] (fig. 17.44(a)) such that after having passed the transmission line, attenuation and enhancement of the high-frequency parts compensate each other. Figure 17.44(b) shows the corresponding eye diagram with and without damping correction.

17.9 Detector dead time

In general, for every readout channel there is a minimum time after registering a signal for which the detector can no longer register a further signal in this channel or at least cannot distinguish a second or third signal from the first one. This *dead time* is usually caused by the detector itself since it needs a certain recovery time after sensing a particle and generating a signal. To give examples, the recovery times of Cherenkov detectors or organic scintillators are comparatively short (typically in the nanosecond range), whereas Geiger counters cannot issue another signal for milliseconds after having fired. A further dead time is caused by the trigger and readout electronics that need a certain amount of time for signal registration and processing. Often no other signal is accepted by the entire detector on any readout channel during this time. In other cases the detector dead time of an individual channel does not necessarily govern the

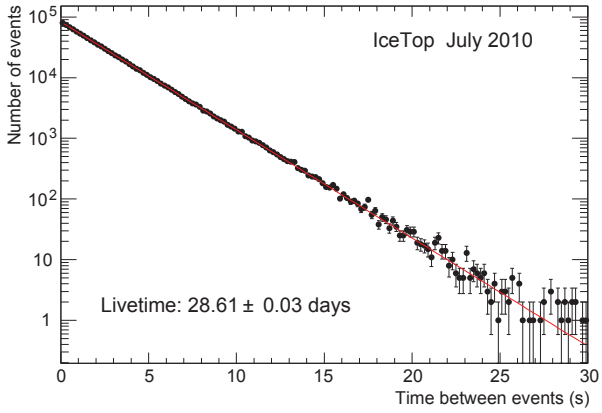


Fig. 17.45 Distribution of time differences [403] between two events registered with the air shower detector IceTop [16].

dead time of the entire detector. Often an experimental measurement consists of many individual signals, for example the hits of a particle track in a tracking detector having many layers, which allows accounting for a channel's dead time as an inefficiency.

Depending on how long a channel is paralysed by a dead time τ and how high the signal rate is, which may either appear in random order or at a fixed repetition rate, this can lead to a significant fraction of time during which the apparatus is not ready for a new measurement. The detector dead time must be determined, if possible experimentally and at best in parallel to the normal data taking. In offline data analysis, the measurements are to be corrected for this dead time.

In the following sections we describe different scenarios for dead times which depend on different probabilities for events to occur and their sequence in time. We assume that the corresponding dead times cannot be reduced by special measures, for example by parallelising trigger decision and data taking by means of pipeline buffers, a method that is often used in complex experiments operating at high input rates. Such, ideally dead-time-free, systems are described in chapter 18.

17.9.1 Equally distributed events

17.9.1.1 Time distribution

In this section events are considered which occur randomly but with equal probability per unit of time, for example in decays of long-lived radioactive isotopes or when observing cosmic rays. In these cases the number of events in a time interval Δt is Poisson distributed. With n being the event rate, the average number in a fixed time interval Δt is given by $N = n\,\Delta t$. The probability that k events appear in this interval is given by the Poisson distribution:

$$P(k|N) = \frac{N^k}{k!}\,\mathrm{e}^{-N}\,. \tag{17.62}$$

At any point in time the probability that the next event occurs after a time t in the interval dt is given by the product of the probability $P(0|N)$ not to see any event in the interval $[0, t]$ and the probability $n\,dt$ to see exactly one event in dt:

$$dp(t) = P(0|N)\,n\,dt = n\,\mathrm{e}^{-N} dt = n\,\mathrm{e}^{-n\,t} dt\,. \tag{17.63}$$

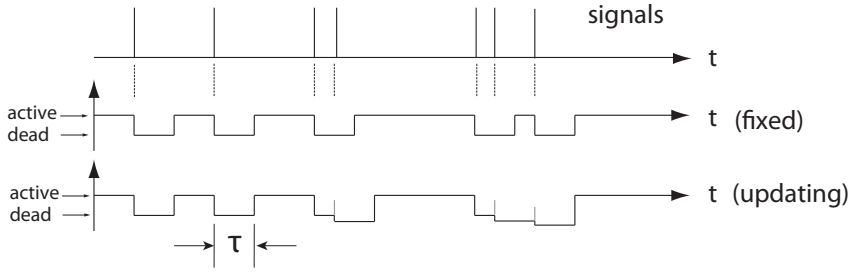


Fig. 17.46 Detector signal response and dead-time behaviour for two idealised systems (adapted from [616]). Top: time sequence of signal events; middle: response of a detector with fixed dead-time interval after a registered event; bottom: response of a detector with updating dead-time interval after a registered event.

Since $dp(t)/dt$ then is the distribution of times after each of which the next event arrives, it is also the distribution of time differences Δt between two events (simply consider that the start time $t = 0$ in the above derivation is assigned to the last respective event):

$$\frac{dp(\Delta t)}{d(\Delta t)} = n\,\mathrm{e}^{-n\,\Delta t}\,. \tag{17.64}$$

Events appearing at random within a time interval (i.e. uniformly distributed in time) therefore follow an exponential distribution in their time differences. Such an exponential behaviour of time differences is shown in fig. 17.45 for air shower events detected by the detector IceTop described in section 16.4.3. The distribution agrees very well with the exponential law of (17.64).

17.9.1.2 Dead times for events uniformly distributed in time

On every event there follows a time interval (the system dead time τ) during which the detector cannot register new events. The system dead time is to be distinguished from the dead-time interval Δt (also called the *dead time*) which depends on additional conditions as we discuss in the following. For this we consider two special cases illustrated in fig. 17.46, see also [616]:

(a) *Detector with a fixed dead-time interval (non-paralysable detector):* In this case the detector system generates a time interval $\Delta t = \tau$ after every registered event, within which the detector is not sensitive to register further events. This time interval is fixed and is not updated if further events appear during this time.

(b) *Detector with updating dead-time interval (paralysable detector):* In this case, if a further event appears during the time the system is dead, Δt will be updated by τ. Since this can happen repeatedly, Δt can in principle become infinitely long (at high rates).

In the example of fig. 17.46, of the seven signal events five are registered in the case of a fixed dead time and four in the case of updating dead time.

Case (a): fixed dead-time interval. Let n be the true average event rate (N events per time t) and m be the measured average rate (M measured events per time t). The average number of events that lie in a dead-time interval τ, and hence are

not registered, is $n\tau$. Since the dead-time intervals appear with rate m the rate of non-registered events is:

$$\frac{N-M}{t} = n - m = m\,n\,\tau\,. \tag{17.65}$$

Solving for n yields the true rate obtained from the measured rate with knowledge of the system dead time τ:

$$n = \frac{m}{1-m\tau}\,. \tag{17.66}$$

Conversely the measured rate as a function of the true rate is

$$m = \frac{n}{1+n\tau} \tag{17.67}$$

and the fractional dead time of the total time of data taking is

$$\frac{n-m}{n} = \frac{n\tau}{1+n\tau}\,. \tag{17.68}$$

Case (b): updating dead-time interval. In this case the dead-time interval is extended if during its duration another event occurs. The high-rate limit means that events are only registered for which the time distance to the previous event is larger than the system dead time τ. The probability for this to happen follows from (17.64):

$$p(\Delta t > \tau) = \int_{\tau}^{\infty} n\,\mathrm{e}^{-n\,\Delta t}\,d(\Delta t) = \mathrm{e}^{-n\tau}\,. \tag{17.69}$$

The measured average rate m for which $t > \tau$ is valid is the product of (17.69) and the average rate n:

$$m = n\;\mathrm{e}^{-n\tau}\,. \tag{17.70}$$

Equation (17.70) cannot be explicitly solved for n. It must be solved either numerically or can be obtained for example by fitting (17.64) to the distribution of time distances between successive events (see section 17.9.3).

Comparison of both cases. Figure 17.47 shows the different behaviour of the measured rate in the two cases described. For small rates ($n \ll 1/\tau$) both curves start tangentially to the diagonal $m = n$. For high event rates, that is, when the influence of dead time becomes stronger and stronger, the measurable rate m for a system with fixed dead time asymptotically approaches $m = 1/\tau$. This means that after the end of every dead-time interval a new event immediately appears leading to a new dead-time interval. By contrast the curve for updating dead time runs through a maximum at $n = 1/\tau$ with $m = n/\mathrm{e} = 0.37\,n$ and then decreases afterwards, whereas for fixed dead time at $n = 1/\tau$ the measured rate reaches the value $m = 0.5\,n$.

For the updating system and high input rate, the measured rate approaches the zero line because due to the permanently updating dead time the detector can hardly register new events. Two true rates can the be assigned to a measured rate m, a low and a high one, which can lead to misinterpretations.

For both dead time models the true event rate should be small compared to $1/\tau$ in order that the measured rate does not differ too much from the true rate. In particular, for updating dead-time intervals the detected rate m even decreases for $n > 1/\tau$ below the rates achievable at lower input rate n. In chapter 18 is explained how these limits can be overcome by parallel processing and buffering of data. This, however, requires significant additional circuit complexity.

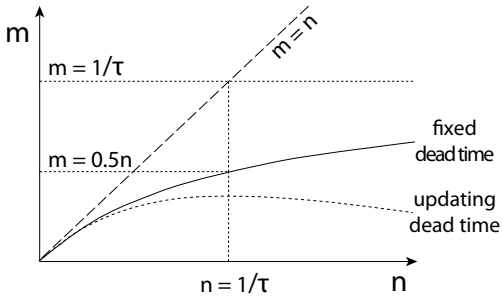


Fig. 17.47 Measured count rate m as a function of the true rate n for detectors with fixed dead-time interval (solid line), updating dead-time interval (dotted line) and without dead time (dashed line).

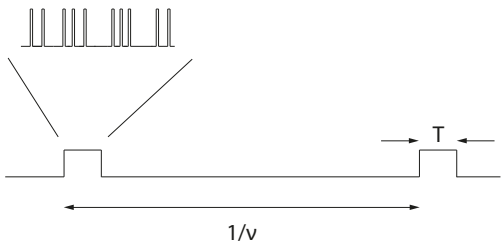


Fig. 17.48 Example of pulsed event generation. A pulse of several events repeats itself with repetition frequency ν. The pulse duration is T. Within T several events are observed.

17.9.2 Dead time for pulsed event generation

If the events to be measured by a detector do not occur in random order but are pulsed in *bunch trains* of duration T, appearing equidistant in time with repetition rate ν (fig. 17.48), then (17.66) and (17.70) may not be strictly applicable. Within a pulse train there can be additional time structures, as is the case for instance in packet structures of accelerator beams. An example are beams extracted from a synchrotron in which particles or X-ray radiation (synchrotron radiation) are distributed in a limited time interval (typically in the millisecond range) around the acceleration maximum with a high-frequency (HF) substructure. Another example is the operation of a linear collider with pulsed HF power. Here the beam particles are accelerated over a time interval T with repetition rate ν in a sequence of bunch trains with distances Δt. The bunches of two linear colliders, accelerated in opposite directions, are brought into collision in an interaction zone where the detector is located. A number of detectable collision events can observed during the crossing of the bunch trains, which can also be zero. A typical set of parameters for the planned International Linear Collider (ILC) [555] is $T = 1\,\text{ms}$, $\nu = 5\,\text{Hz}$ with about 1300 bunches with time distance $\Delta t = 554\,\text{ns}$ in every pulse [168].

In what follows the substructures shall be assumed negligible relative to the pulse duration. Furthermore we assume as before that no additional measures like buffering and parallelisation are taken to reduce the dead time (this would not be true for the ILC). The quantities to consider are (as in the previous section) the true rate n and the measured rate m (each taken with respect to the total period $1/\nu$) and the corresponding number of events per pulse $N = n/\nu$ and $M = m/\nu$. One can then distinguish the following cases [616]:

$\tau \ll T$*:* If the detector's system dead time τ is much smaller than the duration of the beam pulses T, then the fact that a coarse bunch structure exists has little effect and (17.67) and (17.70) can be used.

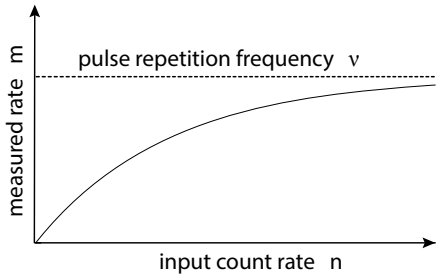


Fig. 17.49 Saturation behaviour of the measured event rate in the case of pulsed event generation according to (17.72) (dead time larger than pulse duration.)

$\tau > T$: If τ is larger than T but smaller than the time between the beam pulses $(1/\nu - T)$, then at most one event per pulse (bunch train) can be detected. Besides, the detector is always ready for detection at the beginning of a new beam pulse in this case. The measured event rate m can at most be equal to the pulse frequency, that is, $m \leq \nu$. Thus the average number of events per pulse is $M = m/\nu \leq 1$, whereas the average number of true events $N = n/\nu$ can be larger than one. The Poisson distribution determines again the probability that a true event per pulse occurs:

$$P(N > 0) = 1 - P(0) = 1 - \mathrm{e}^{-n/\nu} \,. \tag{17.71}$$

Since the detector is always ready at the beginning of a beam pulse it will register an event when at least one occurs. Hence

$$m/\nu = 1 - \mathrm{e}^{-n/\nu} \,. \tag{17.72}$$

The measured rate approaches the pulse repetition frequency (fig. 17.49). Solving for n we obtain

$$n = \nu \ln\left(\frac{\nu}{\nu - \mathrm{m}}\right) \,. \tag{17.73}$$

Equation (17.73) is valid for $T < \tau < (1/\nu - T)$. Neither the exact value of the system dead time τ nor the dead time behaviour (fixed or updating) play a role. The maximally measurable counting rate m_{max} in this case is equal to ν, the pulse repetition rate, since at most one event can be recorded per pulse. Thus it does not pay off to increase the input rate to values much larger than n by increasing the pulse intensity. The measured rate reaches 63% of the maximum rate already at $n = \nu$, 86% at $n = 2\,\nu$ and 95% at $n = 3\,\nu$.

$\tau \lesssim T$: For the case that detector dead time and pulse duration are of the same order a more involved treatment is needed, beyond the scope of this book, see [320, 978].

17.9.3 Methods of dead time determination

The dead time of detectors must be taken into account in quantitative analyses of experimental data by a dead-time correction. Since in most cases the dead time is not known and can also not be estimated with sufficient precision, it must be measured.

17.9.3.1 Uniformly distributed events

In this section, examples for estimation of the dead time will be discussed for events which take place randomly but which are on average uniformly distributed in time.

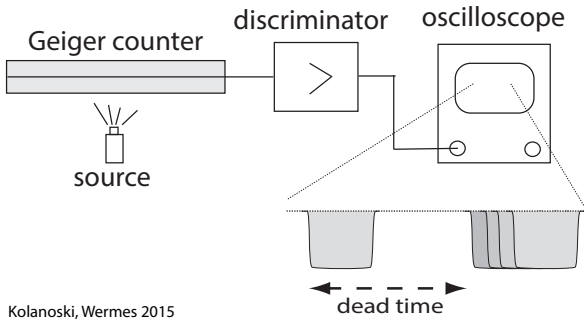


Fig. 17.50 Determination of the dead time of a Geiger counter by observation of signals from a radioactive source in an oscilloscope.

Such random sequences can be found for example in radioactive decays of long-lived isotopes, in background radiation and in cosmic rays. Dead time scenarios in detecting such events have been discussed in section 17.9.1.

Dead time estimation by use of an oscilloscope. The dead time of a simple counter system, for example a Geiger counter with a fixed dead time, can be measured very easily with an oscilloscope. The method is illustrated in fig. 17.50. A counting tube registers the radiation of a radioactive source. By a discriminator those signals with pulse heights above a certain threshold are converted into logic pulses which can be given on the input of an oscilloscope. The scope is triggered on an arbitrary pulse of the sequence such that this pulse and the subsequent pulses are displayed on the scope's screen (self-triggering mode). The time between the triggering edge and the first visible edge of a succeeding pulse is the dead time of the detector. This method to measure the dead time is valid for the different scenarios discussed in section 17.9.1 provided the event rate is high enough to allow clear recognition of the dead-time gap on the oscilloscope. On the other hand the rate must not be so high that the measurement becomes dead-time saturated, that is, no subsequent pulses are visible on the oscilloscope. The dead time determined this way is that of the specific system, in this example for the combination of counter and discriminator.

Dead time determination from decay curve. In the decay-curve method a short-lived radioactive isotope[20] is used having a decay constant λ and decaying with the rate

$$n(t) = n_0\,\mathrm{e}^{-\lambda t} \tag{17.74}$$

where n_0 is the true rate at the beginning of the measurement. Inserting $n(t)$ in (17.66) or (17.70), respectively, yields

$$m(t)\,\mathrm{e}^{\lambda t} = n_0 - n_0\tau\,m(t) \qquad \text{(fixed dead-time interval)}\,, \tag{17.75}$$

$$\ln \mathrm{m(t)} + \lambda \mathrm{t} = \ln \mathrm{n_0} - \mathrm{n_0}\tau\,\mathrm{e}^{-\lambda \mathrm{t}} \qquad \text{(updating dead-time interval)}\,. \tag{17.76}$$

By plotting $m(t)\,\mathrm{e}^{\lambda t}$ versus $m(t)$ one obtains a straight line with slope $-n_0\tau$ and offset n_0 in the case of a fixed dead time; for updating dead time, plotting $\lambda t + \ln m(t)$ versus $\mathrm{e}^{-\lambda t}$ results in the same slope with $\ln n_0$ as offset.

Dead time determination from time difference distributions. If time stamps of statistically distributed individual events are measured, the dead time of a detector

[20] Suited are for example long-lived isomers like ^{116m}In, a β^- emitter with a half-life of 54.3 min.

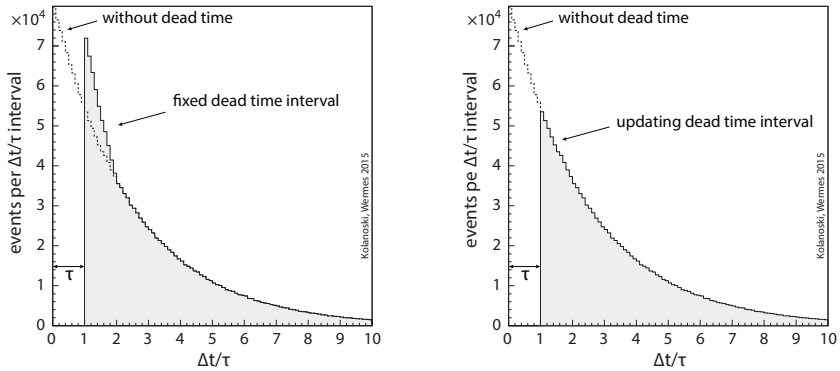


Fig. 17.51 Distribution of measured time differences for the cases: (a) fixed dead-time intertval and (b) updating dead-time interval (simulation). See discussions in the text.

system can be determined from analysis of the distribution of time differences $\Delta t = t_{i+1} - t_i$ between two successive events. The principle is the same as that using an oscilloscope for observation of successively appearing events (fig. 17.50). The method can, however, be applied more generally, in particular also for low event rates as for example in cosmic ray experiments.

Without detector dead time these time differences are exponentially distributed (eq. (17.64) and fig. 17.45),

$$\frac{dN}{d(\Delta t)} = N_0\, n\, \mathrm{e}^{-n\,\Delta t}\,, \tag{17.77}$$

where N_0 is the total number of events (the integral) and n the event rate. Due to dead times after registration of an event, smaller time differences are suppressed and the distribution deviates from a pure exponential shape. Assuming that the length of the system dead time τ is constant we analyse the two cases considered in section 17.9.1.2:

(a) fixed dead-time interval of length τ after a registered event: The event rate respecting the dead time is given in (17.67).

(b) fixed dead time τ after every event (i.e. not only after registered events): This means that after any event the remaining dead time is extended and is always at least τ (updating of the dead-time interval). The rate for dead-time updating is given in (17.70).

For the distribution of time differences including dead times, case (b) is the simpler case: no time difference smaller than τ occurs. All other time differences are not affected (fig. 17.51(b)). The number M of still registered events is given by the integral of (17.77) from τ to ∞:

$$M = N_0\, n \int_{\tau}^{\infty} \mathrm{e}^{-n\,\Delta t} d(\Delta t) = N_0\, \mathrm{e}^{-n\,\tau}\,, \tag{17.78}$$

in agreement with (17.70).

Case (a), that is, a fixed dead-time interval τ only after a *registered* event, is more complex than (b). In every dead-time interval there are on average $n\tau$ events that are not registered. The time difference between the last of these not registered events

to the first event occurring right after the interval can be shorter than τ. Since the latter event (appearing after τ) is registered it contributes to the counting rate m. Measured, however, is the time difference to the last registered event, which is τ plus the time of the event after the end of the dead-time interval. By contrast the same event would not be registered in case (b) when having a time distance smaller than τ to the previous event, since this previous event would have extended the dead-time interval. To find the distribution of these additional time differences we consider the distribution of time distances y of last events inside the dead-time interval convoluted with the distribution of the distance z of the first event after the dead-time interval, with the condition $\Delta t' = y + z < \tau$. The distribution of y, and also of z, each in an interval τ, is given by eq. (D.4) on page 829 for $k = 1$ (first event). The distribution of the (additionally) registered time differences $\Delta t = \tau + z$ arises from the product of the probability distributions for y and z with integration over the unobserved variable y:

$$\begin{aligned} f(\Delta t)d(\Delta t) &= \int_0^{2\tau-\Delta t} \left(n\,\mathrm{e}^{-ny}\right)\left(n\,\mathrm{e}^{-n(\Delta t-\tau)}\right) dy\; d(\Delta t) \\ &= \begin{cases} n\left(\mathrm{e}^{-n(\Delta t-\tau)} - \mathrm{e}^{-n\tau}\right) d(\Delta t) & \text{for } \tau < \Delta t < 2\tau\,, \\ 0 & \text{else}\,. \end{cases} \end{aligned} \tag{17.79}$$

The resulting time difference distribution is shown in fig. 17.51(a). Besides the gap at $0 \le \Delta t \le \tau$ as in fig. 17.51(b), an enhancement of events occurs between $\Delta t = \tau$ and 2τ. The size (normalisation) of this enhancement can be derived from (17.67) and (17.78), the measured event rate with dead time:

$$\widetilde{M} = \frac{N_0}{1+n\tau} = N_0\,\mathrm{e}^{-n\,\tau} + \Delta M \quad\Longrightarrow\quad \Delta M = N_0\left(\frac{1}{1+n\tau} - \mathrm{e}^{-n\,\tau}\right). \tag{17.80}$$

Here $\widetilde{M}$ is the measured number of events composed of the part from the integral from τ to ∞ as in (17.78) plus the enhancement contribution ΔM. Hence the distribution of the measured time differences in case (a) arises by adding the original distribution (17.77) above τ and the enhancement proportional to $f(\Delta t)$ calculated in (17.79):

$$\begin{aligned} \frac{dN}{d(\Delta t)} &= N_0\left(n\,\mathrm{e}^{-n\,\Delta t} + \frac{1}{1+n\tau} f(\Delta t)\right) && \text{for } \Delta t > \tau\,, \\ &= 0 && \text{else}\,. \end{aligned} \tag{17.81}$$

The coefficient $f(\Delta t)$ has been determined such that the integral over this term yields ΔM, as calculated in (17.80).

17.9.3.2 Electronic determination of the inactive time of a large detector system

If not only the dead time of an individual detector is concerned, but that of a larger system consisting of many detector components usually being hooked to one common computer, then the dead time is determined by an 'electronic stopwatch'. Such systems are discussed in chapter 18. Schematically, they consist of the detector, a triggering system deciding if an event should be read out or not, and the data taking system (see fig. 18.1). We describe here in short the determination of the inactive time in 'classical'

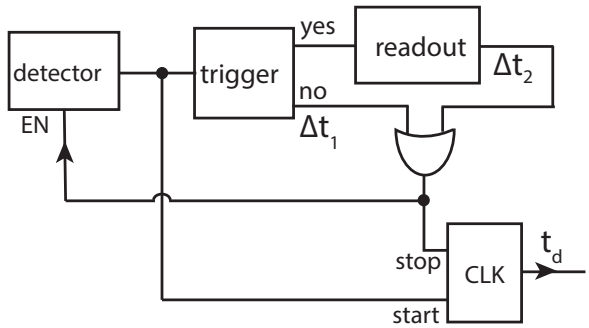


Fig. 17.52 Detector dead time in experiments with trigger (schematic). The detector is reactivated depending on the trigger decision (EN). The dead time t_d is measured for every event. It is determined either by the trigger time or by the sum of trigger and readout time.

systems without additional, dead time reducing electronics. For the more general case we refer to chapter 18.

The detector signals (or a selected subset of signals) are input to a trigger system (fig. 17.52) at a certain time t_0. At the same time a clock is started measuring the time until the trigger decision. During this time the detector is deactivated for data taking. If the trigger decision is negative the detector is reactivated, the clock is stopped and the measured time is stored as dead time. If the trigger decision is positive the dead time is extended by the time needed to read out the detector to the computer (readout time). A 'ready' signal ends the process.

It should be noted that in this scheme the individual channel dead times do not explicitly enter the consideration. They enter as inefficiencies of the individual channels in the analysis, either by determining them explicitly, as discussed above for the case of a counter tube, or by exploiting redundant information. For example, the inefficiency (containing the dead time) of a particular channel in a tracking detector can be determined using information from other channels concerned with the track that crosses many detector layers. In cosmic ray experiments, with events uniformly distributed in time, to measure for example the cosmic ray flux, the integrated dead time must be subtracted from the total time the detector is recording data.

In accelerator experiments the event rate is in general not constant. In fixed-target experiments the event rate is proportional to the (pulsed) beam intensity (eq. (3.4)) and in collider experiments it is proportional to the luminosity (eq. (2.2)). To measure cross sections these normalising quantities must be known. Dead time effects are implicitly taken into account by measuring them in parallel during data taking as discussed (for luminosity determination) in section 2.2.1. Dead times are implicitly taken care of if the measurement of normalisation quantities such as the luminosity suffer the same dead times as the total experiment.

17.10 Signal fluctuations and electronic noise

No measurement is perfect. Signal generation as well as signal processing in the readout chain suffer from imperfections, caused by statistical fluctuations and by systematic errors during the measurement process. The latter effects depend on the individual set-up, the measuring environment and the operation mode of the measurement apparatus. They can be treated here only in part. In this section we mainly consider statistical fluctuations in the signal generation and in the readout electronics. The closer a potential noise source can be to the front end of a readout chain (i.e. before amplification), the larger is its possible contribution to the total noise of the apparatus.

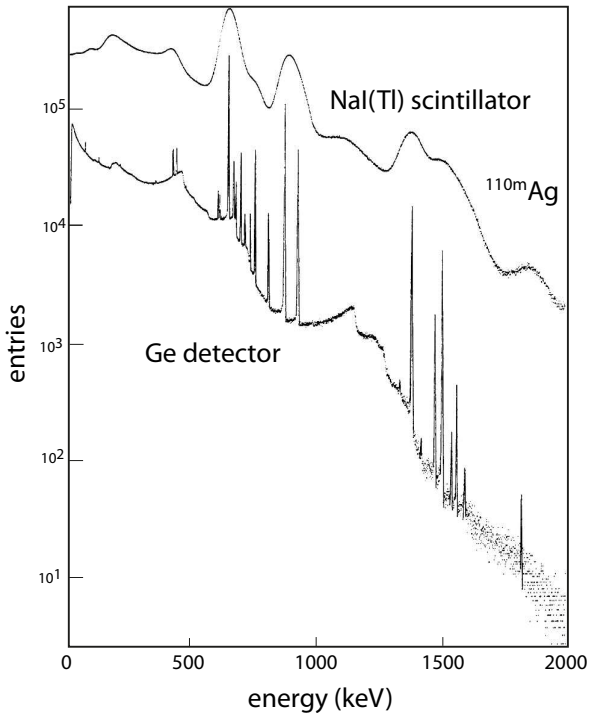


Fig. 17.53 Illustration of the influence of measurement fluctuations: γ spectrum, obtained using a NaI(Tl) scintillation detector in comparison to a Ge detector. The identification of sharp individual lines becomes possible by the better resolution of the Ge detector (smaller signal fluctuations and/or smaller electronic noise). Adapted from [782], with kind permission of IEEE.

17.10.1 Fluctuations in the measurement process

Fluctuations during the measurement process can never be completely avoided. With a suitable choice of the apparatus and the design of the set-up they can, however, be minimised. The fluctuations can have multiple origins:

- statistical fluctuations in the process of signal generation,
- fluctuating contributions of background processes,
- fluctuations in the signal processing electronics (electronic noise),
- pick-up of low- or high-frequency external signals (*pick-up noise*),
- coherent noise: oscillations in the electronic system (caused for example by ground loops),
- quantisation noise introduced by signal digitisation,
- temperature fluctuations

and many more. We concentrate in this section on purely statistical fluctuations (a) during the generation of the signal (signal fluctuations, in some systems also called Fano noise) and (b) in the electronic processing of the signal (electronic noise). The focus remains, as done so far in this chapter, on noise in detector systems detecting charge, for example semiconductor or gas-filled detectors (see also [905] and [837]).

Figure 17.53 shows an example of how measurement fluctuations influence the capability of a detector to resolve signals (see also [905]). Compared are γ lines of the meta-stable isomer ^{110m}Ag, once taken using a NaI(Tl) scintillator detector with PMT readout and once using a semiconductor detector (Ge) with charge integrating electronic readout [782]. The germanium detector has a much better resolution, mainly due to its smaller statistical fluctuations in the signal generation process.

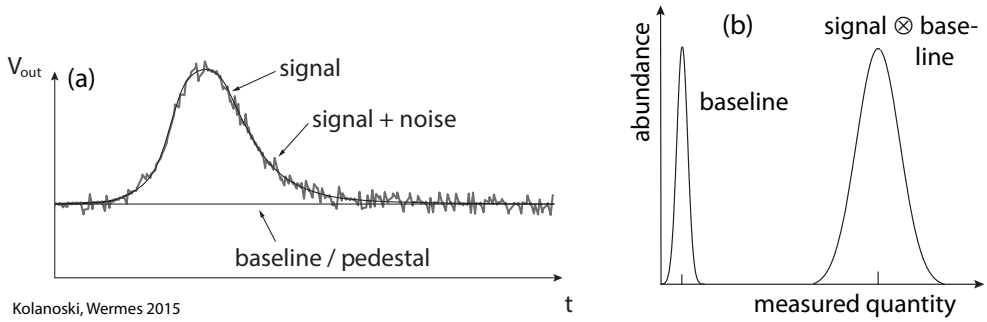


Fig. 17.54 (a) Schematic representation of a measurement acquisition (e.g. of a voltage proportional to a charge) with *baseline/pedestal* and noise contributions. (b) Abundance distribution of signals of the type shown in (a).

The output pulse of a detector often looks like that shown in fig. 17.54(a). If many such signals are obtained and the measurements (e.g. the integrated charge or the pulse height) are filled in a histogram, a spectrum of measurements is obtained, as in fig. 17.54(b). It shows the distribution of the *baseline* or *pedestal*, being the output value in absence of an input, and of the signal value, whose spread contains signal as well as baseline fluctuations which add quadratically.

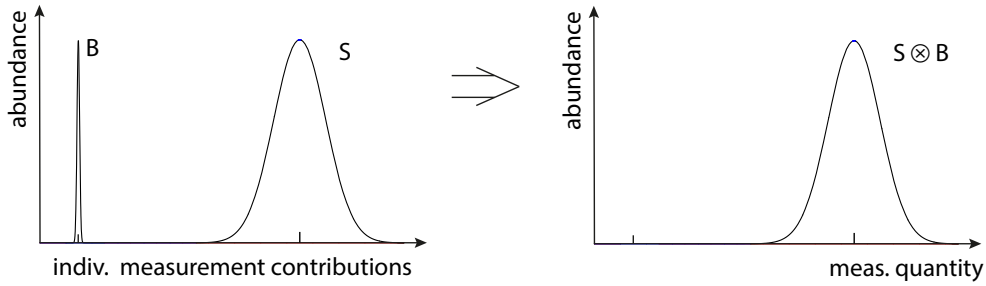
To illustrate which roles are played by fluctuations of signal and electronics we distinguish two cases:

(1) The width of the baseline distribution (B) is much smaller than the width of the signal distribution (S): $\sigma_S \gg \sigma_B$, as shown in fig. 17.55(a). An example for this situation is a NaI(Tl) scintillator crystal (see section 13.3), read out by a photomultiplier (see section 10.2.1). The total width is governed by the statistical fluctuations in the signal process, a generic property of the detector.

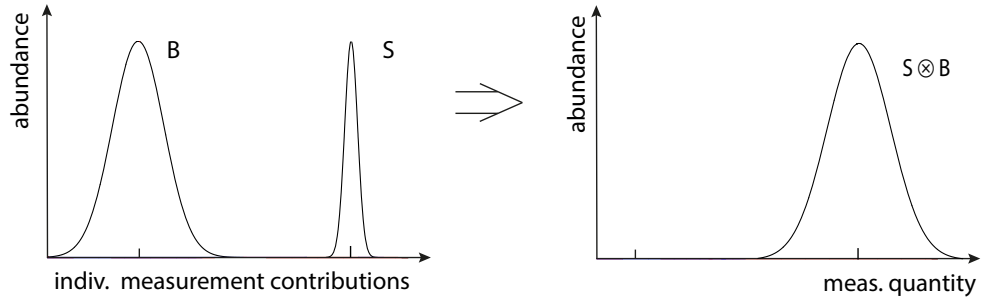
(2) The width of the baseline distribution (B) is much larger than the width of the signal fluctuations (S): $\sigma_B \gg \sigma_S$, see fig. 17.55(b). A prominent example for this situation are semiconductor detectors. In this case electronic noise largely determines the obtainable resolution. Its minimisation therefore plays an important role for such detectors and hence has a large influence on the design of the readout electronics.

To exemplify which element in a signal generation and readout chain represents the dominant contribution to the energy resolution, we consider in the following scintillating crystals used for instance in clinical positron-electron tomography (PET) (shown in fig. 2.12 on page 20). Two 511 keV gamma quanta from the annihilation of the e^+ with a nearby shell electron are collinearly emitted in opposite directions. With the help of fig. 17.56 we can identify the element in the generation process or readout chain which dominates the energy resolution:

– In a NaI(Tl) crystal about 43 000 scintillation photons are generated per MeV of deposited energy (see table 13.3 on page 520). A γ quantum of 511 keV hence generates $N_\gamma \approx 22\,000$ (optical) photons in a crystal.
– Of these about $N_\gamma \approx 13\,000$ (60%) reach the photocathode of the photomultiplier.
– Assuming a quantum efficiency of 20% we then have about $N_e \approx 2\,500$ photoelectrons that exit the photocathode of the PMT.



(a) $\sigma_S \gg \sigma_B$.



(b) $\sigma_S \ll \sigma_B$.

Fig. 17.55 Fluctuation contributions to a measurement spectrum for the cases (a) if $\sigma_S \gg \sigma_B$, as for example typical for NaI(Tl) scintillation detectors with PMT readout, and (b) if $\sigma_S \ll \sigma_B$, being typical for example for silicon microstrip detectors (with relatively large detector capacitance) with charge amplification readout. For both, on the left hand side the individual fluctuation contributions are shown and on the right-hand side the measurements containing both contributions. The resolution is given by quadratic addition $S \otimes B$ of the widths of baseline (B) and signal (S) distributions, respectively. In (a) the resolution is dominated by the fluctuations of the signal process, whereas in (b) the electronic noise fluctuations are large in comparison to the signal fluctuations and hence dominate the total resolution (see also [905]).

– For a typical PMT gain of 10^6 about $N_e = 2.5 \times 10^9$ electrons reach the anode leading to a current signal in the mA range assuming a typical pulse duration of NaI(Tl) detectors.

The achievable energy resolution σ_E/E is determined by the smallest number in this chain, that is, the one contributing the largest statistical uncertainty. The other quantities contribute much less to the total uncertainty due to quadratic error propagation. In this case it is the number of photons at the first dynode that are dominating the error:

$$\begin{aligned} \sigma_{\mathrm{dyn1}} &= \sqrt{N_{\mathrm{dyn1}}} = \sqrt{2500} = 50 = 2\%\ N_{\mathrm{dyn1}} \\ \Rightarrow \qquad \sigma_{\mathrm{sig}} &= \sigma_{\mathrm{dyn1}} \times (\text{amplification}) \approx 5 \times 10^7\,. \end{aligned}$$

The statistical fluctuations of the signal are much larger in this case than the electronic noise responsible for the width of the baseline (B). For PMTs this typically is in the range $10^{-(4-5)}$ times the signal [495] and is hence negligible here for the resolution of the detection system.

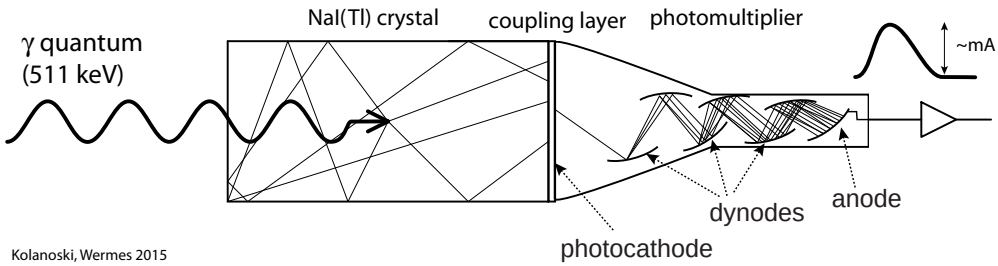


Fig. 17.56 Readout chain in a NaI(Tl) crystal–photomultiplier system. Scintillation light generated in the NaI crystal is converted into electrons at the photocathode. The electric signal is amplified over several amplification stages (dynodes) and extracted at the anode.

For the inverse case $\sigma(S) \ll \sigma(B)$ we take as an example the detection of a 60-keV γ ray from a ^{241}Am radioactive source detected with a Si semiconductor detector. The number of electron–hole pairs is

$$N_{e/h} = \frac{E_\gamma}{\omega_i} = \frac{60\,000\,\text{eV}}{3.65\,\text{eV}/(\text{e/h})} \approx 16\,500 \;\; \text{e/h pairs}$$

with ω_i being the average energy required for the production of an e/h pair. The signal fluctuations are computed as:

$$\begin{aligned} \sigma_{\text{sig}} &= \sqrt{N_{e/h}\, F}\, e^- \qquad (F = \text{Fano factor, see section 17.10.2}) \\ &= \sqrt{N_{e/h} \cdot 0.1}\, e^- = 40\, e^- \,. \end{aligned}$$

Relative to the signal size this yields an uncertainty of $\sigma_{e/h}/N_{e/h} = 0.2\,\%$ which is to be compared with the relative error of the first example of 2%. The noise in semiconductor detectors depends in the first instance on the electrode capacitance at the input of the amplifier (see section 17.10.3.3). For pixel and microstrip detectors it typically lies in the range of $\mathcal{O}(100\, e^-)$ for pixels to $\mathcal{O}(1000\, e^-)$ for strips. It is hence much larger than the signal fluctuations σ_{signal} and thus determines the achievable resolution of the system.

17.10.2 Signal fluctuations and Fano factor

Besides the (usually negligible) natural line width of a decaying nuclear level, fluctuations in the signal generation process are usually traced back to statistical fluctuations of the quanta (photons, charge carriers, phonons) created in the primary process underlying the Poisson statistics:

$$P(k, \mu) = \frac{\mu^k}{k!} \mathrm{e}^{-\mu} \tag{17.82}$$

with $P(k, \mu)$ being the probability for k events occurring when the average event number is $\mu > 0$. The variance σ^2 of the Poisson distribution is given by μ as well. For large mean event numbers $\mu = \langle N \rangle$ the Poisson distribution turns into a Gaussian. In practice this can already be assumed for N larger than about 10. The fluctuation of the measured number of events is

$$\sigma = \sqrt{N}\,,$$

and the relative error is $\sigma/N = 1/\sqrt{N}$.

If the energy of a particle is completely absorbed in a detector,[21] as for example an X-ray quantum in silicon, a better than $1/\sqrt{N}$ resolution results from the constraint that the total deposited energy is fixed. Quantitatively this is accounted for by the *Fano factor* F derived as follows.

The absorbed energy is distributed over a generally small number of energy loss processes whose energy sum must be the incident particle energy. A larger energy deposit onto the lattice (phonon excitation) conditions less charge carriers to be liberated. In this simple example with only two participating processes for energy deposition (phonons or charge carriers) the statistical error would be given by the binomial distribution with variance $\sigma^2 = N\,p\,(1-p) < N$, with p being the probability for phonon excitation, for example, which is smaller than the poissonian variance. The result of the energy conservation constraint, when apportioning the absorbed energy, is a better energy resolution than imposed by the Poisson error $\sqrt{N}$ of the charge carriers.

The general computation of the Fano factor is complicated. We consider here as a simple system a silicon detector for which the discussion can essentially be reduced to two energy loss mechanisms (see also [905]): the generation of electron–hole pairs and lattice excitations (phonon excitations). For the creation of an e/h pair at least the band-gap energy (in silicon $E_G = 1.1\,\text{eV}$) is needed. For every event, however, the deposited energy is subdivided differently for the generation of e/h pairs or for lattice excitations such that on average the energy of $w_i = 3.65\,\text{eV}$ is needed to create one e/h pair.

Let us assume that in a process a fixed energy E_0 be deposited with every event in a detector, for example the energy of an X-ray or γ quantum from a radioactive source. This energy is available for the creation of N_p phonon excitations and of $N_{e/h}$ electron–hole pairs. We thus have

$$E_0 = E_i\,N_{e/h} + E_x\,N_p\,, \tag{17.83}$$

where E_i and E_x are the (assumed fixed) energies necessary for one individual ionisation and one individual phonon excitation, respectively

The energy E_0 can split arbitrarily between ionisations or lattice excitations. Since E_0 is fixed, however, and since therefore for every absorbed quantum the same energy is deposited, then in every absorption process a fluctuation of a larger E_0 portion into phonon excitation ($E_x\,\Delta N_p$) must be compensated by a correspondingly smaller E_0 portion for ionisation ($E_i\,(-\Delta N_{e/h})$), where ΔN_p and $\Delta N_{e/h}$ are the number fluctuations of phonons and e/h pairs for one individual event, respectively:

$$E_x\,\Delta N_p - E_i\,\Delta N_{e/h} = 0\,.$$

Averaged over many absorption processes of the energy E_0 therefore yields:

$$\begin{aligned} & E_x\,\sigma_p = E_i\,\sigma_{e/h} \\ \Rightarrow \quad & \sigma_{e/h} = \sigma_p\frac{E_x}{E_i} = \sqrt{N_p}\frac{E_x}{E_i}\,, \end{aligned}$$

[21] Note that for a particle traversing a detector and depositing only a fraction of its energy (dE/dx) this constraint does not apply.

where Gaussian statistics has been assumed to be approximately applicable. Recasting (17.83):

$$N_p = \frac{E_0 - E_i N_{e/h}}{E_x}$$

and with the average number of e/h pairs,

$$N_{e/h} = \frac{E_0}{\omega_i}\,,$$

it follows that

$$\sigma_{e/h} = \frac{E_x}{E_i}\sqrt{\frac{E_0}{E_x} - \frac{E_i}{E_x}\frac{E_0}{\omega_i}} = \sqrt{\frac{E_0}{\omega_i}}\sqrt{\underbrace{\frac{E_x}{E_i}\left(\frac{\omega_i}{E_i} - 1\right)}_{=F\ \text{(Fano factor)}}} = \sqrt{\frac{E_0}{w_i}\,F}$$

$$\Longrightarrow \sigma_{e/h} = \sqrt{N_{e/h}\,F}\,. \tag{17.84}$$

Hence the resulting resolution $\sigma_{e/h}$ is better than $\sqrt{N_{e/h}}$ expected from Poisson statistics for independently generated e/h pairs, by the factor $\sqrt{F}$. Therefore the Fano factor F can also be defined as:

$$\text{Fano factor } F = \frac{\text{observed variance}}{\text{Poissonian variance}}\,.$$

For silicon one can estimate F from (17.84) using $E_x = 0.037\,\text{eV}$, $E_i = E_G = 1.12\,\text{eV}$ and $w_i = 3.65\,\text{eV}$ yielding $F_{Si} \approx 0.08$. The Fano factor is temperature dependent. Measurements (e.g. in [368] at 90 K) and theoretical computations [81] yield Fano factors between $F_{Si} \approx 0.084$ and 0.115. As can be seen from (17.84) the smallness of the Fano factor for Si is mainly due to the small ratio of the contributing process energies E_x/E_i. For (unrealistic) E_x values larger than 0.5 eV F would become larger than one.

For complete absorption of the energy of the incident particle the signal noise in silicon is therefore only about 30% of that expected from the sheer number fluctuations of created e/h pairs:

$$\sigma_{\text{signal}}(\text{Si}) \propto \sigma_{e/h} \approx 0.3\sqrt{N_{e/h}}\,.$$

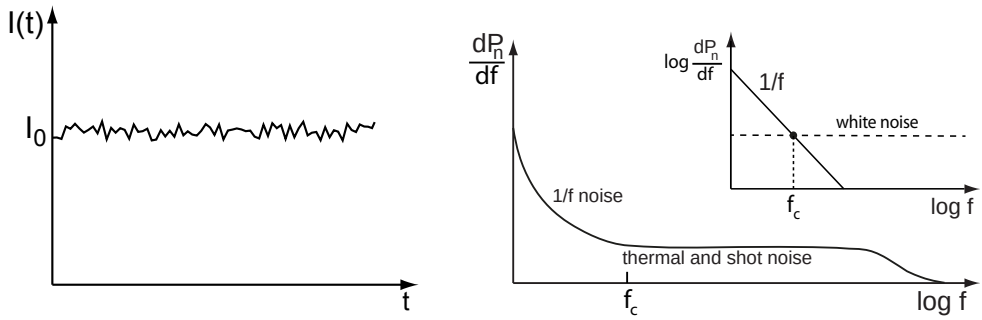
For different semiconductors and for argon gas one finds in [81, 588] the following theoretical values for Fano factors which are, however, largely verified experimentally:

Material	Si	Ge	GaAs	CdTe	diamond	Ar	liq. Ar
Fano factor	0.115	0.13	0.10	0.10	0.08	0.20	0.107–0.116

For detectors with more and also more complex signal generation processes, as for example scintillators, for which exciton processes also play a role (see section 13.3 on page 517), Fano factors larger than one ($F > 1$) can even occur [361].

17.10.3 Electronic noise

Statistical fluctuations in all electronic devices and circuits cause voltage/current fluctuations commonly termed (electronic) *noise*. Physical origins are for example thermal



(a) Current noise as a function of time.

(b) Spectral noise density (schematic) as a function of frequency.

Fig. 17.57 (a) Electronic noise in the time domain. (b) Typical noise power spectral density in the frequency domain. f_c is the corner frequency, marking the transition between 1/f and white noise (insert: double logarithmic representation).

motion of electrons, stochastic currents or charge appearence/disappearance, temporal and spatial fluctuations due to the quantisation of charge carriers, generation and recombination of electron–hole pairs as well as effects (interferences) outside of the circuit itself. In this section we discuss electronic noise due to statistical fluctuations in the processing of the signal, as opposed to noise in signal generation discussed in section 17.10.2.

Noise effects constitute a limit in every measurement process below which signals cannot be measured (or only under special conditions) because they are drowned in the noise. We first discuss the fundamental noise sources separately, then we describe their consequences for detectors and present methods for electronic noise reduction.

17.10.3.1 Noise sources

Noise is stochastic fluctuation about a set value, in electronics normally a voltage or a current. Since the noise average about a measurement value $\langle I \rangle = I_0$ is zero, the mean squared error (variance) $v = \sigma^2$ about I_0 is used to quantify noise (cf. fig. 17.57(a)):

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} \int_0^T \big(I(t) - I_0\big)^2 dt \,, \tag{17.85}$$

where σ^2 is the measure for noise intensity and σ is termed *rms noise* (voltage or current). Often it is convenient to describe noise activity by means of a frequency spectrum given by the Fourier transform of the time-dependent fluctuations. We define the spectral noise power density dP_n/df, specifying the power in a frequency interval of 1 Hz at the frequency f. The total noise power then is:

$$P_n = \int_0^\infty \frac{dP_n}{df} df \,. \tag{17.86}$$

Since the total noise power must be finite the noise power density (being constant for thermal and shot noise over a large frequency range) must eventually decrease from a certain frequency onwards. A possibly contained DC contribution manifests itself as a δ function at $f = 0$.

In electrical systems noise is usually described in terms of squared voltage or current spectral densities. The relation with the noise power density is established by the assumption of a fixed resistance R:

$$dP_n/df = \frac{1}{R}\, d\langle v^2\rangle/df = R\, d\langle i^2\rangle/df\,. \tag{17.87}$$

In language usage the quadratic expressions without R are also termed power densities. The current power density corresponds to the variance σ^2 in (17.85). The unit of noise power density is A^2/Hz for current noise and V^2/Hz for voltage noise, respectively. The rms noise quantity $\sigma = \sqrt{\sigma^2}$ hence characterises noise in units of $\mathrm{A}/\sqrt{\mathrm{Hz}}$ and $\mathrm{V}/\sqrt{\mathrm{Hz}}$, respectively.

If the noise power is frequency independent in a frequency band, we speak of *white noise*; noise contributions decreasing with increasing frequency approximately as $1/f$ are called 1/f noise (also called *flicker noise* or *pink noise*). Both noise contributions are illustrated schematically in fig. 17.57(b). In frequency spectra one sometime finds characteristic lines or frequency bands which are caused by electromagnetic interferences, such as the well-known 50 Hz or 60 Hz line of electric supply networks. These electromagnetic interferences are not generally classified as noise but are often called *pick-up noise*.

As mentioned, noise is caused by fluctuations. A current of N charge carriers between two electrodes at a distance d,

$$i = \frac{Nev}{d}\,,$$

can fluctuate independently in number N and in velocity v:

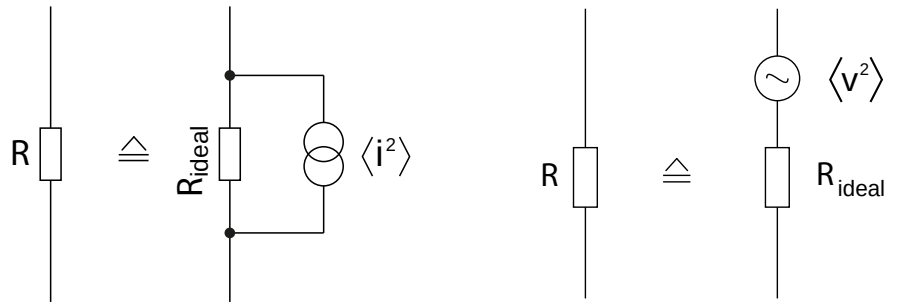
$$(di)^2 = \left(\frac{ev}{d}dN\right)^2 + \left(\frac{eN}{d}dv\right)^2\,,$$

expressed by the total differential di composed of the quadratic sum of both terms. In what follows we mainly distinguish three physical noise sources:

- thermal noise,
- shot noise,
- 1/f noise .

Thermal noise has its origin in thermal velocity fluctuations of charge carriers (Brownian motion). Shot noise and—at least in MOSFETs—also 1/f noise are caused by fluctuations in number density. Thermal and shot noise have a white frequency distribution over a wide frequency range; 1/f noise, as the name says, has a spectral behaviour as $1/f$ or—more general—as $1/f^\alpha$ with $\alpha = 0.5\ldots2$ (fig. 17.57(b)).

A multitude of names used for electronic noise discussions in various different contexts unnecessarily undermines an easy understanding. The above-mentioned three noise sources, however, are the fundamental physical sources that need to be taken care of in electrical circuits. Further nomenclature with somewhat different noise phenomena, such as *burst or random telegraph noise* or *avalanche noise* and others, are often in their origin related to 1/f noise and also follow a falling $1/f$ or $1/f^2$ spectral shape below a characteristic corner frequency f_c (see e.g. [472, 540]). They can be neglected for most applications, in particular for the detector readout addressed here. If circuitry-wise the position of a noise source is important for its impact one also finds



(a) Equivalent circuit with parallel current noise source.

(b) Equivalent circuit with serial voltage noise source.

Fig. 17.58 Equivalent circuits of a noisy resistor.

the expressions *parallel noise* and *serial noise*, respectively, meaning parallel or serial usually to the input amplifier.

In the following paragraphs we just quote the formulae for the respective main noise power densities. In appendix I the physical reasoning for these formulae describing the main noise sources is given.

Thermal noise. In any field-free conductor or semiconductor the free charge carriers move due to their thermal kinetic energy without any preferred direction. If an electric field is applied this thermal motion is superimposed on an existing current motion. If the expectation value of this fluctuation is measured one finds for a conductor with resistance R without an applied electric field (see appendix I):

$$d\left\langle i^2\right\rangle_{\text{therm}} = \frac{1}{R}\frac{dP_n}{df}\,df = \frac{1}{R}\,4kT\,df \tag{17.88}$$

with the Boltzmann constant $k = 8.617 \times 10^{-5}$ eV/K.

Thermal current noise, also called *Johnson noise* or *Nyquist noise* after its discoverers, is independent of the amount of current flowing through R. Furthermore it is frequency independent up to very high frequencies (white noise). In an equivalent circuit a noisy resistor is represented by an ideal resistor with a current noise source in parallel (fig. 17.58(a)) or equivalently with a serial voltage noise source (fig. 17.58(b)):

$$d\left\langle v^2\right\rangle_{\text{therm}} = R\,\frac{dP_n}{df}\,df = R\;4kT\,df\,. \tag{17.89}$$

Equations (17.88) and (17.89) are the so-called Nyquist formulae. At room temperature we have $4kT \approx 100$ meV $\approx 1.65 \times 10^{-20}$ V C. Hence in a 1 kΩ resistor we find a current independent thermal noise of

$$\sqrt{\frac{\Delta\left\langle i^2\right\rangle}{\Delta f}} = 4\frac{\text{pA}}{\sqrt{\text{Hz}}}\,.$$

Shot noise. Shot noise is a statistical charge carrier fluctuation occurring when charge carriers are emitted independently of each other over a potential barrier into a volume. An example is the emission of electrons in a vacuum tube where they must overcome the work function of the cathode. Another example is electron–hole

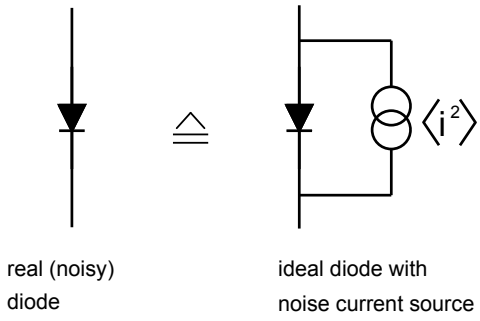


Fig. 17.59 Equivalent circuit of a noisy diode (pn boundary) with noise current source.

pair generation in semiconductors where the band gap energy must be overcome. Recombination of charge carriers also contributes to shot noise. The term *shot noise* originates from the sound of pattering pellet shots, a characteristic sound that could actually be heard in early electron tube radios. The Schottky relation [868] yields the following result for the spectral current noise power density (consult appendix I):

$$d\langle i^2\rangle_{\text{shot}} = 2eI_0\, df\,, \tag{17.90}$$

where I_0 is the average current in the noisy system (e.g. the leakage current in semiconductor detectors). Shot noise is hence directly proportional to the current I_0 and therefore often simply called 'current noise'. The frequency spectrum is 'white' as well.[22]

Noisy diodes can be described in an equivalent circuit diagram by an ideal diode and a constant current noise source in parallel (fig. 17.59). In contrast to thermal noise, which at a finite temperature is always present, even without a current, a current source is necessary for shot noise to occur. For example, for a current of $I_0 = 1\,\text{mA}$ one obtains an rms shot noise current of:

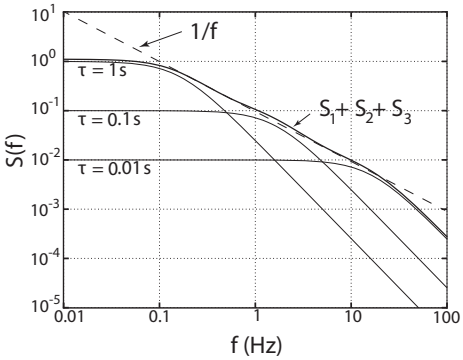
$$\sqrt{\frac{\Delta\,\langle i^2\rangle}{\Delta f}} = \sqrt{2eI_0} = 18\frac{\text{pA}}{\sqrt{\text{Hz}}}\,.$$

Shot noise at $I_0 = 50\,\mu\text{A}$ hence corresponds to thermal noise in a $1\,\text{k}\Omega$ resistor at room temperature.

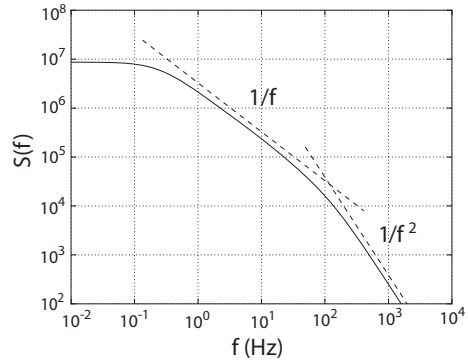
1/f noise (flicker noise). Under 1/f noise we understand in a fairly general sense all noise contributions featuring a 'non-white' frequency spectrum according to $1/f^\alpha$, with $\alpha = 0.5\ldots2$–3. It was discovered by Johnson in 1925 [576] in an experiment using vacuum tubes that he performed to test Schottky's shot noise theory of 1918 [867, 868]. At low frequencies, however, the noise in Johnson's experiment did not show a white behaviour but increased towards small frequencies. A first theoretical ansatz was published by Schottky in 1926 [869] (see also appendix I).

To this day there is no theory which completely describes the phenomenon of 1/f noise; 1/f noise does not only occur in electronics devices. It is also observed in many other systems [708, 969]. The velocity of ocean currents [931], the frequency of Earth rotation, traffic flow, membrane potential of brain cells, music [962], timing devices

[22] This only holds for frequencies small compared to the inverse of the pulse duration. Only δ pulses strictly generate white noise for $f \to \infty$.



(a) Superposition of three spectra with very different time constants.



(b) Superposition of many spectra with uniformly distributed time constants.

Fig. 17.60 Examples for appearance of $1/f$ behaviour in frequency spectra. In (a) three typical $1/f^2$ spectra of trap–release processes are superimposed with time constants each differing from the previous one [708]. The summed spectrum shows a characteristic $1/f$ behaviour. In (b) the noise power spectra (in arbitrary units) of 10 000 trap–release processes are superimposed having different, uniformly distributed $(1/\tau)$ time constants over three orders of magnitude in time. The lines represent $1/f$ and $1/f^2$ trends of the spectral density. Over a range of 3–4 orders of magnitude a 1/f behaviour governs the spectrum [708].

from sand to atomic clocks and even the human voice exhibit 1/f behaviour for various phenomena [795,969]. Hence the question arises if 1/f noise is a fundamental statistical phenomenon or if it is a special property of a certain system.

For electronic systems capture/release processes with different time constants are usually identified as the physical cause of 1/f noise (see appendix I). In particular in MOSFETs charge carriers moving in the transistor channel can be trapped at the gate–substrate interface for some time. So-called trapping centres are occupied by a charge carrier which is trapped for some (short) time. The spectral density for a single trapping process is proportional to $\tau^2/(1+(2\pi f\tau)^2)$ which for $2\pi f\tau \gg 1$ turns into a $1/f^2$ behaviour. Superposition of only few $1/f^2$ spectra with different trapping time constants already creates a summed spectrum which over a wide frequency range shows a $1/f$ behaviour.

Figure 17.60(a) illustrates that three trapping processes with time constants of $\tau_1 = 1\,\mathrm{s}$, $\tau_2 = 0.1\,\mathrm{s}$ and $\tau_3 = 0.01\,\mathrm{s}$ superimposed already exhibit a 1/f spectrum over two decades. For many trap–release processes the 1/f range can extend over 3–4 decades, as shown in fig. 17.60(b) for a superposition of 10 000 trap–release processes with uniformly distributed time constants over three orders of magnitude [708].

The noise power density of 1/f noise is generally written as

$$d\left\langle i^2\right\rangle = K_\alpha \frac{1}{f^\alpha}\, df \qquad (\alpha = 0.5\ldots 2)\,. \tag{17.91}$$

The constant K_α must be determined for the respectively considered process (e.g. charge trapping in MOS transistors), usually using device simulations (see also the following section). The frequency where the 1/f noise spectrum equals the white noise spectrum is called the *corner frequency* f_c (see fig. 17.57(b) on page 784).

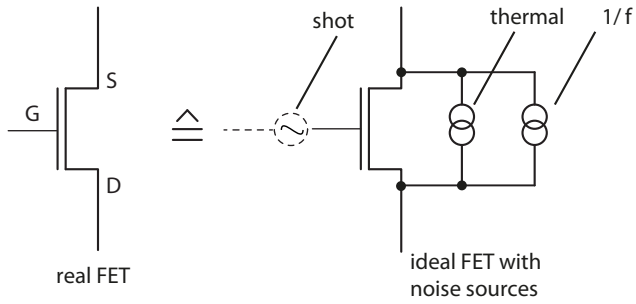


Fig. 17.61 Current (thermal and 1/f) and voltage (shot) noise sources for a FET input transistor. The dashed line indicates that the (voltage) noise source is external to the FET.

17.10.3.2 Noise in a (MOS)FET amplifier

Regarding noise performance in detector readout systems the foremost element in the signal chain, usually the preamplifier, plays the crucial role. In most cases one can even curtail this role down to a single circuit component, the input transistor which very often is a MOS field effect transistor (FET, see section 8.3.5 on page 297). We discuss noise sources in such a transistor in order to consider their effects on the noise performance of a readout system including preamplifier and pulse shaper in the next section (see also [837]).

A FET input transistor has the following noise current sources (fig. 17.61):

– thermal noise in the FET's channel,
– 1/f noise in the FET's channel,

and, if there is an external (voltage) noise source at the gate input[23], as for example introduced by a detector causing leakage current, there is in addition

– shot noise at the input (gate) of the FET.

In FETs, in particular in MOSFETs, current shot noise in the transistor itself is negligible since ideally no current enters the input (gate) and the transistor channel is viewed as a resistor (no potential barrier) exhibiting thermal and 1/f noise, but no shot noise.

Often it is useful to consider noise contributions at the output of a circuit element as effective noise contributions at the input. The two current noise contributions parallel to the transistor channel are equivalent to voltage noise contributions in series with the input (fig. 17.62):

$$\left\langle i_D^2 \right\rangle = \left\langle (g_m v_{\text{in}})^2 \right\rangle \,, \tag{17.92}$$

where $g_m = \partial I_D / \partial V_{GS}$ is the transconductance of the transistor with I_D = drain current and V_{GS} = gate-source voltage. The resistance of the channel depends, however, on the transistor geometry, that is, on the space charge region and on the channel doping, and is also position dependent (the channel narrows down towards the drain implant). The resistance R in (17.89) is for a FET replaced by $1/(\alpha g_m)$. In the saturation region (of the transistor output characteristic) g_m depends on the square root of the drain current, whose size is given by the transistor geometry:

$$g_m = \partial I_D / \partial V_{GS} \propto \sqrt{I_D} \propto \sqrt{W/L} \,. \tag{17.93}$$

[23]Notice that parallel shot noise (current) acts as serial voltage noise at the gate input (see also discussions on page 791).

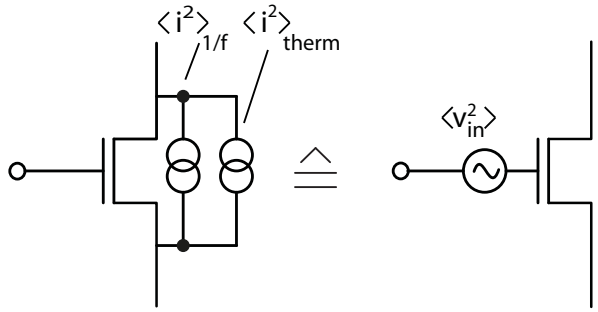


Fig. 17.62 Equivalent circuit for current noise contributions for a FET amplifier.

Here W and L are width and length of the gate. For JFETs and long channel MOSFETs one finds $\alpha = 2/3$ [472]. For short channel MOSFETs α is a function of $V_{GS} - V_T$ with V_{GS} = gate-source voltage, V_T = threshold voltage.

Hence with (17.92) we have for the thermal channel noise contribution at the input (gate) of the transistor:

$$\frac{d\left\langle v_{therm}^2\right\rangle}{df} = 4kT\frac{2}{3}\frac{1}{g_m}\,. \tag{17.94}$$

The 1/f noise in the channel arises from trapping and release of charge carriers at the interface between oxide and substrate. It depends on the gate area (width W × length L). Different expressions are used for a description. One is to describe 1/f channel noise as a noise voltage at the input of the MOSFET as follows:[24]

$$\frac{d\left\langle v_{1/f}^2\right\rangle}{df} = K_f\frac{1}{C'_{ox}WL}\,\frac{1}{f}\,, \tag{17.95}$$

where $C'_{ox} = \frac{3}{2}\frac{C_{GS}}{WL} \approx \epsilon_0\epsilon/d$ is the oxide capacitance per unit area (d = oxide thickness, C_{GS} = gate-source capacitance) and K_f is the *1/f noise constant* which depends on the technology (especially the feature size). K_f is related to K_α in (17.91). For MOS transistors in 250 nm technologies K_f is about 30×10^{-25}J for nMOS transistors and about 10–25 times smaller for pMOS transistors [689]. 1/f noise is the smaller the more homogeneous the transistor is fabricated. The inverse proportionality to the gate area $W \times L$ can be understood if one accepts as the physical cause trapping and release of charges q. The Poisson variance of the trap–release processes is then given by the number of existing traps and hence proportional to the area, while the resulting squared voltage fluctuations are proportional to $q^2/C^2 \propto 1/(WL)^2$, together yielding the inverse area dependence. Also, 1/f noise is the smaller the larger the conducting volume in the channel, since the influence of interactions with the interface on the current then decreases. For operation in 'strong inversion' (see page 294 in section 8.3.5) therefore, 1/f noise in the channel can be neglected.

The two noise sources in the transistor channel (thermal and 1/f noise) can hence be represented as voltage noise sources at the input as follows (see also fig. 17.62):

[24] In the literature one also finds different descriptions, based on different computer modelling, for example $\left\langle v_{1/f}^2\right\rangle = K'_f\frac{1}{{C'_{ox}}^2WL}\frac{1}{g_m^2}\frac{1}{f}\,df$ with a different constant K'_f and different dependences on C'_{ox} and g_m [278]. Here we use a common expression based on modelling by the program package PSPICE [269].

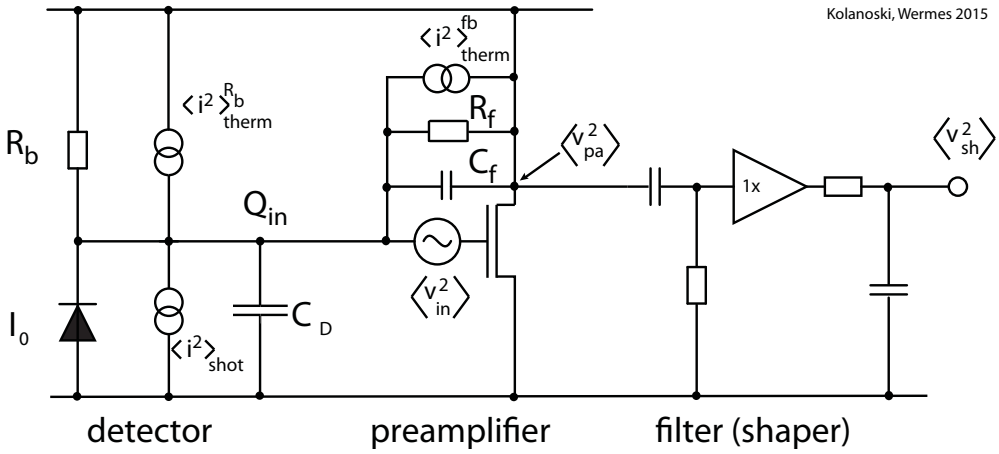


Fig. 17.63 Noise equivalent circuit of a detector-preamplifier-filter system. The detector is represented here as a diode with leakage current I_0 and capacitance C_D. The supply potentials at 'top' and 'bottom' are represented by horizontal lines.

$$\frac{d\left\langle v_{in}^2 \right\rangle}{df} = 4kT\frac{2}{3}\frac{1}{g_m} + K_f\frac{1}{C'_{ox}WL}\frac{1}{f}\,. \tag{17.96}$$

Depending on the operation mode of the amplifier, in particular on the chosen feedback, the voltage noise at the input $\left\langle v_{in}^2 \right\rangle$ has different consequences on the output. A readout scheme often used in particle physics detectors (see e.g. fig. 17.11 on page 729) uses an additional filter (shaper) after a charge integrating amplifier. We discuss such a scheme in the next section.

17.10.3.3 Noise in a detector–preamplifier–filter system

An often used readout chain of a detector is a preamplifier-filter(amplifier) system whose noise properties we will discuss in this section (see also section 17.3 and [837]). As a preamplifier we assume a charge-sensitive amplifier (CSA) with feedback through a capacitor C_f which is discharged by a resistor R_f, as shown in fig. 17.11. We further assume that the filter restricts the bandwidth of the system more strongly than the preamplifier, such that the transfer function of the preamplifier can be neglected. We also consider here only the usually dominant noise sources which are serial, respectively, parallel to the FET amplifier input. With the help of fig. 17.63 they can be identified as:

1. serial voltage noise, whose physical origin is thermal current noise in the channel of the input transistor,
2. serial voltage noise, whose physical origin is 1/f noise in the transistor channel,
3. and parallel current noise, caused by the detector leakage current.

Further noise sources which can be identified in the noise equivalent circuit diagram of fig. 17.63 are to a large extent negligible:

– Parallel thermal current noise $\langle i^2 \rangle^{R_b}_{\text{therm}}$, caused by the bias resistor R_b, becomes negligible if R_b is chosen sufficiently large (compare with (17.88)).
– (Serial) shot noise at the input of the FET caused by current into the gate (not drawn in fig. 17.63) is—particularly in MOSFETs—negligle.

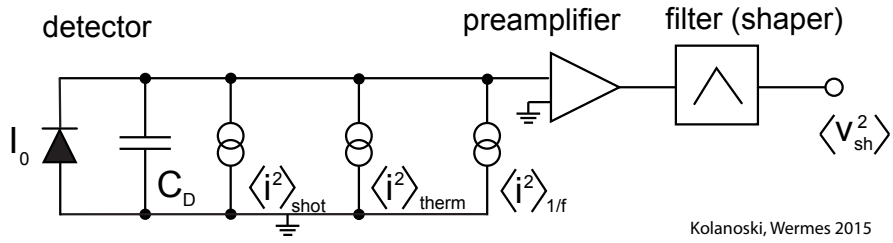


Fig. 17.64 Noise equivalent circuit of the readout chain where the main noise sources are arranged as current sources parallel to the preamplifier. For simplicity the bottom potential is grounded. The supply voltages are not relevant for noise considerations.

– Parallel thermal current noise $\langle i^2 \rangle^{fb}_{\text{therm}}$, caused by the feedback resistor R_f, is usually small (see subsequent discussion).

We analyse the influence of the dominant noise sources on the preamplifier–filter system of fig. 17.63 in two steps, first at the output of the preamplifier (v^2_{pa}) and then at the output of the shaper (v^2_{sh}).

It is also useful to treat the serial voltage noise sources (1/f and thermal noise) as equivalent parallel current noise via the capacitance C_D at the input of the preamplifier. This is possible via the relationship

$$\langle i^2_{in} \rangle \approx \langle v^2_{\text{in}} \rangle \, (\omega C_D)^2 \qquad (\text{with } \omega = 2\pi f) \tag{17.97}$$

since a voltage fluctuation serial to the amplifier input is equivalent to a parallel current fluctuation through all capacitances attached to that node. And we assume that C_D is large compared to the input capacitance of the FET ($\approx C_{GS}$) and the feedback capacitance C_f. The equivalent circuit diagram then takes the form as shown in fig. 17.64.

Following what was said previously and with (17.96) and (17.90), the current noise contributions at the input of the preamplifier add as

$$\frac{d\langle i^2_{\text{in}} \rangle}{d\omega} = \underbrace{2eI_0 \, \frac{1}{2\pi}}_{\text{'shot'}} + \underbrace{\frac{K_f}{C'_{ox} W L} \frac{1}{\omega} (\omega C_D)^2}_{\text{'1/f'}} + \underbrace{\frac{4kT}{g_m} \frac{1}{3\pi} (\omega C_D)^2}_{\text{'thermal'}}, \tag{17.98}$$

where I_0 is the average leakage current. This noise current, flowing through the feedback capacitance C_f, generates a noise voltage behind the preamplifier:

$$\langle v^2_{\text{pa}} \rangle = \langle i^2_{\text{in}} \rangle \left(\frac{1}{\omega C_f} \right)^2 . \tag{17.99}$$

Hence we obtain:

$$\frac{d\langle v^2_{\text{pa}} \rangle}{d\omega} = \frac{eI_0}{\pi \omega^2 C_f^2} + K_f \frac{1}{C'_{ox} W L} \frac{C_D^2}{C_f^2} \frac{1}{\omega} + \frac{4}{3\pi} \frac{kT}{g_m} \frac{C_D^2}{C_f^2}$$

$$= \sum_{k=-2}^{0} c_K \omega^k \tag{17.100}$$

with

$$c_{-2} = \frac{e}{\pi} I_0 \frac{1}{C_f^2}\,, \qquad c_{-1} = K_f \frac{1}{C'_{ox} W L} \frac{C_D^2}{C_f^2}\,, \qquad c_0 = \frac{4}{3\pi} kT \frac{1}{g_m} \frac{C_D^2}{C_f^2}\,. \tag{17.101}$$

Notice that in (17.100) the shot noise contribution increases with the detector leakage current I_0 and that all noise contributions decrease with increasing feedback capacitance C_f. The latter, however, implies smaller gain. Furthermore, the direct dependence of the 1/f and thermal noise contributions on the detector capacitance C_D (i.e. total input capacitance) is striking. For this reason small electrode pixel detectors can generally be operated with lower noise figures than strip or pad detectors, which typically have larger electrodes and hence larger capacitances.

The thermal noise contribution $d\langle i^2_{\text{therm}}\rangle/df = 4kT/R_f$ introduced by the feedback resistor R_f was argued above to be small and hence has been neglected in further discussions. It can now be easily estimated, because it acts on the preamplifier input in a very similar way as the leakage current shot noise contribution, yielding at the output a voltage noise contribution which in its form is analogous to the first term in (17.100):

$$\frac{d\left\langle v^2_{\text{pa}}\right\rangle_{R_f}}{d\omega} = \frac{2kT}{R_f} \frac{1}{\pi\omega^2 C_f^2}\,. \tag{17.102}$$

Its magnitude is usually small in comparison to the other contributions discussed, in particular to the leakage-current-induced shot noise contribution of a typical semiconductor detector. Similarly, the thermal noise contribution caused by the bias resistor R_b (see fig. 17.63) is usually negligible for large enough resistance, as already mentioned.

Effect of the filter stage (shaper). We now consider the influence of the shaper stage on the noise performance of the system. This stage can be a second gain stage but can also be simply a gain-one stage providing only a bandwidth limitation (see section 17.3). Second stage gain should—when carefully designed—not introduce additional noise after first stage amplification and hence should be negligible. As mentioned above, we also neglect bandwidth limitation by the preamplifier which—as a general rule—is not dominant. For a concrete example calculation, which nevertheless maintains quite some generality, we take a $(CR)^N(RC)^M$ filter (section 17.3) with $N = M = 1$, that is, a CR–RC filter composed of a high and a low-pass stage decoupled by a unity gain buffer (fig. 17.63).

In the time domain the output of a CSA can be approximated by a step function as long as the rise time is small and the discharge time $\tau_{\text{pa}} = C_f R_f$ of the feedback capacitor is large compared to the filter time τ (see fig. 17.11 on page 729). The band-pass filter transforms the step function into a pulse given by (17.23)

$$v_{sh}(t) = A \frac{t}{\tau} \mathrm{e}^{-\frac{t}{\tau}} \tag{17.103}$$

with filter time constant $\tau = RC$ and $t \geq 0$.

In the frequency domain the step function is a $1/s$ function with Laplace variable s and the filter transfer function is given in (H.10) of appendix H as:

$$H(s) = A\,\frac{s\tau}{(1+s\tau)^2}\,, \qquad A = \text{gain factor.}$$

For the noise power density transfer depending on frequency ω we need the squared transfer function $|H|^2$ (with $s \to i\omega$):

$$|H(\omega)|^2 = A^2\left(\frac{\omega\tau}{1+\omega^2\tau^2}\right)^2. \tag{17.104}$$

The noise power at the output of the shaper amplifier $\langle v_{sh}^2\rangle$ is obtained by integration of the noise power density spectrum (17.100) at the filter input, multiplied by (17.104):

$$\langle v_{\rm sh}^2\rangle = \int_0^\infty \frac{d\langle v_{\rm pa}^2\rangle}{d\omega}\,|H(\omega)|^2\,d\omega\,.$$

For our system with three dominant noise sources, a quasi ideal, that is, not predominantly bandwidth-limiting preamplifier, and a $(CR)^1(RC)^1$ filter with transfer function given by (17.104) we obtain using (17.100):

$$\begin{aligned}\langle v_{\rm sh}^2\rangle &= \sum_{k=-2}^{0}\int_0^\infty c_k\omega^k\,|H(\omega)|^2\,d\omega \qquad (17.105)\\ &= A^2\frac{1}{2}\sum_{k=-2}^{0} c_k\tau^{-k-1}\,\Gamma\left(1+\frac{k+1}{2}\right)\Gamma\left(1-\frac{k+1}{2}\right),\end{aligned}$$

where the Γ function obeys the relations

$$\Gamma(x+1) = x\Gamma(x), \qquad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}, \qquad \Gamma(1) = 1\,.$$

Hence we obtain

$$\langle v_{\rm sh}^2\rangle = \frac{\pi}{4}A^2\left(c_{-2}\,\tau + \frac{2}{\pi}c_{-1} + c_0\frac{1}{\tau}\right), \tag{17.106}$$

where the individual terms correspond as before to the contributions from shot, thermal and 1/f noise. Shot noise increases linearly with the filter time τ, whereas thermal noise decreases as $1/\tau$; 1/f noise is not influenced by the chosen CR–RC filtering.

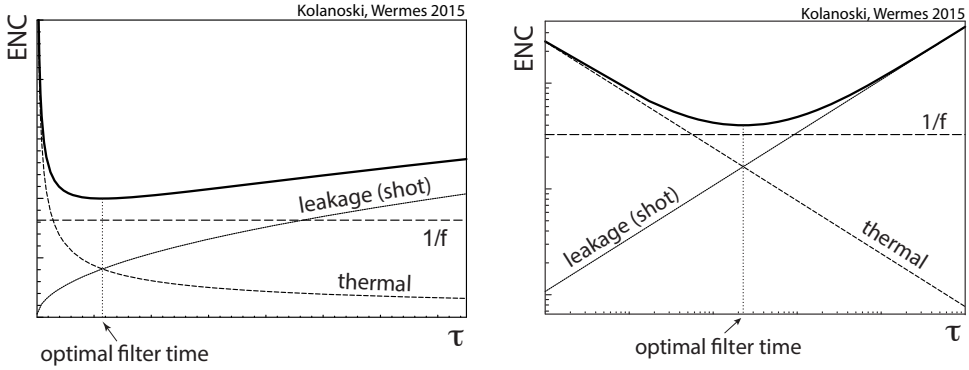
In order to assess this result we compare it to a 'signal' at the input of the system. A charge signal of one electron appears behind the CR–RC shaper as a voltage signal

$$v_{\rm sig} = \frac{A}{2.71}\,\frac{e}{C_f} \tag{17.107}$$

with overall gain factor A (this includes the preamplifier gain), Euler's number 2.71 ..., and the elementary charge $e = 1.6\times10^{-19}$ C. Here we used for the reference voltage $v_{\rm sig}$ the peak value of the amplitude $A\,\frac{1}{2.71}$ (compare eq. (17.103)).

The quantity *equivalent noise charge* ENC

$$\text{ENC} = \frac{\text{noise output voltage (V)}}{\text{output voltage of a signal of 1 e}^-\ \text{(V/e}^-)}\,,$$



(a) Linear representation. (b) Double logarithmic representation.

Fig. 17.65 Equivalent noise charge (ENC) in a typical detector–preamplifier–filter system as a function of the filter time (shaping time). The system and the assumptions made are described in the text. Shown are the contributions of the three main noise sources as well as the total noise. For given system parameters there exists an optimal filter time τ_{opt}.

or

$$\mathrm{ENC}^2 = \frac{\langle v_{\mathrm{sh}}^2 \rangle}{v_{\mathrm{sig}}^2} \tag{17.108}$$

is often used to quantify the noise performance of a charge measurement.

Due to its normalisation to the charge of one electron at the input, ENC is expressed in units of (equivalent) 'electrons' (e^-). Using (17.106) and (17.101) we obtain

$$\mathrm{ENC}^2(e^{-2}) = \frac{(2.71)^2}{4e^2}\left(eI_0\tau + 2C_D^2\,K_f\frac{1}{C'_{ox}WL} + \frac{4}{3}\frac{kT}{g_m}\frac{C_D^2}{\tau}\right). \tag{17.109}$$

In (17.109) the characteristic dependence on important system parameters, the detector capacitance C_D and the filter time constant τ, becomes evident:

$$\mathrm{ENC}^2 = a_{\mathrm{shot}}\,\tau + a_{1/\mathrm{f}}\,C_D^2 + a_{\mathrm{therm}}\,\frac{C_D^2}{\tau}. \tag{17.110}$$

They must be optimised in a readout system. Figure 17.65 shows ENC as a function of the filter time τ for a typical detector system (here a semiconductor detector). An optimal filter time exists for which noise is minimal:

$$\tau_{\mathrm{opt}} = \left(\frac{a_{\mathrm{therm}}}{a_{\mathrm{shot}}}\,C_D^2\right)^{1/2} = \left(\frac{4kT}{3\,eI_0g_m}\,C_D^2\right)^{1/2}, \tag{17.111}$$

The double logarithmic representation of fig. 17.65(b) shows this most clearly.

The dependences of the different noise contributions on τ can be understood from the derivation of (17.109):

– Shot noise, which is parallel current noise to the input, is proportional to the detector leakage current I_0 and increases with the filter time τ, since I_0 is effectively integrated over τ by the CSA–shaper system. While still being frequency independent (white) at the CSA input, $\langle v_{pa}^2 \rangle_{\mathrm{shot}}$ develops a $1/f^2$ dependence behind the preamplifier as described by (17.100), and a 1/f dependence after the shaper corresponding to a linear dependence on τ.

– Thermal noise in the transistor channel, while still being 'white' behind the preamplifier, is strongly reduced by the bandwidth limitation through the filter, leading to a decrease with $1/\tau$ after the shaper.
– For the 1/f noise part in the input transistor channel one would naively expect a larger contribution for large τ values (corresponding to small frequencies). This contribution, however, is cancelled by the bandwidth reduction by about the same factor, such that at the shaper output any τ dependence is no longer present.

In a system composed of detector, preamplifier, and filter one can determine the individual noise contributions by measuring the noise as a function of the filter time τ (fig. 17.65). The reader should also note the direct dependence of the noise on the geometry of the input transistor. The 1/f noise contribution ($\propto \frac{1}{WL}$) and also the thermal noise ($\propto \frac{1}{g_m}; g_m \propto \frac{W}{L}$) directly depend on width W and length L of the gate of the amplification transistor.

Using $C'_{ox} = 6\,\mathrm{fF}/\mu\mathrm{m}^2$ (typical for CMOS technologies (130 nm - 28 nm) employed in detector readout chips at the LHC experiments and their upgrades) we can write (17.109) in a convenient form:

$$\mathrm{ENC}^2\,(e^{-2}) = 11\,\frac{I_d}{\mathrm{nA}}\,\frac{\tau}{\mathrm{ns}} + 740\,\frac{1}{WL/(\mu\mathrm{m}^2)}\,\frac{C_D^2}{(100\,\mathrm{fF})^2} + 4000\,\frac{1}{g_m/\mathrm{mS}}\,\frac{C_D^2/(100\,\mathrm{fF})^2}{\tau/\mathrm{ns}}\,. \tag{17.112}$$

Examples. Using (17.112) we give three examples for typical detector readouts based on charge integrating amplification and filtering, but with different electrode geometries and leakage current contributions, for which the assumptions made in this section are largely valid. The ENC values obtained agree with corresponding measurements.

Pixel detector. As an example featuring small electrodes and correspondingly small input capacitances we choose a silicon pixel detector (section 8.7) with parameters $C_D = 200\,\mathrm{fF}$, $I_0 = 1\,\mathrm{nA}$, $\tau = 50\,\mathrm{ns}$, $W = 20\,\mu\mathrm{m}$, $L = 0.5\,\mu\mathrm{m}$, $g_m = 0.5\,\mathrm{mS}$, where we assumed a typical leakage current before the detector received substantial radiation damage. With (17.112) an equivalent noise charge of

$$\mathrm{ENC}^2 \approx (24\,e^-)^2\Big|_{\mathrm{shot}} + (17\,e^-)^2\Big|_{1/\mathrm{f}} + (25\,e^-)^2\Big|_{\mathrm{therm}} \approx (40\,e^-)^2$$

is estimated, which compares well with figures given for example in [837, 484].

Strip detector. For a typical silicon microstrip detector (see section 8.6.2) after radiation damage one obtains with $C_D = 20\,\mathrm{pF}$, $I_0 = 1\,\mu\mathrm{A}$, $\tau = 50\,\mathrm{ns}$, $W = 2000\,\mu\mathrm{m}$, $L = 0.4\,\mu\mathrm{m}$, $g_m = 5\,\mathrm{mS}$:

$$\mathrm{ENC}^2 \approx (750\,e^-)^2\Big|_{\mathrm{shot}} + (200\,e^-)^2\Big|_{1/\mathrm{f}} + (800\,e^-)^2\Big|_{\mathrm{therm}} \approx (1100\,e^-)^2\,.$$

We assumed in this example a large leakage current, typical for a Si-strip detector after irradiation with a large particle fluence ($\gtrsim 10^{14}\mathrm{n}_{eq}/cm^2$) and no noise-increasing radiation effects for the chip electronics. Note that the larger capacitance of strips compared to pixels leads to a much worse noise performance which can only be partially compensated by allowing more power in the amplification transistor leading to a larger g_m.

Liquid argon calorimeter. As an example of a detector with a large electrode capacitance we take a liquid argon calorimeter cell with typical values as given by the ATLAS electromagnetic calorimeter (see section 15.5.3.2 on page 621) in the central region. With the parameters $C_D = 1.5\,\mathrm{nF}$, $I_0 = < 2\,\mu\mathrm{A}$, $\tau = 50\,\mathrm{ns}$, $W = 3000\,\mu\mathrm{m}$, $L = 0.25\,\mu\mathrm{m}$, $g_m = 100\,\mathrm{mS}$, one obtains:

$$\mathrm{ENC}^2 \approx (1000\,e^-)^2\Big|_{\mathrm{shot}} + (15000\,e^-)^2\Big|_{1/\mathrm{f}} + (13500\,e^-)^2\Big|_{\mathrm{therm}} \approx (20200\,e^-)^2\,.$$

Here only a small (negligible) parallel shot noise (leakage current) contribution is assumed, which is typical for liquid argon calorimeters. The ENC obtained is consistent with measurements in [297]. The noise level must be compared to the signal, which in this case is given by the ionisation charges produced by the particles of a given electromagnetic shower (see also section 15.2). Depending on the energy of the shower the signal seen by an electrode of the liquid argon calorimeter is much larger than the noise level quoted above. To set the (low) scale, a minimum-ionising muon yields a signal-to-noise ratio of four [124].

Noise optimisation of system parameters.

- Noise in detector readout systems, as discussed in this section, only depends on few quantities in a significant way: the detector capacitance C_D (including additional parasitic capacitances), the filter time τ, the transconductance g_m of the input transistor as well as the leakage current of the detector delivered to the preamplifier. When treated independently of each other and of other constraints, τ and g_m should be chosen large while C_D and I_0 should be as small as possible. However, these parameters also influence other properties of the system such as the readout speed (g_m, τ and C_D) and the power consumption (g_m). Capacitance C_D and leakage current I_0 are detector properties, C_D is determined by design, largely driven by the detector granularity, whereas I_0 also depends on the operation environment (temperature, radiation damage).
- For a given input capacitance an increase of the transconductance g_m (and hence of the drain current of the amplifier, $g_m \propto I_D$) reduces the thermal noise, which in many applications dominates. However, power consumption and generated heat also increase. In a system with a large readout channel density this is a big disadvantage and should be minimised since cooling and heat removal entail more passive material in the path of particles and hence an increase of multiple scattering (see section 3.4) and secondary interactions.
- By *capacitance matching* we understand size matching of amplifier input capacitance C_i and detector capacitance C_D in order to maximise the signal-to-noise ratio (SNR). This is possible under the premise that thermal noise dominates and other noise sources can be neglected in relative comparison which, however, is the case for many applications. Then, in a frequency band Δf, the following signal-to-noise relationship holds:

$$\begin{aligned}\left(\frac{S}{N}\right)^2 &= \frac{(Q_S/C_{tot})^2}{\langle v^2_{\mathrm{therm}}\rangle} = \frac{Q_S^2}{(C_D + C_i)^2}\,\frac{3g_m}{8kT\,\Delta f} \\ &= \frac{3Q_S^2}{8kT\,\Delta f}\left(\frac{g_m}{C_i}\right)\frac{1}{C_i\,(1 + C_D/C_i)^2}\,. \end{aligned} \tag{17.113}$$

For 'normal' operation of the amplification transistor in so-called 'strong inversion' (see page 294 in section 8.3.5) g_m is proportional to the gate width W as is C_i, such that the ratio g_m/C_i is constant. According to (17.113) SNR is then maximised for $C_i = C_D$. This statement must, however, be taken with care and is more complex in detail (see e.g. [905], page 252 ff.). Especially the assumption that $g_m/C_i =$ constant is not given for other modes of operation of the input transistor or when using bipolar transistors.

– By cooling the system, which is usually done, in particular for semiconductor detectors and electronics, mainly the noise contribution introduced by the leakage current, which exhibits an exponential temperature dependence, is reduced. Input transistor thermal noise, by contrast, only decreases linearly with T by cooling. The temperature dependence of 1/f noise strongly depends on the fabrication process [313].
– The dependence of the equivalent noise charge on the detector capacitance C_D is approximately linear as long as channel noise of the FET, that is, the thermal and 1/f noise contributions, dominate as is the case for examle for larger detector capacitances. Commercial CSAs are hence characterised by the following *figure-of-merit* quotation at a given power and filter time τ:

$$\mathrm{ENC} = A\,e^- + B\,e^- \times \frac{C_D}{\mathrm{pF}}\,. \tag{17.114}$$

For (a typical) example ENC $= 200\,\mathrm{e}^- + 20\,\mathrm{e}^-/\mathrm{pF}$ at 2 mW power consumption and a filter time of $\tau = 200\,\mu\mathrm{s}$.

18
Trigger and data acquisition systems

18.1 Overview and requirements

The quantities measured by detectors are usually analog signals or counting rates which are in general electronically available, with few exceptions (see chapter 6). Usually one wants to process the data in digitised form using computers. In this chapter we discuss the interfaces between the so-called front-end electronics at or near the detector and a computer or a computer system. In order to limit the transfer rates of the interfaces and the sizes of the permanent storage media to what is necessary, the interesting events are usually preselected by *triggers*. Therefore the data acquisition (DAQ) and the trigger systems are closely connected and have to be coordinated.

A simple 'classical' scheme of a trigger–DAQ system is shown in fig. 18.1. From a selection of fast signals, for example from scintillators as in fig. 7.25(a), a 'pretrigger' is formed on a time-scale of about 100 ns or less and the experiment is stopped if the trigger decision is positive. The trigger starts the conversion electronics which converts the analog signals into digital information. Whether the converted data are finally read out or not may be decided by a more complex trigger. The readout into a computer proceeds via interfaces which convert the data formats according to a specific protocol into computer-readable formats. Today, however, such a simple scheme is only used by small experiments or in test set-ups.

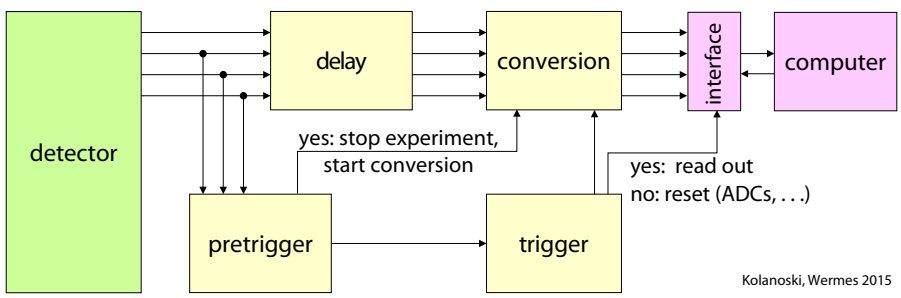


Fig. 18.1 Classical scheme of a trigger and data acquisition system; see description in the text.

Table 18.1 Typical rates and data flows of experiments at various machines (see table 2.2). Listed are luminosities, frequencies ν_{beam} of beam crossings, number of interactions N_{pile} or, respectively, triggers per beam crossing ('pile-up' multiplicity), the event rate ν_{storage} that can be written to a permanent storage device, the number N_{channel} of readout channels and the data volume per event D_{event}.

Beams	Machine	Luminosity ($\text{cm}^{-2}\,\text{s}^{-1}$)	ν_{beam} (MHz)	N_{pile}	ν_{storage} (Hz)	N_{channel}	D_{event} (kB)
e^+e^-	LEP	10^{31-32}	0.05	10^{-2}	1	10^5	100
ep	HERA	10^{30-31}	10	10^{-3}	10	10^5	100
e^+e^-	B-factory	10^{33-34}	120–240	10^{-5}	200	10^5	100
$\text{p}\bar{\text{p}}$	Tevatron	$> 10^{32}$	2.5	6	80	10^5	150
pp	LHC	10^{34}	40	25	200	10^7	1500

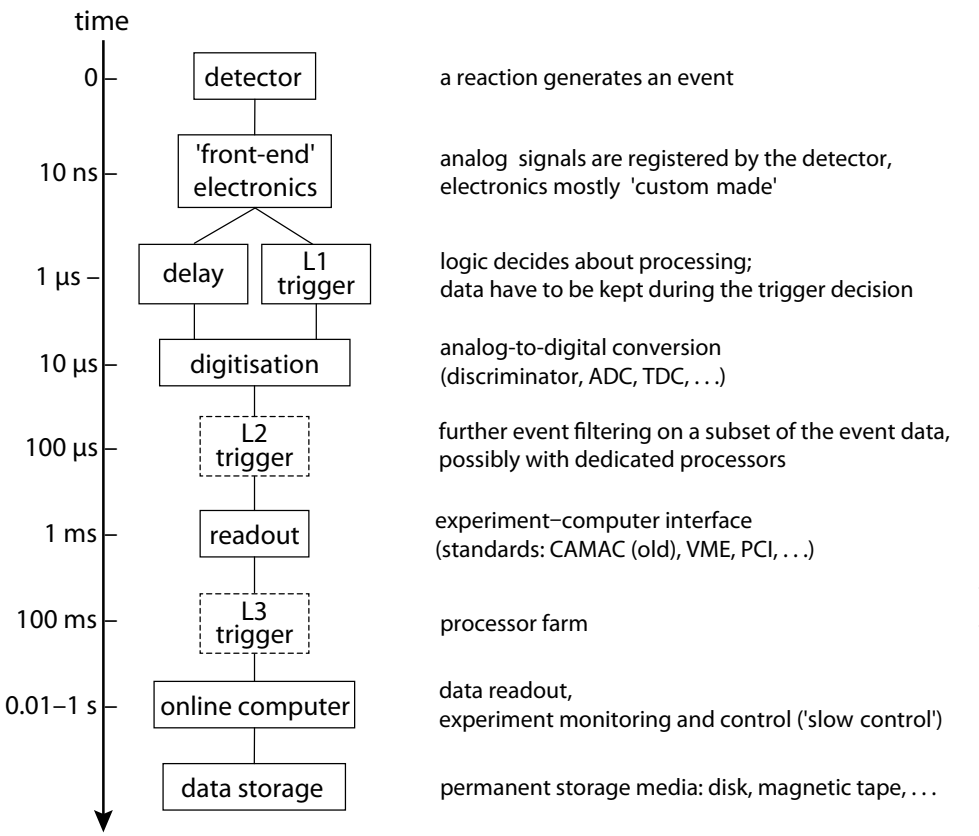


Fig. 18.2 Typical time structure of a trigger and data acquisition system in a high-energy experiment. In each specific case the time sequence can be quite different, in particular also the number of trigger levels (here L1–L3) can vary.

The requirements on the trigger and data acquisition systems depend very much on the size and complexity of the experiment. In particular the event and background rates and their time structure are essential parameters, for example if they are continuous or pulsed in certain time intervals. In recent decades the capabilities of data acquisition and processing have grown at a rapid rate and presumably will further continue to grow with the development of computers, networks and consumer electronics. Accordingly, in this chapter we will rather discuss general principles since quantitative performance numbers are often very short-lived. The given quantitative examples usually correspond to the technical possibilities in the starting phase of the respective experiment. Experiments that have a long lifetime often update their data acquisition system and computer capacities after an appropriate time.

Trigger requirements in particle physics experiments. Particle physics experiments at high-energy accelerators often have particularly high requirements on the data acquisition systems because typically the interesting events are rare and therefore high interaction rates are desirable. Table 18.1 shows typical examples of rates and data flow that experiments at various accelerators (section 2.2.1, table 2.2) are confronted with. In particular the LHC experiments feature a large discrepancy between the interaction rate yielding a high information flow at the detector *front-end* and the amount of data which can finally be read out and permanently stored:

- Data flow in the detector electronics (*front-end*): $\geq 100\,\text{Tbit/s}$. To illustrate this number, for example, 10^4 flash ADC channels (see section 17.7.1) with a resolution of 8 bit and a sampling rate of 100 MHz (all parameters are typical) yield a data flow of 10 Tbit/s.
- In contrast, for the permanent storage data flows in the order of only 100 Mbit/s may be reached.

In this example a reduction by a factor of at least 10^6 would be necessary. The trigger systems, which achieve such a suppression of uninteresting events and at the same time a high efficiency for the detection of interesting events, belong to the biggest challenges in the development of a detector system.

Figure 18.2 shows a characteristic scheme of a trigger and data acquisition system. Details of the time sequences and the number of separate levels can be very different for specific systems. In the following we want to discuss the essential properties of such systems:

- The analog data accrue within nanoseconds (at most microseconds) at the front-end of the detector.
- On a first trigger level (*level 1*, L1) a fast decision has to be made if an event is to be accepted. During the decision time (*trigger latency*) the data have to be kept on hold, which can be achieved by delay cables, electronic delays or *electronic pipelines* (series connected, clocked storage cells, see fig. 18.4), the choice depending on the complexity of the system.
- The data of the accepted events are then digitised (if not already done) and written into a buffer storage.
- A possible second trigger level has access to the digital data, however, usually only to a restricted amount of data that contain information about regions which have been recognised as interesting by the L1 trigger (*regions-of-interest*, RoI). The restriction to RoIs over several trigger levels is a general concept that allows one to keep data flows as low as possible and the reconstruction algorithms as simple as possible.

– After the second (or sometimes also a third) level the data from all storages of all detector components are combined to form a complete event record (*event building*).
– At this point a higher level trigger could access the complete data and could apply another filter.

On each level of the trigger–DAQ system the deployed hardware has been adapted to the requirements. Common characteristics are:

– In general the L1 trigger is assembled from dedicated hardware which in large experiments is usually a custom development. Integrated electronic circuits with a large number of logic gates with freely programmable connectivity become increasingly popular (*programmable logic devices*, PLDs). In smaller experiments one frequently resorts to modularly structured decision logic (e.g. modules employing the NIM standard; see section 18.4.1).
– For the digitisation again one can use custom developments or modular standards (section 18.2.1). Typical features of standardised readout systems (see section 18.2.1) are modules, for example ADCs, TDCs, DACs, ..., inserted into a common crate which features a backplane with address and data buses that connect the modules with each other and a control unit. The control unit has a link to a computer and can thus transfer data between the modules and the computer in both directions (read/write).
– At the second and higher trigger levels the triggers and event filters are usually run on 'processor farms', which are most economically realised by standard PC architecture (*commodity computers*). For parallel processing the events are distributed onto the farm nodes, which requires efficient management and network systems.
– In general the data are first stored to disks which may be later copied to other permanent storage devices.

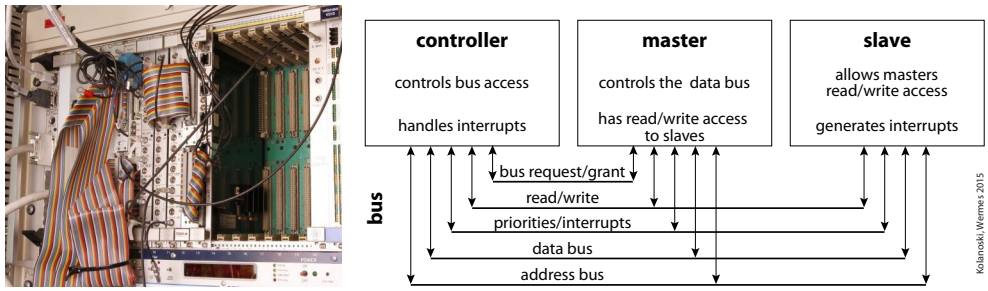
18.2 Data acquisition

In chapter 17 the detector-near signal processing has been discussed. Figure 17.2 shows a chain from the front-end electronics at the detector to the digitisation of the measured quantities, like pulse heights, charges and times. In the following we will talk about the transfer of these data via an interface into a computer. The interfaces are standardised systems or, in particular in large experiments or for special applications, also custom developments.

Besides the event data, which have to be transferred fast and with high rate in *real time*, there are also 'slow' data on status, monitoring and control of the detector (*slow control*), which are partly sent with each event or are collected by parallel, asynchronous systems.

18.2.1 Standardised readout systems

In order to make detector signals available for further electronic processing, an interface between detector and computer is necessary. As an example for an interface one can imagine a digital storage device with two-sided access, a so-called *dual-port memory*. From the detector side data are written into the memory which are read out by the computer from the other side. A *protocol* has to be agreed upon by both sides in order to fix the data format and the memory addresses. For the data acquisition of random events the detector has to communicate with the computer in real time, requiring the specification of a respective signal exchange in protocol and hardware.



(a) (b)

Fig. 18.3 VME data acquisition system: (a) photograph of a VME crate with plug-in modules (Source: DESY); (b) block diagram of a VME system depicting the communication structure of control and master/slave units via the VMEbus (adapted from [655]).

Since such interfaces are needed for many different experiments and applications and since custom developments are very costly, very early on standards for interfaces were defined which are generally usable due to a modular structure. The interface is a connection between a data line, called 'bus', on the detector side and a data line (or bus) on the computer side. For a flexible and efficient deployment of a readout system the standardisation of the communication software is also an important requirement.

The first standard with very broad and long-term applications was CAMAC (*computer automated measurement and control*), which was first described in 1969 in a EURATOM report [381] and was defined as the standard EUR4100 in 1972 by the ESONE committee.[1] The broad international acceptance was guaranteed because from the beginning it was internationally negotiated, amongst others with the NIM committee[2] in the USA. Being an electronic system that was designed in the 1960s, CAMAC has survived a surprisingly long time. Even today CAMAC systems are still employed, although no longer in the large experiments of particle physics. Restrictions come from the limited speed of the data transfer (24 bit parallel at about 1 MHz), the limited address space and most of all from the lack of support for multi-processor operation. In the course of time various systems have been proposed as alternatives, however, only few of these established themselves in the long run.

VMEbus. Since the 1980s the VMEbus standard (Versa Module Eurocard bus) was introduced and in 1987 defined under ANSI[3]/IEEE[4] 1014-1987 [552]. In contrast to CAMAC the VME standard supports the employment of multi-processor systems (the VME development was closely connected to the introduction of microprocessors).

The VME bus is a *backplane bus* in a crate which houses plug-in modules (see fig. 18.3(a)). Most modules serve as analog-to-digital converters, such as ADCs, TDCs or DACs (section 17.7.1). The analog data enter a module from the front side and the digital data are transported on the backplane. Data can be both sent and received by the detector. Received signals are mostly for detector control. The modules can act autonomously, often using their own processors. For each assembly it can be defined

[1] ESONE = European Standards on Nuclear Electronics.

[2] NIM = Nuclear Instrumentation Module (standard for modular electronics for signal and trigger processing).

[3] ANSI = American National Standards Institute.

[4] IEEE = Institute of Electrical and Electronics Engineers.

which module has the coordinating function (*master*) and which modules are subordinate (*slaves*), see fig. 18.3(b). The interfacing to the external computer is usually done by a special *controller* module.

The bus is divided into sub-buses: data bus, address bus, interrupt and priority bus, arbitration bus and utility bus. Data and addresses can be transferred as 32-bit words, in an extended version up to 64-bit words. In the original versions the data transfer rates were 40 Mbyte/s and 80 Mbytes/s for 32-bit and 64-bit words, respectively. The protocol is asynchronous, which means that the course of actions is not controlled by a common clock but rather by the exchange of control signals. Interrupt signals are sent in real time to the system on the priority bus. The arbitration bus configures the *master–slave* relation of the different processors in the system. The utility bus provides the power supplies delivering the required voltages.

VMEbus systems are employed in particle physics experiments and also in other science as well as in industrial, military, aerospace and telecommunication applications. Extensions of the original standard achieve transmission rates of several gigabytes per second.

ATCA, μTCA. In the field of telecommunication the requirements on speed and switching frequencies are continuously growing. The standards developed for these applications have properties which make them interesting for experiments with high data flow. For example, for the upgrade of the LHC experiments standards like AdvancedTCA, abbreviated ATCA (advanced telecommunications computing architecture) are considered to substitute the current modular bus systems like VME. The ATCA standard has been developed by a group of computer manufacturers, PICMG (PCI Industrial Computer Manufacturers Group) [767], for the next generation of communication technology. The specification with the designations PICMG 3.X follow the most up-to-date trends in the development of technologies of high-speed connections, of next-generation processors and in the improvement of reliability, operability and maintainability of such systems. In particular, the data transfer rate is typically 40 Gbit/s and thus orders of magnitude higher than for example for VMEbus systems. Another important advantage is the possibility of hard- and software replacements without the need to run down the system ('hot-swapping').

With the realisation of the specifications ATCA found a broad distribution in telecommunication and now also finds applications in industrial automation, in the military domain, in aerospace systems and in research. Due to the broad distribution components are readily available at relatively moderate prices which make such systems very attractive, in particular also for particle physics experiments.

The specification MicroTCA or μTCA has, as compared to ATCA, a reduced *form factor*, a term which refers to the mechanical dimensions of crates and plug-in modules. Especially for the requirements in particle physics and accelerator control (see e.g. [455]) the development 'xTCA for Physics' offers additional properties and options for ATCA and μTCA.

For the application of ATCA systems in particle experiments another important aspect enters the game: due to the high data transmission speed and the use of highly flexible, programmable logic devices (like FPGAs, see page 813), ATCA modules can be employed for front-end readout and triggering, a domain which used to be governed by custom-made electronics. Application examples are a development of a first-level calorimeter trigger for an LHC experiment described in [933] and of a readout system for a neutrino underground experiment described in [897]. At the time of writing, both

examples are still in the prototype phase. The future developments can be followed by the 'IEEE Real Time Conference' series.

18.2.2 Computer buses, interfaces and networks

With the advent of personal computers (PCs) new possibilities arose for data acquisition, in particular for the readout of single instruments, small apparatuses and test set-ups. The instruments can be connected through interfaces to an internal bus of the computer as a peripheral device, such as a printer or a display screen. The communication usually proceeds through specific driver software which is in general delivered together with the instrument. Extended versions of computer buses are also employed in larger experiments, as will be described further down.

PCI. A very common standard is the PCI bus (peripheral component interconnect). Because of the broad use for computers and consumer electronics the systems are reasonably priced and readily available, and the standard is continuously improved, in particular by rising transfer rates. The just-discussed (section 18.2.1) TCA standard is also PCI based.

Meanwhile interfaces connecting PCI to other backplane buses like CAMAC and VME are also provided. For more voluminous apparatuses the PCI standard can be used as an extension of the internal PCI bus of the computer, which is then called PXI (PCI eXtensions for Instrumentation). The ending 'XI' is generally used for denoting extensions of buses to instrumentation, for example, there is for VME the VXI bus and for LAN (Ethernet, see below) the LXI bus.

Ethernet. For large experiments the Ethernet standard has a particular importance for the interconnection of distributed processor systems (farms). For example, the trigger–DAQ sequence in fig. 18.2 shows that at higher trigger levels the data are processed and filtered on *processor farms*. It is beneficial to resort to commercial solutions for a *local area network* (LAN) distributing events to farm nodes.

The dominating LAN technology is Ethernet, which was introduced in 1980 and was standardised in 1983 under IEEE 802.3 [553]. The standard describes a whole family of computer networks with specifications for hardware and protocols [987]. The currently specified transfer rates are 10 Mbit/s, 100 Mbit/s (*Fast Ethernet*), 1 Gbit/s (*Gigabit Ethernet*), 10, 40 and 100 Gbit/s. Correspondingly fast are the switches for the distribution of data (*gigabit switches*).

A priori Ethernet is not 'deterministic', meaning that it has no fixed latencies, which makes it as such not suitable for real-time applications. While this excludes in most cases the Ethernet use at the lowest level, that is, the interface to the detector, this restriction is less relevant for the distribution of data to processors. For data distribution it is more important that the maximally possible latency period is limited and that latency fluctuations can be evened by buffer storages.

By means of additional specifications Ethernet can become capable of real-time operations, for example fulfilling the requirements of industrial automation and control of distributed systems, typically in the range of milliseconds (see e.g. [1009]). A development of a fully deterministic Ethernet with a timing accuracy below 1 ns is known under the project name *White Rabbit* [888]. With such a system the time-critical synchronisation of experiment and accelerator components that are distributed over wide areas is possible [569]. Real-time applications are also supported by LXI (*LAN eXtensions for Instrumentation*) for interconnecting measurement devices using Ethernet.

In addition LXI–web interfaces for measurement devices render internet-based control and monitoring possible.

Characterisation of a bus or a network. Speed and capacity of a bus or a network are determined by the latency and the bandwidth.[5] The latency is the time needed by an information to arrive from the sender at the receiver. This includes contributions from pure cable propagation delays, but also delays in network components that serve the data preparation according to the protocol and the data distribution. In a large experiment, usually the cable delays on the first trigger level are most critical. At a typical signal speed of 5 ns/m a signal already needs 0.5 µs for 100 m cable length, which has to be compared to 1–3 µs for a fast trigger decision. The latency is also determined by the data transfer protocol, for example, whether the data bits are sent serially or in parallel, in simplex or duplex mode (transfer in one direction or in both at the same time) or whether a *handshake* protocol requires that a transfer has to be acknowledged by the receiver. In particular the latter can limit the maximal reach of a connection or it may require an extension via a *repeater*.

The (digital) bandwidth of a network is the information which flows from sender to receiver per time, measured for example in bit/s or byte/s. At a given clock frequency for the signals the bandwidth can be improved by increasing the number of parallel lines. For example, a backplane bus transfers many bits in parallel, while bits are serially sent according to the Ethernet standard. In the serial case the assignment of bits to data words has to be controlled by the protocol (which adds to the latency). For fast applications at the detector front-end one would try to reduce the protocol overhead to a minimum. Safety and reliability of data transfer requires more overhead (like acknowledgement of received data). The difference between the physical bandwidth and the actually achieved one may vary between some 10% and several 100%.

18.3 Trigger systems

18.3.1 Classical trigger scheme

A *classical* trigger scheme is shown in fig. 18.1. Signals of a 'fast' detector are logically combined yielding a *pretrigger* if a candidate event is recognised. The pretrigger stops the experiment and, if applicable, starts the digitisation of the pending data. The data flow has to be delayed so that a more complex trigger logic can decide if the converted data should finally be read out into an online computer. The limitations of such a system are:

- At some number of channels the amount of necessary delay cables exceeds a practical limit.
- If, however, the number and/or the lengths of delay cables has to be limited the trigger logic in general has to be less sophisticated than might be desirable to obtain sufficient discrimination power.
- For high-rate experiments the dead times (section 17.9) introduced by the interruptions of data recording due to the trigger latencies could strongly diminish the efficiency.

[5] Here the term 'bandwidth' refers to the *digital bandwidth* of digital information transfer. In contrast the theoretically possible bit rate is determined by the *analog bandwidth*, that is, the frequency spectrum of the system.

18.3.2 Requirements on a trigger

With the continuously growing capabilities of electronics and data processing experiments have tended to become ever more complex. At the beginning of particle physics, the experiments detected single tracks using magnet spectrometers with very limited solid angle coverage (a few millisteradian). For example, in electron–proton scattering measurements for the determination of form factors and structure functions only the electron was detected and not the recoiling nucleon or nucleus, see fig. 2.9 on page 14. In such a case a trigger could have a very simple structure, for example, an electron trigger could be a coincidence of subsequent scintillation counters and a threshold Cherenkov detector (as in fig. 2.9). Modern experiments at the large accelerators (see table 18.1) feature high rates, large particle multiplicities and solid angle coverage close to 4π steradian. The challenge here is to recognise in the detector as fast as possible characteristic event topologies of the interesting physics processes. This results in high requirements on the trigger systems, as we will elaborate on in the following.

A trigger system has to reduce the input rate of events to the storage rate capability. In this procedure the interesting events should be stored without losses and the uninteresting ones should be rejected early and efficiently. The trigger decision should be fast in order to avoid dead times and thus the loss of data.

To give an example we take the relatively simple situation described as case (a) in section 17.9.1: events, randomly distributed in time and fixed dead times per event (i.e. not updated if a new event comes during a dead time). To start with we refer to (17.67) but change the notation substituting

$$n \to \nu_{in},\ m \to \nu_{meas},\ \tau \to \Delta t_{dt} = \Delta t_{trig} + \epsilon_{RO}\,\Delta t_{RO}$$

with the definitions

ν_{in}	input rate, depending on luminosity (i.e. beam rates and/or target densities), cross section and background;
ν_{meas}	triggered event rate;
Δt_{trig}	time necessary for the trigger decision;
Δt_{RO}	time necessary for read out (depends on number of channels, parallelism, ...)
ϵ_{RO}	average fraction of measured events that lead to a readout
Δt_{dt}	average dead time for trigger and readout.

Then the dead-time portion of the total time of data recording is, according to eqs. (17.65)–(17.68), given by

$$\eta_{dt} = \frac{\nu_{in} - \nu_{meas}}{\nu_{in}} = \frac{\nu_{in}\,\Delta t_{dt}}{1 + \nu_{in}\,\Delta t_{dt}} = \nu_{meas}\,(\Delta t_{trig} + \epsilon_{RO}\,\Delta t_{RO})\,, \tag{18.1}$$

with $\Delta t_{dt} = \Delta t_{trig} + \epsilon_{RO}\,\Delta t_{RO}$. If the product $\nu_{in}\,\Delta t_{dt}$ is much smaller than 1 (dead time per event small compared to the time between events) then the dead-time fraction is just $\nu_{in}\,\Delta t_{dt}$, else if $\nu_{in}\,\Delta t_{dt}$ is large the dead-time fraction approaches 1, $\eta_{dt} \to 1$. In the following we want to discuss what to do if the dead-time fraction becomes large, possibly close to 100%.

18.3.3 Trigger architecture

Special trigger architectures have been developed which reduce or even avoid dead times of trigger–DAQ systems. Such architectures are particularly important for high rate experiments.

Trigger level. Even if trigger and readout latencies are relatively long, dead-time-free systems can be achieved. As a first step the decisions are divided into levels where the event rates are reduced stepwise with an increasing complexity and precision of the decision algorithms. If the input rate ν_{in}^i of the ith level becomes smaller than the rate at the $(i-1)$th level the decision time Δt^i (level i latency) can become correspondingly longer.

Quantitatively one has to consider different cases for the product $\nu_{in}^i \, \Delta t^i$:

(a) $\nu_{in}^i \, \Delta t^i \ll 1 \;\Rightarrow$ The experiment can be stopped for the trigger decision and the dead-time fraction is $\eta_t^i = \nu_{in}^i \, \Delta t^i \ll 1$.

(b) $\nu_{in}^i \, \Delta t^i \lesssim 1 \;\Rightarrow$ The system can work on the events in a dead-time-free mode only if the events occur in about equal time intervals. This can be achieved using an intermediate storage which continuously supplies events. Such an intermediate storage is called *derandomising buffer*.

(c) $\nu_{in}^i \, \Delta t^i \gg 1 \;\Rightarrow$ decisions have to be paralleled.

Parallelisation of trigger decisions. With $\nu_{in}^i \, \Delta t^i \leq n$, where n is a number > 1, we obtain three possibilities of parallelisation and combinations thereof:

1. taking n trigger decisions in parallel;
2. averaging decisions over n events;
3. dividing the decision algorithm in n time slices and pushing an event by means of a 'trigger pipeline' through the algorithm;
4. combining steps 1 to 3.

For example, at the LHC the input rate for the first trigger level is $\nu_{in}^1 = 40\,\text{MHz}$ (= beam crossing frequency). With typical decision times on the first level of about $\Delta t^1 = 1\,\mu\text{s}$ there are about 40 new inputs during a decision time. Hence a trigger pipeline according to item 3 of the above list has to decompose the trigger algorithm into at least 40 separated steps such that at any given time $n \geq 40$ events (= beam crossings) are simultaneously worked on.

18.3.4 Data buffer

During the trigger decisions the data of the events that are not yet rejected have to be kept available in data buffers. The options are:

- Delay cables. Problem: not practicable for a large number of channels and/or for long delay times.
- Electronic delays, for example by sampling a pulse and keeping the maximum (*sample and hold*, section 17.4). Problem: a continuous clocking is not possible.
- Data pipelines which can be realised as integrated circuits (fig. 18.4):
  - analog pipelines: analog storage cells in *switched capacitor arrays*.
  - Digital pipelines: shift registers, FIFO (*first-in-first-out*), ring buffers or 'dual-port memory' (e.g. in flash ADCs, section 17.7.1). Problems: sufficiently large dynamic

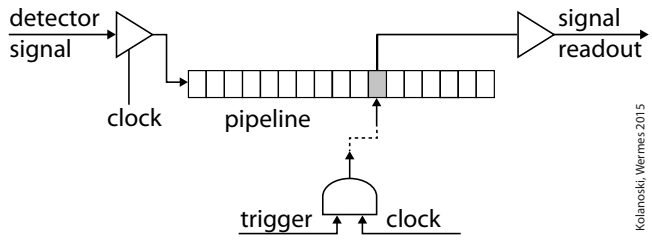


Fig. 18.4 Principle of a pipeline. Signals of a detector channel are written into a device with multiple storage cells under the control of a clock, which is synchronised with the accelerator frequency in the case of periodic machines, and are then pushed into the following cell at the next clock pulse. Then the storage contains the temporal sequence of signals in this channel within a time period which corresponds to the depth of the pipeline. Through a synchronised trigger pulse the storage cell that contains the triggered event can be accessed and read out.

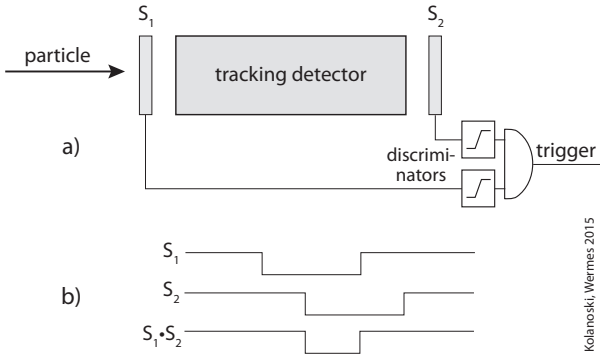


Fig. 18.5 (a) Scheme of a trigger with two counters which are hit in coincidence. Such counters are often organic scintillators (section 13.5.1) which can provide temporal precision of the coincidences of few nanoseconds. (b) Discriminator signals of the counters and the resulting coincidence signal yielding the trigger.

range is costly (especially for calorimetry), power consumption is larger than for analog storages.

18.4 Realisation of triggers

In this section we give a short overview on how one can select 'interesting' events and reject uninteresting events and background by employing logical combinations of detector signals. The principal concept goes back to the development of the 'coincidence method' by W. Bothe and others in the 1920s (section 2.1, fig. 2.2). The method implies the electronic search for temporally coincident signals which lie above a threshold and fulfil further criteria. An often requested criterion is a certain spatial pattern of the corresponding signal sources, for example, a signal pattern generated by a particle that passes several layers of a tracking detector (see fig. 18.6).

18.4.1 Simple counting coincidences

In the simplest case a trigger is formed from an arrangement of two or more counters, for example Geiger counters as in figs. 2.2(a) and 6.2(b) or scintillation counters as in figs. 7.13 and 7.25. The scheme with two counters and a tracking detector is again depicted in fig. 18.5. In such an arrangement the area and distance of the counters define the solid angle region for a triggering particle. A somewhat more complex trigger logic is formed from the counters in figs. 16.8 and 13.19, where in addition to counter

coincidences one explicitly requires that so-called *veto counters* are not hit within a certain time interval.

If two counters (as in fig. 18.5) feature uncorrelated, uniformly distributed background signals with rates ν_1 and ν_2 then the rate of random coincidences is given by

$$\nu_{12} = \nu_1 \, \nu_2 \, \Delta t_{12} \, . \tag{18.2}$$

Here $\Delta t_{12} = \Delta t_1 + \Delta t_2 - \Delta t_{\text{coin}}$ is the time span within which an overlap of both signals leads to a coincidence. This time span is the sum of both signal lengths Δt_1, Δt_2 minus the overlap time Δt_{coin} which is minimally required for yielding a coincidence. If, for example, the individual background rates are $\nu_1 = \nu_2 = 1\,\text{kHz}$ with $\Delta t_{12} = 10\,\text{ns}$, then the resulting background rate is 0.01 Hz. Each additional (independent) counter i in the coincidence condition lowers the output rate by a factor $\nu_i \Delta t_i$. A high background can therefore be optimally suppressed by defining a coincidence condition that is temporally as sharp as possible and has as many contributions from different detector components as possible.

18.4.2 Logic electronics: NIM standard

The logic of trigger decisions can be configured using standard electronics. Already since the 1960s the NIM standard has commonly been used in nuclear and particle physics [737]. This standard defines a system of *crates* with plug-in modules for various electronic tasks, cables, connectors and the levels of logic signals (see section 17.7.4).

The crates are similar to those for the readout systems described in section 18.2.1. However, in general the modules in a NIM crate cannot communicate with each other via the backplane. The backplane usually carries only utility lines (mostly DC voltages ± 6, ± 12 and ± 24 V) and there is no connection to an external computer. A large variety of NIM modules is available, for example discriminators, signal generators and logic modules for different binary functions. On the other hand analog-to-digital converters (ADCs, TDCs, ...) became uncommon for NIM crates but are more naturally deployed in readout systems with digital communication buses (section 18.2.1).

The original NIM standard for cables is a coaxial cable with a wave impedance of $50\,\Omega$. Meanwhile other cable types are also permitted, in particular those that allow for a higher packing density (like *twisted pair* cables, see section 17.8 on page 761). The NIM level for the logic values '0' and '1' are defined as currents corresponding to voltages of about 0 and $-0.8\,\text{V}$, respectively, across the required 50-Ω termination. Besides NIM levelsother standardised logical levels may also be accepted in a NIM system (e.g. TTL, ECL, ...). Table 17.2 in section 17.7.4 gives an overview of standard logical levels.

In the previous examples the trigger should only recognise a single particle. In detectors with large solid-angle acceptance, measuring many-particle final states, the requirements on a trigger are correspondingly larger. The complex requirements of trigger tasks found a variety of inventive solutions, which we will only exemplarily discuss in the following.

18.4.3 Pattern recognition on trigger level

In general a detector consists of a number of detector cells, which can be as many as of the order of 10^8. An event leaves a certain pattern of hit cells (see fig. 18.6). A trigger has the task to decide whether the event should be accepted by comparing the hit cell ensemble with each of the possible patterns.

A naive solution of this task would be to scan through all possible cell combinations and compare them to a list of predefined patterns. For a large cell number this would obviously very soon lead to an unacceptably large number of combinations. Take as an example a tracking detector with (only) four layers, each with 100 cells (e.g. drift chamber cells or silicon strip electrodes). The number of combinations of one cell in each layer is then 10^8 and with twice the number of layers it would be 10^{16}, illustrating that this procedure quickly becomes very inefficient with an increasing number of cells.

Information reduction. A basic rule in pattern recognition says that the number of possible combinations should be reduced to those absolutely necessary as early in the trigger chain as possible. Respective methods of information reduction include:

- Reduction of the resolution for trigger purposes by combining detector cells as long as an interesting signature can still be recognised. For example, one can combine neighbouring cells of a tracking detector or a calorimeter such that track or jet multiplicities, respectively, can still be estimated.
- Restriction to 'roads' or *search windows* in the detector. Only those combinations of trigger cells in different detector layers are considered which could lead to physically possible patterns. For example, calorimeter cells are combined to form trigger towers which point to the interaction point (in case of an accelerator experiment), as shown in fig. 15.17 for the D0 detector.
- Exploiting neighbourhood relations of local hits. For example, in tracking detectors one often combines hits in two or three layers to form doublets or triplets of track elements if they fulfil the condition of being tangent to a possible track.
- Iterative refinement of search windows using the Kalman filter method corresponding to fig. 9.9 in section 9.3.6. Starting at a point in parameter space (e.g. of a track) one searches in the nearest neighbourhood for points which could match a pattern corresponding to a model for the searched object. At each step the search window can be narrowed.

The examples show that the configuration of a trigger system should be considered already in the design phase of a detector because trigger logics can be simplified by suitable construction measures. A good example for this principle is the design of trigger towers in fig. 15.17. Similarly one should consider in the design phase how the efficiency of a foreseen trigger can be determined. This is usually solved by planning for several redundant triggers, enabling a mutual monitoring. For example, in an event that was triggered by the calorimeter one can check if existing tracks have been recognised by a likewise installed track trigger.

The recognition of complex patterns by a trigger system on a time-scale of typically microseconds requires considerable electronic effort. Depending on the number of trigger inputs (detector or trigger cells) the number of combinations to be tested can be very high. If the trigger has several trigger levels the L1 level is usually restricted to a few robust decisions which sufficiently reduce the rate such that more complex decisions are possible on higher levels. For example, in LHC experiments, calorimeter and muon triggers are most suited for the L1 level. The calorimeter triggers require large energy deposits in local clusters as signatures for jets, and muon triggers require muons with large transverse momentum relative to the beam axis.

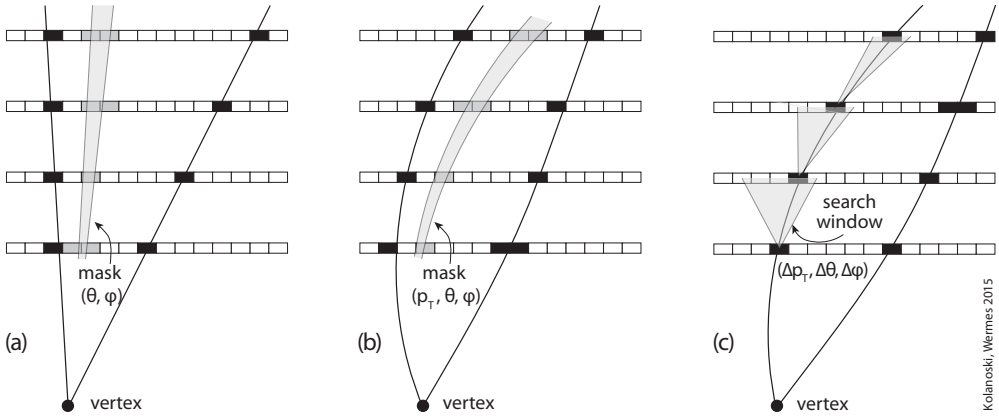


Fig. 18.6 Scheme of a track trigger. The black cells are the measured patterns which are in this case created by two tracks. (a) Straight tracks; (b, c) tracks curved by a magnetic field. In (a) and (b) a mask is stored for each possible pattern of single tracks. The light-grey regions are examples for 'roads' which in the case of straight tracks cover a certain direction region and in the case of curved tracks cover a direction–momentum region. The grey cells are assigned to a mask covering a road. The regions of different masks have to overlap in order to achieve full trigger efficiency. In (c) the scheme of a trigger based on the Kalman filter method is shown. Starting from a reference point (a hit cell) in the first detector layer one searches for a hit in the next layer in a window which—under certain conditions—could have been crossed by a track which goes through the reference point. A condition for the size of the window could be that the track comes from a known interaction region (usually for accelerator experiments). If in the next layer a hit is found the prediction for the next layer can be improved and the search window can become narrower.

18.4.4 Track trigger

Here we discuss track triggers on a low trigger level where the time constraints are in the range of microseconds and where complex pattern recognition is only possible with very specific electronics.

Concepts. The tasks to be fulfilled are schematically displayed in fig. 18.6. A track leaves a hit pattern in the detector. The trigger has now to find out whether a certain pattern can be identified as due to one or several tracks.

Mask processor. A common trigger algorithm employs masks which are compared to a measured hit pattern (fig. 18.6(a, b)). Each mask corresponds to tracks in a certain covered region in parameter space, here given by directions and momenta, the latter if there is a magnetic field. The size of the region depends on the desired resolution which in turn has to be adapted to the particle density in the detector. Many different, partly overlapping masks have to be defined in order to find preferentially true tracks with high efficiency. For a fast trigger decision it is advantageous if the assignment of masks to patterns can be processed in parallel steps, for example separately for detector sections.

Kalman filter. A quite different approach is the Kalman filter method which is depicted in fig. 18.6(c) (see also section 9.3.6). Starting from a seed point the parameters of a track model are successively adjusted to match hits in the subsequent detector

layers. With the found hits the track parameters are then improved and the search window in the next layer can be refined. The stepwise procedure and the fact that at each step the trigger processes only local hit data make the Kalman filter algorithm very suited for a trigger that works in a pipeline modus.

An example for a Kalman filter application is the fast track trigger of the HERA-B experiment [144]. Within the L1 latency of only 12 µs and in an environment with very high particle density, tracks from lepton pairs are found and their invariant mass is determined. The mass resolution is so good that leptonic decays of J/ψ mesons can be selected with reasonable purity and efficiency.

Electronic logic devices. If the trigger algorithms necessary to sufficiently suppress background cannot be run with an acceptable latency of the trigger system then the individual decision steps have to be parallelised according to the options listed on page 808. The solutions are manifold and are usually realised employing the fastest electronic components available at a time. Electronic components for such triggers are for example:

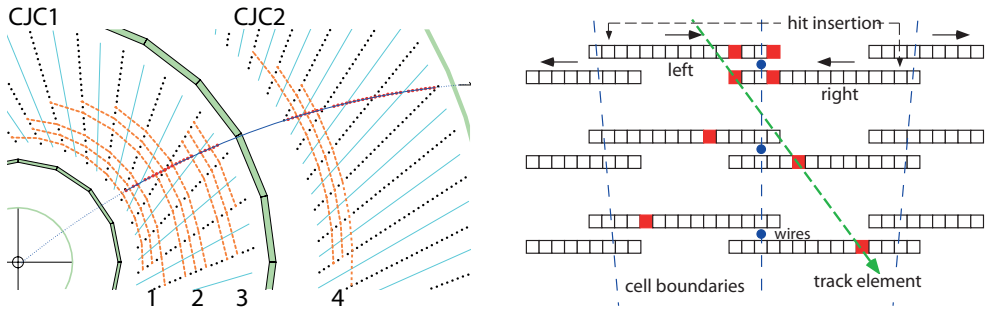
FIFO (*first-in-first-out shift register*): Data are filled into a chain of storage cells, pushed through the chain on clock pulses and finally drop out after a maximal number of steps, the storage depth. If the shifting follows a continuous clock each storage cell carries besides the storage content also the information of the time when the cell was filled. Shift registers are for example employed as data pipelines for the bridging of trigger latencies (see fig. 18.4).

LUT (*look-up table*): A stored table containing pre-calculated results of time-consuming calculations. For example, such a table can be used to assign to a measured track pattern the track parameters like direction angles, intercept and curvature. In programmable logic circuits the logic functions are often stored as LUTs, a procedure which offers faster processing as compared to a software evaluation and at the same time a larger flexibility as compared to a hard-wired implementation.

CAM (*content addressable memory*) or associative memory: In contrast to a RAM (*random addressable memory*), which is a more standard type of memory where an address is assigned to a storage cell, in a CAM the input patterns are compared to stored contents. A match between input and content can be signalled by setting a bit or it can initiate the transfer of further information related to the pattern (*encoded output*). While in a RAM a match between a data word with a storage content has to be serially searched for, in a CAM the comparison is done in parallel, rendering the solution of search tasks particularly fast. For example, in data networks CAMs are used to assign destination addresses to data packages arriving at a node. In particle physics associative memories are used for the pattern recognition of tracks and clusters. The principle of the electronic realisation is described in [349] and the application in a trigger of the CDF vertex detector in [348].

DSP (*digital signal processors*): Processor chips which are specialised on the fast processing of signals with fast input and output of data.

FPGA (*field programmable gate array*): A freely programmable integrated circuit, a variant of programmable logic circuits (PLDs) with highest flexibility for trigger algorithms. An FPGA comprises gates, memories, look-up-tables, processors (mostly DSPs), multiplexers and standard connectors for input and output. In par-



(a) (b)

Fig. 18.7 The 'Fast Track Trigger' of the H1 experiment [143]. (a) Cross section of a quadrant of the central jet chamber perpendicular to the colliding beams. The signal wires and cathode planes are indicated by dots and lines, respectively. The *trigger layers*, each consisting of three wire layers, are numbered 1–4. (b) Delineation of a *trigger cell* in one trigger layer (three wire layers) with the shift registers that marks the position of hits. For each shift register there is a mirrored register because of the left–right ambiguity. Shift registers of neighbouring cells cover overlapping regions in the drift chamber in order not to lose tracks in the transition region. Track segments are nearly linear links between hits in the three wire layers. Source: DESY/H1 Collaboration.

ticle physics the fast triggers are mostly based on the logical processing of the signals in FPGAs which are correspondingly configured. The free programmability by so-called *firmware* enables a fast adaptation to the experimental conditions.

Example: 'Fast Track Trigger' of H1. As a typical example for a track trigger we consider the *Fast Track Trigger* (FTT) of the H1 experiment at HERA[6] [143].

At HERA the time period between beam crossings is 96 ns. For the first level trigger H1 allows for a latency of 2.3 µs during which the data are stored in a pipeline. At this level the FTT coarsely reconstructs tracks in two dimensions while at the second level with a latency of 23 µs all tracks are reconstructed with high precision in all three dimensions. This allows one to identify particle resonances at the third trigger level within not more than 100 µs.

The H1 FTT uses tracks in the central jet chamber (CJC = *Central Jet Chamber*) which consists of two concentric cylindrical drift chambers (CJC1, CJC2) with a total of 56 wire layers in a jet chamber structure as in fig. 7.41(c). Figure 18.7(a) shows a cross section of a chamber quadrant.

Trigger concept. The trigger works on 12 of the 56 wire layers which are grouped in 4 *trigger layers* with 3 wire layers each. The trigger layers are marked in fig. 18.7(a). In a first step track segments, which are composed of signals from the 3 wire layers of a trigger layer, are searched for in parallel in each trigger layer. The track segments are found by comparing the signal patterns with masks which are identified by mean values of the covered corresponding inverse transverse momentum $\kappa = 1/p_T$ and the azimuth angle ϕ ranges. In a further step the track segments found in the four trigger layers with the same or similar (κ, ϕ) values are combined to form a track if at least two segments match. Finally an L1 trigger is issued if certain predefined conditions

[6] There are various other examples with similar requirements on event rate and trigger latency, for example the *Extremely Fast Tracker* (XFT) of the CDF experiment [939].

on the multiplicity of found tracks and their topological and kinematical relations are fulfilled. At higher trigger levels the L1 tracks are refined and then used for more detailed reconstructions [143].

Electronic implementation. The drift chamber signals of the trigger layers are digitised with a 80-MHz clock frequency by a flash ADC (FADC, section 7.10.6). The FADC data are transferred to an FPGA where a digital signal processor executes a charge–time analysis (Q–t analysis). For signals above a noise threshold the data are written into a shift register with a clock frequency of 20 MHz. The reduced clock rate is achieved by combining four FADC cells. The shift register then contains the time sequence of the hits relative to a beam crossing time. With a known space–drift-time relation (section 7.10.7) and for a given beam crossing time each storage cell contains the position information of the hit. Thus for each wire the drift distance from cathode to anode is mapped onto the cells of the shift register, as depicted in fig. 18.7(b). Because of the left–right ambiguity at the wire (see section 7.10.4) there are always two registers containing the mirror-pairs of positions. Further shift registers account for overlaps of neighbouring cells. Possible combinations of hits in the three wire layers (see fig. 18.7(b)) are transferred by the FPGA as a pattern onto a CAM. The CAM then issues the (κ, ϕ) values of the pattern if it matches one of the track patterns stored in the CAM. Since the drift time is longer than the time between beam crossings the track search has to be restarted at each beam crossing such that several searches have to be processed in parallel. For different beam crossings a cell in the shift register is mapped onto another position in the drift chamber.

For the linking of track segments *histogramming* is employed. In the FPGA the (κ, ϕ) values of the found segments are inscribed into a two-dimensional histogram with 16×60 pixels.[7] A fast search algorithm (*peak finder*) looks for histogram cells with at least two entries. The corresponding segments are then linked to a track. At the same time the left–right ambiguity is resolved because segments with a wrong left–right assignment in one or more cells do not match (see fig. 7.41(c)). In addition segments in general match only for the correct beam crossing time which is thus determined by the resulting trigger as well. The linking process is executed within 130 ns only. The total time required for the FTT on the L1 level is 0.5 µs.

18.4.5 Calorimeter trigger

Calorimeter triggers are particularly suited to suppress background because background events typically do not deposit much energy. In collider experiments events with high *transverse energy* are a signature for interesting reactions. The term 'transverse energy' (E_T) denotes the transverse component of a vector that points from the interaction point to the energy deposition and has the magnitude of the deposited energy (hence corresponding to the transverse component of a momentum if one neglects masses). Formally one has $E_T = E \sin\theta$, where θ is the polar angle of the cluster with the energy E. The calorimeter energy can additionally be divided into electromagnetic and hadronic energy leading to triggers for electrons and for hadrons or jets, respectively. From the measurement of all energies and momenta the *missing transverse energy* E_T^{miss} of an event can be determined, which is an important trigger quantity when searching for exotic processes.

[7] The transformation of the hits in the position space into the space of the parameters (κ, ϕ) corresponds to a Hough transformation (see e.g. [691]).

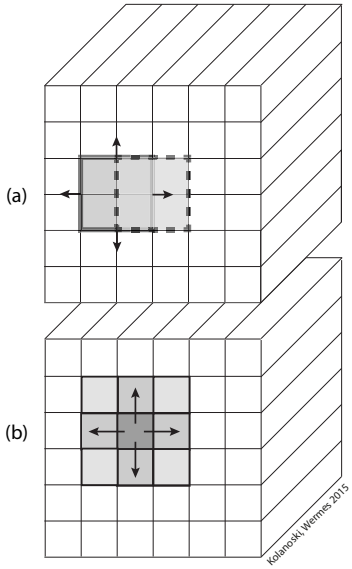


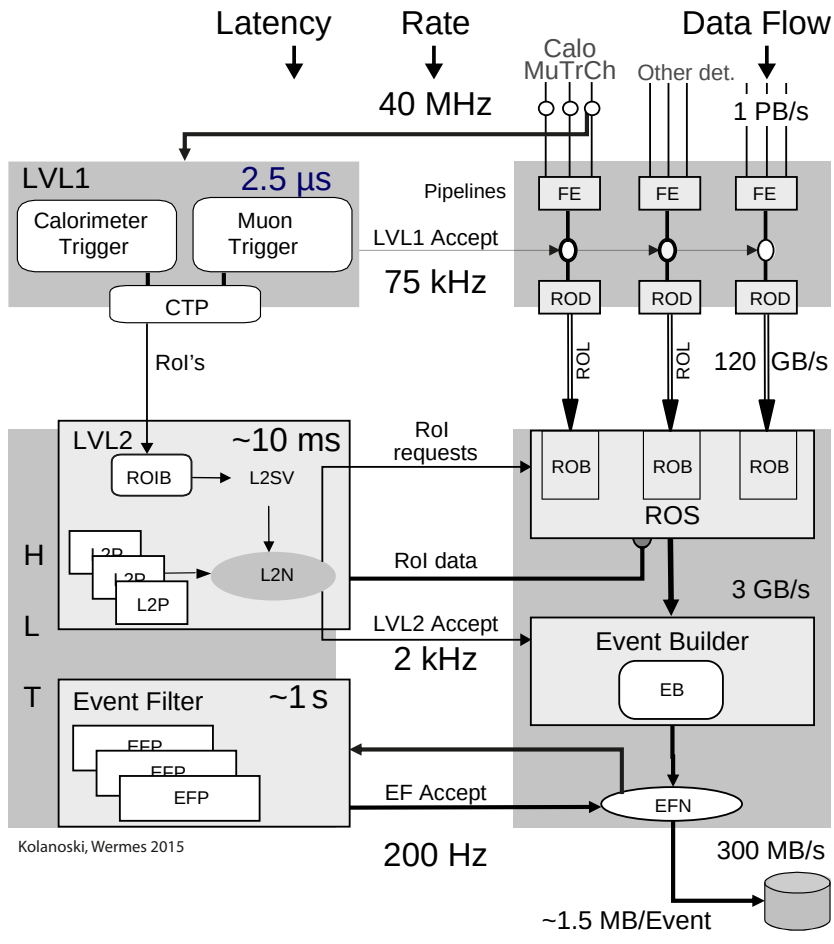
Fig. 18.8 Fast cluster search in a calorimeter for trigger forming. (a) Search in four neighbouring cells for which the sum of energies is above a trigger threshold. The search implies the *sliding window* method where a (4×4) window is slid over the cells in steps of one cell. (b) Search for a single cell which has an energy above a threshold (for the single cell). A cluster is identified if after adding the energies of the neighbouring cells the energy sum lies above a threshold for a cluster. Direct neighbours could be the four cells with a common edge or the eight cells with either a common edge or a common corner.

The search for energy clusters in calorimeter cells to form a fast trigger can proceed about as in fig. 18.8. In (a) the method of *sliding windows* is shown. The energies in four or more cells are summed and the sum is checked for passing a threshold. The search window is shifted such that no inefficiencies occur at the transitions. In (b) the cells are individually scanned and the sum of the energies of the considered cell and their nearest neighbours is checked for passing a cluster threshold. If the calorimeter is also longitudinally segmented then successive cells are usually combined to form towers (*trigger towers*) which point in the direction of the interaction region (see fig. 15.17).

18.5 Example: a trigger–DAQ system at the LHC

To conclude this chapter we want to present as an example of a trigger and data acquisition system in a high-rate experiment the TDAQ system of the ATLAS experiment at the LHC. The system is displayed in fig. 18.9 [4]. At a beam crossing rate of 40 MHz, corresponding to a time of 25 ns between crossings, about 25 events per beam interaction (design luminosity) and high particle multiplicities the input rate of 40 MHz has to be reduced by a factor of more than 10^5 to about 200 Hz. This is only possible with a multi-level trigger with parallelisation of the trigger decisions and an intermediate buffering of the data. The architecture of the TDAQ system therefore follows a three-level trigger hierarchy, similar to the one sketched in fig. 18.2.

The left hand side of fig. 18.9 shows the trigger sequence with the respective latencies and the consecutive rate reduction and the right-hand side the data flow. The first trigger level (LVL1) is a hardware trigger which searches for energy clusters in the calorimeter and tracks in the muon trigger chambers. The *high level triggers* (HLTs) consist of processor farms which stepwise filter the events. The level LVL2 has access only to data of the *regions of interest* (RoIs), denoting regions in which signals have contributed to the LVL1 trigger. Finally, on the third trigger level, called 'event filter', all data prepared by the 'event builder' are available.



Legend:

Calo	Calorimeter	MuTrCh	Muon Trigger Chambers
LVL1/2	Level 1/2 Trigger	FE	Front End Pipelines
CTP	Central Trigger Processor	ROD	Readout Driver
RoI	Region of Interest	ROIB	RoI Builder
ROL	Readout Link	ROB	Readout Board
L2SV	L2 Supervisor	ROS	Readout System
L2P	L2 Processor	L2N	L2 Network
HLT	Higher Level Trigger	EF	Event Filter
EFP	EF Processor	EFN	EF Network

Fig. 18.9 Trigger and data acquisition system of the ATLAS experiment: on the left the three trigger levels and on the right the corresponding data flow. See the detailed description in the text.

First level. On this level the data of each detector channel are filled at a clock rate of 40 MHz into a pipeline with a depth of 2.5 µs, corresponding to the latency of the trigger. The trigger system obtains with the same clock rate data from the calorimeters and the muon trigger chambers. The muon trigger is formed by signals from *resistive plate chambers* (RPCs, see section 7.7.3) which deliver fast signals with position and time information. The first-level trigger provides candidates for muons, jets and electrons (respectively photons) for several different energy or momentum thresholds as well as the total energy and the missing transverse momentum, derived

from the sum of the muon momenta and the magnitudes and directions of the energy clusters. Trigger objects are labelled by their pseudo-rapidity η (eq. (2.8)) and the azimuth angle ϕ before they are transferred to the *central trigger processor* which combines the information to make the LVL1 trigger decision.

The events are accepted at a rate of about 75 kHz. The corresponding data are transferred from the pipelines via optical links to the *readout drivers* (RODs) which are housed in VME crates (section 18.2.1). There the data are processed and compressed before they are written into the *readout buffers* (ROBs), which use a PCI bus (section 18.2.2). Simultaneously the (η, ϕ) values of the RoIs are calculated, which are used on the second level, and are given to the *RoI builder*.

Second level. On this level the construction of the RoIs is refined on a processor farm. A *supervisor* coordinates the distribution of the individual events for parallel processing via Ethernet links. The LVL2 latency is about 10 ms and the output rate about 2 kHz.

Third level. If an event is accepted by LVL2 then the data are passed to the *event builder*. On a processor farm (physically the same as the LVL2 farm, called 'HLT farm') the events are fully reconstructed and the results are filtered yielding a reduction to 200 Hz. At this rate the events, featuring average sizes of 1.5 MB, are written to a permanent storage medium.

Appendix A
Dosimetry and radioactive sources

In this appendix short overviews of the most important measures and units in dosimetry as well as of some radioactive sources frequently used for detector testing will be given. More details and references can be found, for example, in the review article 'Radioactivity and radiation protection' in [746].

A.1 Dosimetry

The activity of a radioactive source is measured in units of becquerel (1 Bq = 1/s). The measure of radiation absorption is the dose D which is the energy E absorbed in a volume with mass M and has the unit gray = Gy (obsolete: rad):

$$D = \frac{E}{M} \qquad [D] = \mathrm{Gy} = \mathrm{J/kg} = 100\,\mathrm{rad}\,. \tag{A.1}$$

The *relative biological effectiveness* (RBE) of absorbed dose depends not only on the particle type and energy but also, for example, on the biological cell type, the dose rate and its distribution in time and space. The RBE of β and γ radiation, being the same and relatively energy independent, is used as reference with RBE = 1. In contrast, α radiation, fission fragments and heavy ions can have RBE values of up to 20.

To obtain a comparable measure for radiation damage of an organ or tissue, independent of radiation composition or other factors, the *equivalent dose* H is defined. It is obtained by summing the absorbed doses D_R of the different absorbed radiation types R, each multiplied by the so-called radiation weighting factor w_R:

$$H = \sum_R w_R \times D_R\,. \tag{A.2}$$

The radiation weighting factors w_R are based on the respective RBE values and are fixed by official regulations. Since w_R is dimensionless the equivalent dose has the same dimension as the dose proper. However, for reasons of differentiation the equivalent dose is given in units of sievert (1 Sv = Gy × RBE, obsolete: rem). The *effective dose*, which is the relevant quantity for radiation damage in human bodies, also accounts for the different effectiveness of radiation in different organs and represents an average of the load on the individual organs and tissues of a human.

Natural radioactivity leads to an average equivalent dose for humans of about 2 mSv per year and in Central Europe additionally about the same amount from medical sources must be added. The lethal dose for humans exposed to total body irradiation is 2–5 Sv.

Table A.1 Compilation of radioactive sources frequently used for detector studies (from [746]). The list contains nuclides with their half-lives $t_{1/2}$ and their dominant decay modes, here β and α decay and *electron capture* (EC). For the monochromatic radiation (α, β, γ) the energies and for the continuous β radiation the endpoint energy are given. The relative percentage of each decay is displayed as well, where for cascade decays the sum of the frequencies can amount to more than 100%. Partly also relevant decays of daughter nuclides are quoted. Usually the α and β decays are accompanied by successive γ decays. Electron capture (EC) is followed by X-ray radiation due to the transition of an outer electron to the former level of the captured electron, as for ^{55}Fe, or the radiation of the transition is directly transferred to an electron which is then monochromatically emitted (*conversion line*, denoted by e^-), as for ^{207}Bi. The energies of the transitions which are most important for detector tests are highlighted in bold.

			α, β		γ		
Nuclide	$t_{1/2}$ (years)	Type	energy (MeV)	(%)	energy (MeV)	(%)	Main use
$^{55}_{26}$Fe	2.73	EC			Mn K X-ray:		
					0.00590	24.4	keV-γ lines
					0.00649	2.86	
$^{57}_{27}$Co	0.744	EC			0.014	9	
					0.122	86	
					0.136	11	γ lines
$^{60}_{27}$Co	5.271	β^-	0.316	100	**1.173**	100	MeV-γ lines
					1.333	100	
$^{90}_{38}$Sr	28.5	β^-	0.546	100			MeV-β
$\longrightarrow$ $^{90}_{39}$Y		β^-	**2.283**	100			
$^{106}_{44}$Ru	1.020	β^-	0.039	100			
$\longrightarrow$ $^{106}_{45}$Rh		β^-	**3.541**	79	0.512	21	MeV-β
					0.622	10	
$^{109}_{48}$Cd	1.26	EC			0.088	4	
					Ag K X-ray:		
					0.022	83	keV-γ lines
					0.025	15	
$^{137}_{55}$Cs	30.2	β^-	0.514	94	**0.662**	85	γ line
			1.176	6			
$^{207}_{83}$Bi	31.8	EC	**0.481 e^-**	2	0.569	98	conversion
			0.975 e^-	7	1.063	75	lines
			1.047 e^-	2	1.770	7	
$^{228}_{90}$Th	1.912	6 α:	**5.341–8.785**		**0.239**	44	
		3 β^-	0.334–2.246		**0.583**	31	α, γ lines
					2.614	36	
(part of the thorium decay chain with α and β^- decays, ending at $^{208}_{82}$Pb)							
$^{241}_{95}$Am	432.7	α	**5.443**	13	**0.060**	36	α, γ lines
			5.486	85			

A.2 Radioactive sources

Radioactive sources are often employed for testing detectors because their use is usually less involved and easier to evaluate as compared to tests using accelerator beams. Some of the frequently used sources are listed in table A.1. Examples for detector signals obtained with radioactive sources are displayed in figs. 3.41, 8.34, 8.68, 13.6, 13.13, 13.21, 13.22 and 17.29.

The α and γ emitters exhibit monochromatic radiation which can be used for the calibration of detector signals. In contrast, the β decay delivers a continuous spectrum up to a maximum energy (endpoint energy). When the nucleus captures an electron from the K shell (EC, *electron capture*) either the free position will be taken by an electron from a higher shell under emission of an X-ray, or the liberated energy will be internally transferred to an electron which will be emitted monochromatically (conversion line).

In the following we summarise the possible applications of the sources listed in table A.1:

α radiation: The isotopes ^{241}Am and ^{228}Th are α emitters providing a high local ionisation density. Because of their short range, the energy deposition of α rays is relatively well localised which can be exploited for surface investigations. For example, the α rays of ^{241}Am have a range of about 40 µm in water or organic scintillator and about 17 µm in silicon.

β radiation: β particles in the MeV region, for example from ^{90}Sr and ^{106}Ru, can penetrate not too thick detectors such as single layers of gas-filled wire chambers or thin (200–300µm) semiconductor detectors. In this case they deliver energy depositions which are similar to those of high-energy particles (see section 3.2). In the case of a continuous β spectrum one can assure that a β particle has passed the detector by placing the source in front of the detector and a tagging detector, for example a scintillator, behind the detector. An electronic coincidence between the tested detector and the scintillator would indicate a particle passage.

γ radiation: The monochromatic photons of the listed nuclides ^{60}Co, ^{137}Cs and ^{228}Th are employed for the calibration of calorimeters (see section 15.4.7). The γ lines of ^{241}Am at 60 keV and of ^{57}Co at 122 keV are preferentially used for the calibration of semiconductor detectors. A minimum-ionising particle passing a 300 µm thick silicon layer deposits an energy of 116 keV, which corresponds approximately to the energy of the ^{57}Co photopeak. The ^{55}Fe line (5.9 keV) has similar properties for gaseous detectors because a minimum-ionising particle deposits in argon, for example, an energy of 2.5 keV on a path of 1 cm.

Conversion electrons: The monochromatic electrons of the nuclide ^{207}Bi are employed, for example, for the pulse height adjustment of organic scintillators. The electrons of the lines near 1 MeV have a range of about 4.5 mm and produce in a 5-mm-thick scintillator about the same signal as a minimum-ionising particle.

Appendix B
Weighting potential of segmented electrodes

B.1 Solving for potentials by conformal mapping

The calculation of the weighting potential of an electrode arrangement as in figs. 5.12 and B.1(a) with one electrode segmented into strips of width a (with negligibly small space between strips) requires the solution of the Laplace equation

$$\Delta\phi(x,y) = 0 \qquad \text{for} \quad 0 < y < 1 \tag{B.1}$$

with the Dirichlet boundary conditions

$$\phi(x, y = 1) = 0, \quad \phi(|x| > a/2, y = 0) = 0, \quad \phi(|x| \leq a/2, y = 0) = 1\,. \tag{B.2}$$

We assume the structure to be infinitely extended in the x and z directions and to be independent of z (here z is the third coordinate in the (x, y, z) system and not the complex variable used in the following). The independence from z permits a two-dimensional treatment in the xy plane. Two-dimensional potential problems can be solved using the method of conformal mapping of the complex plane, see the respective literature on mathematical methods in physics and electrical engineering (e.g. [718], chapter 4 and 10, and [700]).

To employ this method we refer to the xy plane as the plane of complex numbers with the mapping $z = x + iy$. A mapping which maps the complex plane $\mathcal{C}_1$ onto the complex plane $\mathcal{C}_2$ is a conformal (angle-preserving) mapping if it is mediated by an analytical function f:

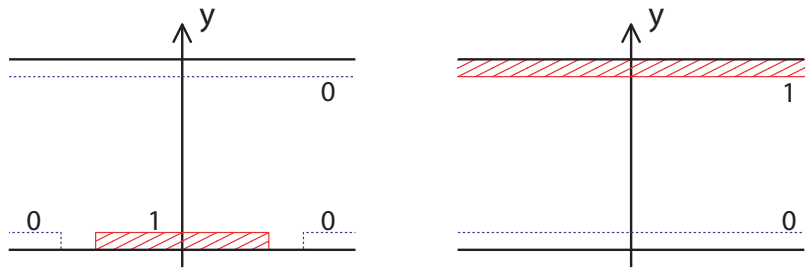
$$z_2 = f(z_1) \qquad z_1 \in \mathcal{C}_1, \quad z_2 \in \mathcal{C}_2\,. \tag{B.3}$$

For a conformal mapping the following argument holds. If $\phi_1(z_1) = \phi_1(x_1, y_1)$ solves the Laplace equation with Dirichlet boundary conditions then also

$$\phi_2(z_2) = \phi_1(z_1(z_2)) \tag{B.4}$$

solves the Laplace equation with the values of ϕ_1 at the points z_1 being equal to the values of ϕ_2 at the points $z_2 = f(z_1)$. This also holds in particular for the corresponding boundary values, a feature which we will exploit for the construction of the conformal mapping of our problem.

Method. The determination of the potential of an arrangement $\boxed{1}$ by conformal mapping onto an arrangement $\boxed{2}$ proceeds as follows:



(a) Geometry [1]. (b) Geometry [2].

Fig. B.1 For arrangement [1] a solution of the Laplace equation (B.1) with boundary conditions (B.2) is searched for. A solution is known for the simple arrangement [2] of an infinitely extended plate capacitor with potentials $\phi(x,0) = 0$ and $\phi(x,1) = 1$. Now the boundary regions in [1] with definite potential values should be conformally mapped onto corresponding regions in [2] with the same potential values.

(1) Search for a simple electrode arrangement [2] for which the Dirichlet boundary problem is solved.
(2) Search for a conformal mapping $z_2 = f(z_1)$ which transforms the boundaries of [1] onto corresponding boundaries on [2] maintaining the same respective boundary values.
(3) By inserting into (B.4) find the desired potential $\phi(x,y) = \phi_1(z_1) = \phi_2(z_2(z_1))$ with $z_1 = x + iy$.

The first step under item (1) is not independent of the second. One would choose for [2] a geometry with a similar topology as in [1] in order to find a suitable conformal mapping.

B.2 Determination of the weighting potential of a strip detector

We want to determine the weighting field of a strip detector, as in fig. B.1(a). As arrangement [2] we choose the infinitely extended plate capacitor with electrodes at $y_1 = 0$ and 1 (fig. B.1(b)), which has a well-known solution for the potential:

$$\phi_2(x_2, y_2) = y_2 \quad \text{or} \quad \phi_2(z_2) = \mathrm{Im}(z_2)\,. \tag{B.5}$$

As the conformal transformation that maps the boundaries of [1] onto the boundaries of [2] in such a way that the strips are mapped onto $y_2 = 1$ and the rest unto $y_2 = 0$, we choose

$$z_2 = \frac{1}{\pi} \ln \left(\frac{\mathrm{e}^{\pi z_1} - \mathrm{e}^{\pi \frac{a}{2}}}{\mathrm{e}^{\pi z_1} - \mathrm{e}^{-\pi \frac{a}{2}}} \right) . \tag{B.6}$$

To verify that this mapping yields the correct transformation of the boundaries one has to calculate $\mathrm{Im}(z_2)$ by inserting $y_1 = \mathrm{Im}(z_1) = 0$ and 1. This evaluation can be easily done using the general formulae for $\mathrm{Im}(z_2)$ derived in the following (eqs. (B.7) and (B.8)).

With (B.6) we have just stated a suitable transformation without a derivation. For a large number of problems a direct derivation of a sought-after conformal mapping is offered by the Schwarz–Christoffel transformation (see e.g. chapter 4 in [718] or

the toolbox for MATLAB [365]). The conformal mapping for the currently discussed case can also be derived by means of the Schwarz–Christoffel transformation. The mathematical foundations are given in the corresponding literature (e.g. [718, 700]).

With (B.6) the sought-after potential is given by

$$\begin{aligned}\phi_1(z_1) = \phi_2(z_2(z_1)) = \mathrm{Im}(z_2) &= \mathrm{Im}\left(\frac{1}{\pi}\ln\left(\frac{\mathrm{e}^{\pi z_1} - \mathrm{e}^{\pi\frac{a}{2}}}{\mathrm{e}^{\pi z_1} - \mathrm{e}^{-\pi\frac{a}{2}}}\right)\right) \\ &= \frac{1}{\pi}\arg\left(\frac{\mathrm{e}^{\pi z_1} - \mathrm{e}^{\pi\frac{a}{2}}}{\mathrm{e}^{\pi z_1} - \mathrm{e}^{-\pi\frac{a}{2}}}\right) = \frac{1}{\pi}\arctan\left(\frac{\mathrm{Im}\left(\frac{\mathrm{e}^{\pi z_1} - \mathrm{e}^{\pi\frac{a}{2}}}{\mathrm{e}^{\pi z_1} - \mathrm{e}^{-\pi\frac{a}{2}}}\right)}{\mathrm{Re}\left(\frac{\mathrm{e}^{\pi z_1} - \mathrm{e}^{\pi\frac{a}{2}}}{\mathrm{e}^{\pi z_1} - \mathrm{e}^{-\pi\frac{a}{2}}}\right)}\right). \end{aligned} \tag{B.7}$$

The evaluation with $\phi(x, y) = \phi_1(z_1)$ and $z_1 = x + iy$ leads to (5.94), as used in section 5.5:

$$\phi(x, y) = \frac{1}{\pi}\arctan\frac{\sin(\pi y)\,\sinh(\pi\frac{a}{2})}{\cosh(\pi x) - \cos(\pi y)\cosh(\pi\frac{a}{2})}. \tag{B.8}$$

When evaluating the potential numerically care has to be taken that the arctan function is mapped onto the domain $[0, \pi]$ by taking the signs both of the numerator and denominator into account.[1] This results in a potential jump by 1 at the positions $(x = \pm a/2, y = 0)$. The graphical representation of this weighting potential is shown in fig. 5.12.

[1] For this purpose the programming languages Fortran, C, C++ and Python, for example, provide the function atan2(y, x).

Appendix C
Diffusion effects in drift chambers

The spatial resolution of a drift chamber is, amongst others, determined by the dispersion of the drifting electrons in the drift direction due to diffusion. This is discussed in section 7.10.8. Here we provide some additional information to this topic (see also discussion of diffusion in section 4.4 and specifically for electrons in gases on page 117).

Width of the diffusion distribution of drifting electrons. The width of a diffusion distribution after a time t is given by the general formula (4.73):

$$\sigma_{\text{diff}}(t) = \sqrt{2Dt} \qquad \text{or} \qquad \sigma_{\text{diff}}(r) = \sqrt{\frac{2D\,r}{v_D}}\,. \tag{C.1}$$

Since for electrons both the diffusion coefficient D and the drift velocity v_D are in general field-dependent,[1] the formulae only hold for constant field (the right-hand side more generally for $D/v_D = const$). With the field dependence of D and v_D, the diffusion of drifting electrons is also in general field dependent (see fig. 4.13). For normal operating conditions of a drift chamber, namely fixed drift gas and constant ambient conditions, this means that only the field dependence remains. This matters if the field varies over a drift cell.

At a constant ratio D/v_D of the diffusion coefficient and the drift velocity the width of the diffusion distribution of the electrons is proportional to the square root of the drift distance r according to (C.1):

$$\sigma_{\text{diff}}(r) = \sqrt{2Dt} = \sqrt{2D\,r/v_D} \propto \sqrt{r}\,, \qquad (D/v_D = const)\,. \tag{C.2}$$

If the ratio D/v_D is not constant an integral over the drift path has to be taken:

$$d\sigma^2_{\text{diff}} = 2D\,dt = 2D\frac{dr}{v_D} \tag{C.3}$$

$$\Longrightarrow \quad \sigma^2_{\text{diff}} = 2\int_0^r D(v_D)\,\frac{dr'}{v_D} = 2\int_0^r \frac{\epsilon_k}{eE(r')}dr' \qquad (D/v_D \neq const)\,. \tag{C.4}$$

In the last expression the relation (4.112) for the characteristic energy ϵ_k which connects the diffusion coefficient D with the mobility μ, $\epsilon_k = eD/\mu$ as well as $v_D = \mu E$ have been used.

The diffusion is smallest in the thermal limit (4.92), that is, for

$$\epsilon_k = kT\,. \tag{C.5}$$

[1] The general formulae for D and v_D in (4.84) and (4.30), respectively, show the dependence on the electron distribution, which for electrons is mainly determined by the electric field. A simplified example for such dependences is given in section 4.6.4.2.

In this case the variance of the diffusion distribution becomes

$$\sigma^2_{\text{diff}} = 2kT \int_0^r \frac{dr'}{eE(r')}\,, \tag{C.6}$$

yielding for the cylinder-symmetric field (7.5) with a $1/r$ dependence:

$$\sigma^2_{\text{diff}} = 2A \int_0^r r'\,dr' = A\,r^2 \qquad \Rightarrow \qquad \sigma_{\text{diff}}(r) \propto r\,, \tag{C.7}$$

where A is a constant. This proportionality to r holds for any constant characteristic energy, not only in the thermal limit.

Position uncertainty due to diffusion. Averaging over all electrons of a cloud minimises the impact of the diffusion on the position uncertainty. Then the position error is the error of the mean value $\bar{r}$ which for N electrons is

$$\sigma^{\bar{r}}_{\text{diff}}(r|N) = \frac{\sigma_{\text{diff}}(r)}{\sqrt{N}}\,. \tag{C.8}$$

In order to actually exploit this $1/\sqrt{N}$ reduction of the error the following conditions have to be fulfilled:

– The isochrones, connecting the points of equal drift time, should run parallel to the track, so that all electrons arrive on average at the same time.
– The electronics must be able to extract an average of arrival times.

These conditions are not generally fulfilled since the isochrones are usually curved and the electronics registers the signal time at the crossing of a given threshold. A certain approximation of the mean value can be reached with the analysis of FADC spectra (see section 7.10.6).

Deriving the signal time from the crossing of a discriminator threshold (*threshold method*) requires at least k electrons that make up for enough charge to pass the threshold. The time measured at the threshold crossing corresponds to the production point of the kth electron. The variance $\sigma^2_{\text{diff}}(r|N,k)$ of the spatial distribution of this point is smaller than the variance $\sigma^2_{\text{diff}}(r)$ of the whole charge cloud with N electrons:

$$\sigma^2_{\text{diff}}(r|N,k) = \frac{\sigma^2_{\text{diff}}(r)}{2\ln N} \sum_{i=k}^{\infty} \frac{1}{i^2} = \frac{\sigma^2_{\text{diff}}(r)}{2\ln N} \left(\frac{\pi^2}{6} - \sum_{i=1}^{k-1} \frac{1}{i^2} \right)\,. \tag{C.9}$$

This formula is an approximation for large N (see the derivation in section 28.6 of [324]). If the electronics triggers on the first electron the formula yields the resolution contribution

$$\sigma_{\text{diff}}(r|N,1) = \sigma_{\text{diff}}(r) \sqrt{\frac{1}{2\ln N}\frac{\pi^2}{6}}\,. \tag{C.10}$$

Appendix D
Ionisation statistics in drift chambers

This appendix provides some more details for the discussion of the influence of ionisation statistics on the position resolution in drift chambers, as discussed in section 7.10.8. Ionisation statistics refers to the statistical distribution of primary ionisation clusters along a track passing a drift cell. If the particle trajectory in a cell of a drift chamber does not run along an isochrone the fluctuations in cluster generation along the trajectory (ionisation statistics) smear out the position determination (see fig. D.1).

We consider a particle which moves in the drift gas in the y direction and whose trajectory has the smallest distance r from the anode at $y = 0$ (fig. D.1). The y coordinate is normalised such that the trajectory spans the values $-1 < y < +1$ in the cell. The average number of ionisation clusters per unit length is n. Since the cluster distribution is symmetric about $y = 0$ the following computations will initially be executed for positive y.

We assume that the track has created m ionisation clusters on the unit interval $0 < y < 1$. The probability that one cluster falls into the interval dy around y and $k-1$ clusters fall below and $m-k$ above y, is given by the probability mdy that a cluster lies in dy multiplied by the binomial distribution for the splitting of the $m-1$ remaining clusters above and below dy:

$$D_k^m(y)dy = \frac{(m-1)!}{(k-1)!\,(m-k)!} y^{k-1}(1-y)^{m-k} m dy = \frac{m!}{(k-1)!\,(m-k)!} y^{k-1}(1-y)^{m-k} dy\,. \tag{D.1}$$

The probability to observe m clusters if n are expected follows a Poisson distribution:

$$P_m^n = \frac{n^m}{m!}\mathrm{e}^{-n}\,. \tag{D.2}$$

Therefore, the distribution $D_k^m(y)$ has to be folded with this Poisson distribution:

$$\begin{aligned} A_k^n(y) &= \sum_{m=k}^{\infty} P_m^n D_k^m(y) = \mathrm{e}^{-n}\frac{y^{k-1}}{(1-y)^k}\frac{1}{(k-1)!}\sum_{m=k}^{\infty}\frac{n^m}{(m-k)!}(1-y)^m \\ &= \mathrm{e}^{-n}\frac{y^{k-1}n^k}{(k-1)!}\underbrace{\sum_{m'=0}^{\infty}\frac{[n(1-y)]^{m'}}{m'!}}_{=\mathrm{e}^{n(1-y)}}\,. \end{aligned} \tag{D.3}$$

This yields the probability that the kth cluster has the distance y from the origin:

$$A_k^n(y)dy = \frac{y^{k-1}}{(k-1)!}n^k\mathrm{e}^{-ny}dy\,. \tag{D.4}$$

It remains to be noted that y is not restricted to the unit interval because A_k^n is a

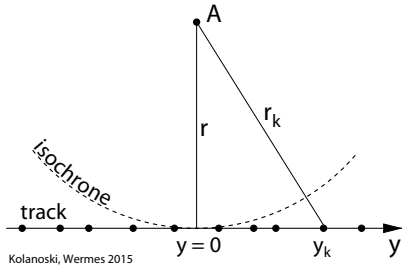


Fig. D.1 Schematic representation of the effect of ionisation statistics for a curved isochrone.

function of ny which is independent of the chosen unit of length (n = number of clusters per length).

The effect of the ionisation statistics depends on how many electrons are necessary to set a time mark (corresponding to the discussion of the effects of diffusion in appendix C). For the kth cluster the mean and the variance of the position become

$$\overline{y_k} = \int_0^1 y\, A_k^n(y) dy \approx \frac{k}{n}\,, \tag{D.5}$$

$$\overline{y_k^2} = \int_0^1 y^2\, A_k^n(y) dy \approx \frac{k(k+1)}{n^2}\,, \tag{D.6}$$

$$\sigma_k^2 = \int_0^1 (y - \overline{y_k})^2\, A_k^n(y) dy = \overline{y_k^2} - \overline{y_k}^2 \approx \frac{k}{n^2}\,. \tag{D.7}$$

The approximations on the right-hand side are valid for $\mathrm{e}^{-n} \approx 0$. With this approximation partial integration yields

$$\int_0^1 y^k \mathrm{e}^{-ny} dy = \frac{k!}{n^{k+1}}\,, \tag{D.8}$$

with which the above integrals can be solved.

Clusters generated at both positive and negative y contribute to the anode signal. Therefore the effective cluster density is $2n$, and then the mean values and variances become

$$\overline{|y_k|} = \frac{k}{2n}, \qquad \sigma_k^2 = \frac{k}{4n^2}\,. \tag{D.9}$$

We now convert the position variance of the kth cluster into the variance of the position of the track. The minimal radial distance r of the track from the anode defines the position of the track in the cell. According to fig. D.1 this radial distance r of the kth cluster is

$$r_k^2 = r^2 + y_k^2 \quad \Rightarrow \quad \frac{dr_k}{dy_k} = \frac{y_k}{\sqrt{r^2 + y_k^2}}\,. \tag{D.10}$$

The mean shift $\overline{|y_k|}$ from the origin will be included in the space–drift-time relation and therefore does not contribute to the error. However, the fluctuation of the kth cluster about $\overline{|y_k|}$ leads to the following contribution to the error in the position determination:

$$\sigma_{ion}^2(r|n,k) = \left(\frac{dr_k}{dy_k}\right)^2_{y_k = \overline{|y_k|}} \times \sigma_k^2 = \frac{k^3}{4n^2(4n^2r^2 + k^2)}\,. \tag{D.11}$$

Appendix E
Position resolution of structured electrodes

A precise measurement of the space coordinates of a signal in a reference plane of a detector is often pursued by employing structured electrodes. Examples are strip or pixel detectors, or more generally all detectors with structured electrodes and multi-channel readout, such as gas-filled detectors with anode wire or structured cathode plane readout or calorimeters with readout of separate cells. In some applications only binary information is given (1 = hit, 0 = no hit), in others the signal is measured proportional to the generated charge as a pulse height or as a pulse integral. A signal, initiated by a particle or by radiation, can appear on a single electrode only or can be distributed over several electrodes. In this appendix we consider the spatial resolution achievable in dependence on the electrode size in relation to the extension of the charge cloud and to the signal-to-noise ratio. Much of the treatment presented here is based on unpublished lectures [413].

In chapter 5 it was shown that the signal measured on the electrodes is caused by electrostatic induction generated by the charge cloud moving in an electric field. For an integrating readout, as is often employed in particle physics experiments, after an integration time which is longer than the arrival time of the drifting charge carriers, a net signal is measured only on those electrodes on which the charge is collected. On the neighbour electrodes a signal also occurs at the beginning of the charge movement, however it disappears again as the charge arrives at the signal electrode. If the induced signal is integrated over time only electrodes collecting (fractions of) the charge cloud have a net 'signal'. During the drift movement to the electrodes the signal charge cloud broadens predominantly by diffusion (see section 4.4). Electrostatic repulsion only plays a negligible role. Diffusion leads to a Gaussian broadening of the charge density distribution at the electrode plane whose width depends on the diffusion constant of the material and the drift distance to the electrode plane (eq. (4.69) on page 102). Several electrodes are 'hit' if the charge cloud is distributed over several electrodes (usually above a threshold), thus generating a 'signal' on them. We consider the case of one-dimensional electrodes (x coordinate) as for example for a strip detector. Extension towards two orthogonal coordinate dimension is straightforward. For equal pitch in both coordinates, the two-dimensional space resolution σ_r with $r = \sqrt{x^2 + y^2}$ is $\sigma_r = \sigma_x \times \sqrt{2}$.

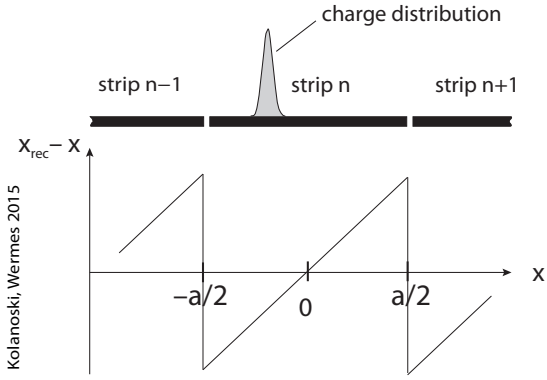


Fig. E.1 Illustration for the spatial resolution with binary readout. For an incoming charge density distribution which is small compared to the electrode width, the reconstructed position x_{rec} is the centre of the electrode. This yields the drawn dependence of the measurement error on the position of the signal.

E.1 Binary detector response

A binary response is given for example if the width of the signal distribution is narrow in comparison to the width of the readout strip and no interpolation between the pixels is done. Only one strip responds upon a particle 'hit'. Figure E.1 illustrates this case. Without loss of generality we place the origin of the coordinate system ($x = 0$) in the centre of the strip. As the reconstructed x coordinate of a hit strip its centre is assumed. The strip pitch from midpoint to midpoint of strips be a. The measurement error Δ_x is the difference between the reconstructed x coordinate (strip centre) x_{rec} and the true x coordinate:

$$\Delta_x = x_{\mathrm{rec}} - x\,. \tag{E.1}$$

The average measurement error vanishes,

$$\langle \Delta_x \rangle = 0\,, \tag{E.2}$$

as long as no systematic contributions influence the measurement. The space resolution is normally defined by the standard deviation of the distribution of the measurement error, that is, the square root of the quadratic deviation from the mean:

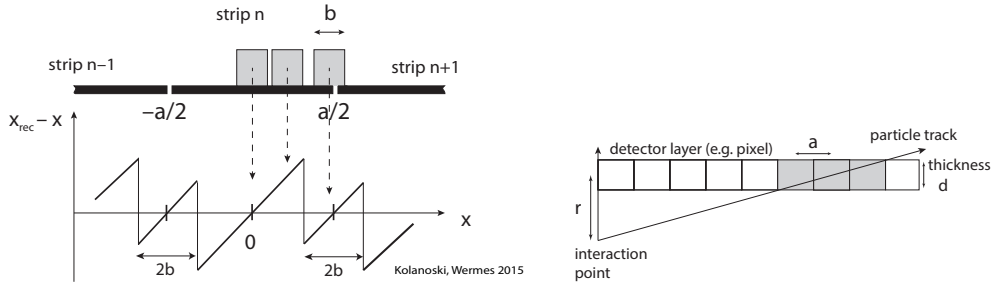
$$\sigma_x = \sqrt{\langle \Delta_x^2 \rangle - \langle \Delta_x \rangle^2} = \sqrt{\langle \Delta_x^2 \rangle}\,, \tag{E.3}$$

where (E.2) has been used.

For the case of a binary response as considered here, the change of the measurement error along the strip width is shown in fig. E.1. It is maximal at the strip boundaries and vanishes in the centre. For a homogeneous particle illumination of the strip at any point x, the probability density of the hit distribution is a uniform distribution $f(x)dx = \frac{1}{a}\,dx$. Since the variance of the quantity x distributed according to a given probability density $f(x)$, is the second moment of the distribution $M_2 = \int_{\min}^{\max} x^2\, f(x)\, dx$, the standard deviation of the measurement error Δ_x for binary response is:

$$\sigma_x^2 = \frac{1}{a}\int_{-a/2}^{a/2} \Delta_x^2\, d(\Delta_x) = \frac{a^2}{12} \;\Rightarrow\; \sigma_x = \frac{a}{\sqrt{12}}\,. \tag{E.4}$$

A variant of this ansatz is the assumption of a box-like charge distribution with a finite width b (fig. E.2(a)). Such a configuration occurs in practice for example for a track passing through a semiconductor detector under a large angle with respect



(a) Box type charge distribution. (b) Track under steep angle.

Fig. E.2 To illustrate the spatial resolution for binary readout of the electrodes for (a) a box-type charge distribution. Three examples of box-like signals (meant to be not appearing at the same time) are indicated. The resolution improves in comparison to fig. E.1 if two strips respond instead of one. (b) An example for box-like charge distributions are tracks entering the detector under very large angles with respect to the vertical.

to perpendicular incidence (fig. E.2(b)). The achievable spatial resolution depends on whether two strips or only one strip respond. In the former case the reconstructed hit coordinate is the average of the two strip centres. At best, if $b = a/2$, this leads to a measurement error which is smaller by a factor of two (fig. E.2(a)).

E.2 Signal partitioning onto multiple electrodes

If the detector response contains information about the amount of detected charge, as given for example for analog readout, a charge *centre of gravity* (c.o.g.) can be determined whose position is a better estimate for the hit coordinate than the centre of the strip. An illustration for the case of a Gaussian distribution of the signal charge on the electrodes is shown in fig. E.3. It is a realistic case if for example a spatially sharp charge deposition broadens (mainly) by diffusion on its way to the electrode plane. The centre of gravity is reconstructed according to

$$x_{rec} = \frac{\sum S_i \, x_i}{\sum S_i} \,, \tag{E.5}$$

where the x_i are the coordinates of the centres of the electrodes (strips) i and S_i the signal heights measured on them. The c.o.g. coordinate is $x_c = x_{rec}$. The signals on the strips are assumed to follow a certain distribution, for example a Gaussian as illustrated in fig. E.3. The signals are then integrals of this distribution over the strip widths.

The achievable resolution depends on the signal distribution function, in particular on its width relative to the strip width. If this ratio is small the resolution approaches the binary resolution of a single strip, $\sigma_x \to a/\sqrt{12}$. In the other limit, if the signal spans over several strips, the resolution becomes very good, approaching zero (without noise). For example, a block-like distribution as in fig. E.2(a) with $b = a$ (block width equals strip width) always yields $\sigma_x = 0$ (again without noise). For a Gaussian distribution this perfect resolution is approached when the width of the charge distribution becomes larger than about half the pitch, $\sigma_x \geq a/2$. This is obvious, since—again without noise—if one perfectly knows the shape of the charge distribution used for the reconstruction (e.g. a Gaussian), the resolution is perfect as soon as more than one

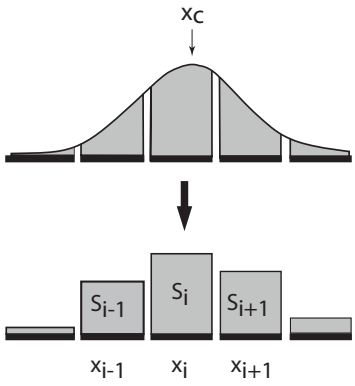


Fig. E.3 Space reconstruction by centre of gravity. x_c is the true centre-of-gravity location of the charge distribution; S_i are the charge signals on the strip electrodes with coordinates x_i.

electrode responds. With this 'intrinsic resolution' approaching zero at some point, the noise on the electrodes will become the dominant contribution to the achievable resolution. That is why we will discuss in the remainder of this appendix the influence of noise on the resolution using structured electrodes.

E.3 Position resolution in the presence of noise

The achievable optimal resolution using c.o.g. depends on the proper matching of the electrode width and the width of the signal charge distribution but also on the signal-to-noise ratio (SNR). Choosing a too small number of electrodes that are hit by an average signal, in the limit $a \gg \sigma$, approaches the binary resolution. On the other hand, a very large number of electrodes firing upon a signal leads to very small signal fractions per electrode relative to the noise. An optimisation is required.

As shown in section 17.10.3, noise on an electrode i is characterised by the variance[1] $\sigma_n^2 = \langle {n_i}^2 \rangle$ (averaged over events k) of the distribution of noise impulses, while the sample mean vanishes $\langle n_i \rangle = 0$. In what follows we assume $\langle {n_i}^2 \rangle$ to be the same for each electrode and that signal S_i and noise n_i be normalised to the total signal, $\sum S_i$, such that for an event the impulse measured on an electrode i is described by $S_i + n_i$. Furthermore we assume for most practical cases that the noise is small compared to the signal, that is, $n_i \ll 1$.

When averaging over noise contributions of different electrodes we distinguish between statistically independent, uncorrelated noise in contrast to (completely) correlated noise. The latter is also called *common mode noise* and can be generated for example by external electromagnetic interference (*pick-up*), which influences many (if not all) electrodes in the same way. For uncorrelated noise we therefore have

$$\langle n_i n_j \rangle = \delta_{ij} \sigma_n^2 \ , \tag{E.6}$$

whereas fully correlated noise on all channels leads to

$$\langle n_i n_j \rangle = \sigma_n^2 \, , \tag{E.7}$$

where σ_n^2 becomes the variance of the common-mode fluctuation.

The midpoints of the electrodes (strips) have coordinates x_i. The origin of the coordinate system is chosen such that $\sum_i x_i = 0$. We further assume for this discussion

[1] The subscript n denotes 'noise' and is not an index.

that the charge distribution is broad compared to the strip width (different to the binary case considered previously) and that ideally (without noise and knowledge of the shape of the charge cloud distribution) the true position of the signal can be perfectly reconstructed as the c.o.g. of the distribution from the sum over the signals on all strips and is in any case negligible compared to the resolution contribution from noise.

In the presence of noise a calculation of the c.o.g. extending (E.5) leads to

$$x_{rec} = \frac{\sum (S_i + n_i)\, x_i}{\sum (S_i + n_i)} = \frac{x + \sum n_i x_i}{1 + \sum n_i} = \left(x + \sum n_i x_i\right)\left(1 - \sum n_i + \mathcal{O}(n_i^2)\right), \tag{E.8}$$

where $x = \sum S_i x_i = x_c$ is the c.o.g. position without noise, which—as described above on page 833—is equal to the true position if the Gaussian charge cloud covers more than one strip. For the right-hand side of (E.8) the numerator has been expanded for small $\sum n_i$. Hence the measurement error is computed as:

$$\begin{aligned} \Delta_x(x) = x_{rec} - x &= \sum n_i x_i - x \sum n_i - \left(\sum n_i x_i\right)\left(\sum n_i\right) + \ldots \\ &= \sum n_i (x_i - x) + \mathcal{O}(n_i^2)\,, \end{aligned} \tag{E.9}$$

and the resolution is obtained as in (E.3) from the mean quadratic error (variance) of the measurement averaging over all $x_{rec} - x$:

$$\begin{aligned} \sigma_x^2 = \langle \Delta_x^2 \rangle - \langle \Delta_x \rangle^2 = \langle \Delta_x^2 \rangle &= \left\langle \sum_{i,j} n_i n_j (x_i - x)(x_j - x) \right\rangle + \mathcal{O}(n_i^3) \\ &= \sum_{ij} \langle n_i n_j \rangle \, \langle (x_i - x)(x_j - x) \rangle + \mathcal{O}(n_i^3)\,, \end{aligned} \tag{E.10}$$

where in the last line the noise and spatial averages have been factorised realising that averaging over all positions in x is independent of the noise averaging over $n_{i,j}$.

For uncorrelated noise we obtain with (E.6) and with $\sum x_i = 0$:

$$\begin{aligned} \sigma_x^2 &= \sum_{ij} \delta_{ij} \sigma_n^2 \, \langle (x_i - x)(x_j - x) \rangle + \ldots = \sum_i \sigma_n^2 \left\langle (x_i - x)^2 \right\rangle + \ldots \\ &= \sigma_n^2 \sum_i \langle x_i^2 - \underbrace{2 x_i x}_{=0} + x^2 \rangle + \ldots = \sigma_n^2 \sum_i \left(\sum_{i=1}^{N} \frac{x_i^2}{N} + \sum_{i=1}^{N} \frac{x^2}{N} \right) + \ldots \\ \Rightarrow \sigma_x^2 &= \sigma_n^2 \left[\left(\sum_{i=1}^{N} x_i^2 \right) + N \langle x^2 \rangle \right] + \mathcal{O}(\sigma_n^3)\,. \end{aligned} \tag{E.11}$$

For correlated noise on all electrodes instead one obtains with (E.7):

$$\sigma_x^2 = \sigma_n^2 N^2 \langle x^2 \rangle + \mathcal{O}(n_i^3)\,. \tag{E.12}$$

The example of a readout with just two strip electrodes illustrates this result, corresponding to a block distribution as in fig. E.2(a) with $b = a$, for which a perfect intrinsic resolution was obtained. The signal splits linearly between the strips, depending on the entrance point x:

$$S_1(x) = \frac{x_2 - x}{a}\,, \qquad S_2(x) = \frac{x - x_1}{a}\,.$$

Both signals satisfy $\sum S_i = 1$ and $x_1 S_1 + x_2 S_2 = x$ with $x_1 + x_2 = 0$ and $x_2 - x_1 = a$.

We compute the space resolution using

$$\sum x_i^2 = x_1^2 + x_2^2 = \frac{a^2}{2} \qquad \text{and} \qquad N\langle x^2\rangle = 2\frac{1}{a}\int_{-a/2}^{a/2} x^2 dx = 2\frac{a^2}{12}\,, \quad \text{(E.13)}$$

obtaining for the uncorrelated case:

$$\sigma_x^2 = \sigma_n^2\left(\frac{a^2}{2} + 2\frac{a^2}{12}\right) = \frac{2}{3}\sigma_n^2 a^2 \qquad \Rightarrow \qquad \sigma_x = 0.082\,\mathrm{a} \;\text{ for }\; \sigma_\mathrm{n} = 0.1$$

and for the fully correlated case

$$\sigma_x^2 = \frac{1}{3}\sigma_n^2 a^2 \qquad \Rightarrow \qquad \sigma_x \approx 6\%\, a \;\text{ for }\; \sigma_\mathrm{n} = 0.1\,.$$

For the uncorrelated case and a typical SNR of 10 ($\sigma_n = 0.1$) we obtain a space resolution of $\sigma_x \approx 8\%\, a$, which is significantly better than the binary resolution of $\sigma_x = a/\sqrt{12} = 29\%\, a$, and is also better than a distribution width which is half the strip width ($b = a/2$) $\sigma_x = \frac{1}{2}a/\sqrt{12} = 14.5\%\, a$ (fig. E.2). Only for SNR < 5.65 does analog charge division no longer improve the resolution compared to a binary measurement.

E.4 A data-driven response treatment

Often for strip or pixel detectors only two electrodes respond for the overwhelming majority of particles passages. For example, this is a typical situation for a microstrip detector with a thickness of 200–300 µm and an electrode pitch between 20 and 50 µm. A perfectly position-proportional splitting of the charge does usually not occur, since for example the response of the electronics is sensitive to the capacitances appearing and the individual gains of the readout channels. We consider two neighbouring electrodes at pitch a with $-a/2 \le x \le a/2$. The electrodes then generate the following 'signals' :

$$S_L(x) = Q\,\eta(x) \quad \text{and} \quad S_R(x) = Q - S_L(x) = Q\,(1 - \eta(x))\,, \qquad \text{(E.14)}$$

where by construction $S_L(x) + S_R(x) = 1$ holds. $\eta(x)$ is the response function of the electrodes and is not dependent on the charge Q. The construction has the following features:

- $0 \le \eta \le 1$.
- $\eta(x)$ is strictly monotonic, that is, $\frac{d\eta}{dx}(x) \neq 0$. This assumption later ensures creating the inverse function η^{-1}.
- The parts of η and $1 - \eta$ are mirror images of each other about $x/a = 0$ and $\eta(x)$ is rotationally symmetric with respect to the midpoint $(x/a, \eta) = (0, 0.5)$.

Examples for response functions $\eta(x)$, which satisfy these conditions, are shown in fig. E.4. Ideal is a perfectly linear response function (fig. E.4(a)).

The big advantage of this ansatz is that $\eta(x)$ can be determined from the measured detector signals themselves [171, 951] using (E.14):

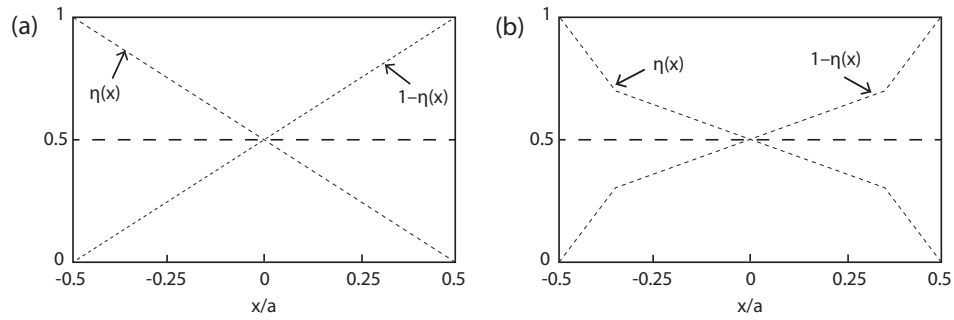


Fig. E.4 Examples of response functions $\eta(x)$ fulfilling the criteria mentioned in the text.

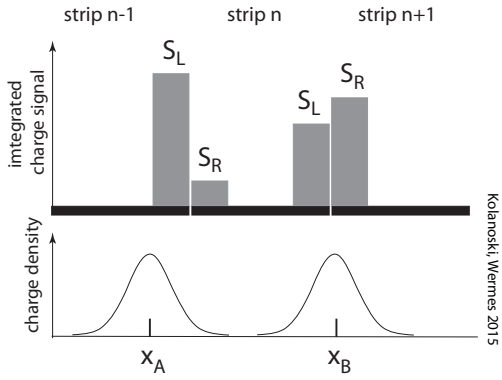


Fig. E.5 Illustration of the η measurement using the example of two signal charge distributions which reach the electrode plane at points x_A and x_B and are detected as proportional to their charge portions on two adjacent strips: S_L and S_R.

$$\eta = \frac{S_L}{S_L + S_R} \,. \tag{E.15}$$

Figure E.5 illustrates the η measurement by means of two signals arriving at the electrode plane. Under the assumption that the electrode plane is illuminated uniformly in x by particles, from the distribution of measured η values one can extract the inverse function η^{-1} using the η distribution $dN/d\eta$ of N measurements. For uniform illumination the ratio of the hit coordinate x to the total electrode width a is equal to the ratio of all $dN/d\eta$ entries in an η histogram up to this point $\eta(x)$, to the total number of histogram entries (see also eqs. (7.64)–(7.66) in section 7.10.7). The reconstructed signal coordinate can thus be determined as

$$x_{rec} = \eta^{-1}\left(\frac{S_L}{S_L + S_R}\right) = \frac{a}{N}\int_0^{\eta} \frac{dN}{d\eta'} d\eta' \,, \tag{E.16}$$

where $\eta^{-1}(S_L/(S_L + S_R))$ denotes the function value of η^{-1} at this η point.

Figure E.6(a) shows a measured η distribution for a silicon microstrip detector with 20 µm strip pitch [171]. Figure E.6(b) shows the relation (found using (E.16)) between reconstructed position x_{rec} and the η value measured for an individual event. The asymmetry in the distribution of fig. E.6(a) results from unequal gains of the two channels. With the η method such and other detector effects are automatically taken into account.

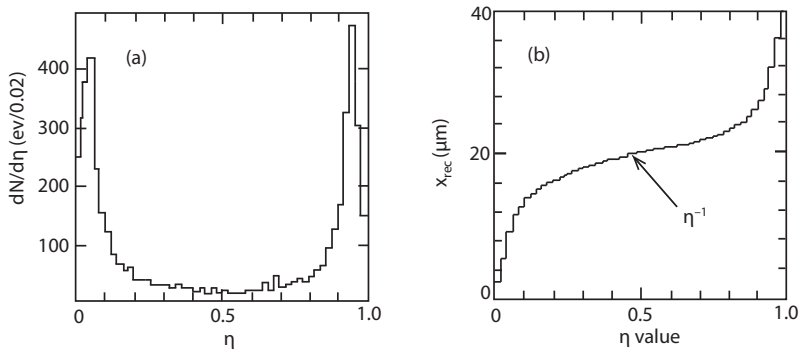


Fig. E.6 (a) Measured η distribution $dN/d\eta$ for a microstrip detector with 20 μm strip pitch and 10 μm strip width (adapted from [171] with kind permission of Elsevier). In (b) the relation between η and the reconstructed space point of the signal x_{rec} is shown, obtained from (E.16).

If noise (n_L and n_R, respectively) is taken into account in the reconstruction, (E.16) changes to (remembering that $S_R + S_L = 1$):

$$\begin{aligned} x_{rec} &= \eta^{-1}\left(\frac{S_L + n_L}{S_L + S_R + n_L + n_R}\right) = \eta^{-1}\left(\frac{\eta(x) + n_L}{1 + n_L + n_R}\right) \\ &\approx \eta^{-1}\Big(\left(\eta(x) + n_L\right)\left(1 - n_L - n_R\right)\Big) \\ &= \eta^{-1}\Big(\eta(x) - \eta(x)n_L - \eta(x)\,n_R + n_L - n_L^2 - n_L n_R\Big) \\ &= \eta^{-1}\Big(\eta(x) + n_L\left(1 - \eta(x)\right) - n_R\,\eta(x) + \mathcal{O}(n_{L,R}^2)\Big) \\ &\approx x + \left.\frac{d\eta^{-1}(s)}{ds}\right|_{\eta(x)}\Big(n_L(1 - \eta(x)) - n_R\,\eta(x)\Big), \end{aligned} \tag{E.17}$$

with $s = \eta(x) + n_L(1 - \eta(x)) - n_R\eta(x)$. In the last step x_{rec} has been Taylor expanded at $\eta(x)$. Finally we exploit the fact that $d\eta^{-1}(s)/ds = 1/\eta'(s)$, with $\eta' = d\eta/ds$, generally holds. Hence we can write the error of the space reconstruction as

$$\Delta_x = x_{rec} - x = \frac{1}{\eta'}\Big(n_L\left(1 - \eta(x)\right) - n_R\eta(x)\Big). \tag{E.18}$$

Because of (E.2) and $\langle n_i \rangle = 0$ the space resolution then results as

$$\begin{aligned} \sigma_x^2 = \langle \Delta_x^2 \rangle - \underbrace{\langle \Delta_x \rangle^2}_{=0} &= \left\langle \frac{n_L^2(1-\eta)^2 - n_R^2\eta^2 - 2n_L n_R\,(1-\eta)\eta}{\eta'^{\,2}} \right\rangle \\ &= \sigma_n^2 \left\langle \frac{1 - 2\eta + 2\eta^2}{\eta'^{\,2}} \right\rangle + 2\langle n_L n_R \rangle \left\langle \frac{\eta^2 - \eta}{\eta'^{\,2}} \right\rangle. \end{aligned} \tag{E.19}$$

Again we assume as in (E.10) that averaging events with respect to their space points and averaging events over noise are independent of each other.

If only statistical, uncorrelated noise ($\langle n_L n_R \rangle = 0$) is present, we obtain

$$\frac{\sigma_x^2}{\sigma_n^2} = \left\langle \frac{1 - 2\eta + 2\eta^2}{\eta'^{\,2}} \right\rangle = 2\left\langle \frac{\eta^2}{\eta'^{\,2}} \right\rangle - \underbrace{2\left\langle \frac{\eta - \frac{1}{2}}{\eta'^{\,2}} \right\rangle}_{= 0, \text{ since } \left(\eta - \frac{1}{2}\right) \text{ is antisymmetric}} = \frac{2}{a} \int_{-a/2}^{a/2} \frac{\eta^2}{\eta'^{\,2}}\, dx\,. \tag{E.20}$$

In the last step the average has been computed by the integral within the boundaries given by the range over which the electrodes respond.

For fully correlated noise (*common mode noise*) with $\langle n_L n_R \rangle = \sigma_n^2$ we obtain from (E.19):

$$\frac{\sigma_x^2}{\sigma_n^2} = \left\langle \frac{1 - 2\eta + 2\eta^2 + 2\eta^2 - 2\eta}{\eta'^{\,2}} \right\rangle = \left\langle \frac{1 - 4\eta + 4\eta^2}{\eta'^{\,2}} \right\rangle = \frac{1}{a} \int_{-a/2}^{a/2} \frac{4\eta^2 - 1}{\eta'^{\,2}}\, dx\,. \tag{E.21}$$

If the derivative of the response function η is small, that is, $\eta(x)$ is rather flat, the resolution is bad. In the simplest case the response function is a linearly falling function from $-a/2$ to $+a/2$: $\eta(x) = \frac{1}{2} - \frac{x}{a}$ (fig. E.4(a)). For this case we obtain assuming uncorrelated noise only:

$$\frac{\sigma_x^2}{\sigma_n^2} = \frac{2}{a} \int_{-a/2}^{a/2} \frac{\left(\frac{1}{2} - \frac{x}{a}\right)^2}{\left(\frac{1}{a}\right)^2}\, dx = \frac{2}{3} a^2\,, \tag{E.22}$$

consistent with the result (E.13) for the c.o.g. method considering a configuration with two strips. For fully correlated noise one finds

$$\frac{\sigma_x^2}{\sigma_n^2} = \frac{2}{a} \int_{-a/2}^{a/2} \frac{4\left(\frac{1}{2} - \frac{x}{a}\right)^2 - 1}{\left(\frac{1}{a}\right)^2}\, dx = \frac{1}{3} a^2\,, \tag{E.23}$$

consistent also with the c.o.g. calculation for two strips (section E.3).

Appendix F
Fitting of track models

In chapter 9 we discuss the fitting of track models to measurement points along a particle trajectory. The track models are functions designed to describe the measurements. They depend on parameters which are to be fitted so that the trajectories are optimally determined. For the optimisation the *least squares method* (LS) is frequently used. If the functions depend linearly on the parameters the optimal parameter estimates can be found by solving a linear system of equations, a procedure called *linear regression*.

In this appendix we briefly introduce the LS formalism in order to prepare the formulae necessary for the discussion of the momentum and direction resolutions in section 9.4. The basic formulae for the resolutions are given in section F.3 of this appendix following the pivotal paper of Gluckstern on tracking resolutions [465]. We abstain from details of the derivations and refer instead to the literature on statistical methods of data analysis in particle physics [566, 246, 224, 155]. A good overview can also be found in the *Review of Particle Physics* [321].

F.1 Least squares method

We consider a sample of measurements and the parametrised description of these measurements:

y_i: measured values at the points x_i, where the x_i are independent variables and as such are known without errors ($i = 1, \ldots, N$);

$V_{y,ij}$: covariance matrix of the measurements y_i; if the measurements are uncorrelated then V_y is diagonal with $V_{y,ii} = \sigma_i^2$ (variance of y_i);

η_i: $\eta_i = f(x_i|\theta)$ is the expectation value of y_i if the dependence on x_i is described by $f(x|\theta)$;

θ_j: parameters ($j = 1, \ldots, m$) of the function f to be optimised such that $f(x_i|\theta) = \eta_i$ describes the measurements y_i as well as possible.

The procedure for applying the *LS principle* is as follows: Determine the estimators $\hat{\theta}$ of the parameters $\theta = (\theta_1, \ldots, \theta_m)$ by minimising the sum of the squares of the relative deviations with respect to their errors. In the general case of a non-diagonal covariance matrix V_y, that is, if the measurements are at least partly correlated, the LS function reads

$$S = \sum_{i=1}^{N}\sum_{j=1}^{N}(y_i - \eta_i)\; V^{-1}_{y,ij}\; (y_j - \eta_j) \quad \left(= \sum_{i=1}^{N} \frac{(y_i - \eta_i)^2}{\sigma_i^2}\,, \text{ if } V_y \text{ is diagonal} \right). \quad \text{(F.1)}$$

F.2 Linear regression

In the following we restrict ourselves to the important case that the fit function $f(x|\theta)$ is a linear function of the parameters $\theta = (\theta_1, \ldots, \theta_m)$:

$$f(x|\theta) = \theta_1\, f_1(x) + \ldots + \theta_m\, f_m(x) = \sum_{j=1}^{m} \theta_j\, f_j(x)\,. \quad \text{(F.2)}$$

The f_j can be arbitrary, thus also nonlinear, functions of x. Then the expectation values for the N measurements y_i are

$$\eta_i = \theta_1\, f_1(x_i) + \ldots + \theta_m\, f_m(x_i) = \sum_{j=1}^{m} \theta_j\, f_j(x_i) = \sum_{j=1}^{m} H_{ij}\, \theta_j\,. \quad \text{(F.3)}$$

The equation defines the $(n \times m)$ matrix H by $H_{ij} = f_j(x_i)$, thereby yielding a compact matrix formulation of the LS function ($\vec{y}$, θ are column vectors containing the measurements and parameters):

$$S = (\vec{y} - H\,\theta)^T\, V_y^{-1}\, (\vec{y} - H\,\theta)\,. \quad \text{(F.4)}$$

From the minimisation condition for S follows a linear system of equations (for details see the references given above), the solution of which yields estimators $\hat{\theta}$ for the parameters:

$$\hat{\theta} = \underbrace{\left(H^T\, V_y^{-1}\, H\right)^{-1}\; H^T\, V_y^{-1}}_{=:\,A}\; \vec{y} = A\vec{y}\,. \quad \text{(F.5)}$$

Thus the parameters $\hat{\theta}$ are given by a linear transformation A of the measurements. By error propagation the covariance matrix of the parameters can consequently be calculated as a linear transformation of the covariance matrix of the measurements:

$$V_\theta = A\, V_y\, A^T = \left(H^T\, V_y^{-1}\, H\right)^{-1}\,. \quad \text{(F.6)}$$

Based on this covariance matrix, we determine in the following the errors of the track parameters for special configurations for which the resolutions are quoted in section 9.4.

Using the fitting function f the estimated y values can be calculated for arbitrary x values:

$$\hat{y} = \sum_{j=1}^{m} \hat{\theta}_j\, f_j(x)\,. \quad \text{(F.7)}$$

The error of $\hat{y}$ is obtained by error propagation:

$$\sigma_y^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial y}{\partial \theta_i}\frac{\partial y}{\partial \theta_j}\, V_{\theta,ij} = \sum_{i=1}^{m}\sum_{j=1}^{m} f_i(x)\, f_j(x)\, V_{\theta,ij}\,. \quad \text{(F.8)}$$

These formulae can be easily programmed using program packages which support matrix operations, like Mathematica, MATLAB or Python.

F.3 Applications: errors of track parameters

In order to come to quantitative estimates of the track parameter resolutions in section 9.4 we make the following assumptions:

- Each track has N measurements, each at a fixed coordinate x_i.
- The measurement errors are uncorrelated and all the same, $\sigma_i = \sigma$.
- The x_i are distributed with equal spacing over a length L centred at $x_c = 0$:

$$x_N - x_1 = L\,, \qquad x_i = x_1 + (i-1)\,\frac{L}{N-1}\,, \qquad x_c = \frac{x_1 + x_N}{2} = 0\,. \tag{F.9}$$

In the following we restrict ourselves to the formulae for the parameter errors under these conditions which can be determined with the help of the covariance matrix of the parameters (F.6). Most of the results have been originally derived by Gluckstern [465].

Fitting a straight line to the measurements. If the measurements are expected to lie on a straight line ($f_1(x) = 1,\ f_2(x) = x$) the fitting function is

$$y = f(x|\theta) = a + b\,x\,, \qquad \text{with} \quad \theta_1 = a,\ \theta_2 = b\,. \tag{F.10}$$

The evaluation of (F.6) yields for the errors of the parameters:

$$\sigma_a^2 = \frac{\sigma^2}{N}, \qquad \sigma_b^2 = \frac{\sigma^2}{L^2}\frac{12(N-1)}{N(N+1)}, \qquad \sigma_{ab} = 0\,. \tag{F.11}$$

The fact that the error matrix is diagonal ($\sigma_{ab} = 0$) is a consequence of the choice of the centre of gravity of the measurement points, x_c, to lie in the origin according to (F.9).

An estimator $\hat{y}_0$ at an arbitrary coordinate x_0 can be calculated using (F.7). According to (F.8) its error is

$$\sigma_y^2|_{x_0} = \sigma_a^2 + x_0^2\,\sigma_b^2 = \frac{\sigma^2}{N}\left(1 + \frac{12(N-1)}{(N+1)}\frac{x_0^2}{L^2}\right). \tag{F.12}$$

Thus the resolution depends on the ratio of the extrapolation distance x_0 from the centre of gravity of the measurements and the length over which the measurement points are distributed. Application examples are discussed in section 9.4.6.

Parabolic approximation of a circular arc. We now want to discuss the parabolic approximation of a circular arc (9.21) as a fitting function for a track in a magnetic field:

$$y = a + bx + \frac{1}{2}cx^2\,. \tag{F.13}$$

This function is linear in the parameters a, b, c so that linear regression can be applied. As in the case of the straight line fit the parameter resolutions can be obtained by evaluating the corresponding covariance matrix of the parameters (F.6):

$$\sigma_a^2 = \sigma^2\frac{3N^2 - 7}{4(N-2)N(N+2)}\,, \tag{F.14}$$

$$\sigma_b^2 = \frac{\sigma^2}{L^2}\frac{12(N-1)}{N(N+1)}\,, \tag{F.15}$$

$$\sigma_c^2 = \frac{\sigma^2}{L^4}\frac{720(N-1)^3}{(N-2)N(N+1)(N+2)}\,, \tag{F.16}$$

$$\sigma_{ab} = \sigma_{bc} = 0\,, \tag{F.17}$$

$$\sigma_{ac} = \frac{\sigma^2}{L^2}\frac{30N}{(N-2)(N+2)}\,. \tag{F.18}$$

As before the error of an estimator $\hat{y}_0$ at an arbitrary coordinate x_0 is determined with the help of (F.8):

$$\begin{aligned}\sigma_y^2\big|_{x_0} &= \sigma_a^2 + x_0^2\,\sigma_b^2 + \frac{1}{4}x_0^4\,\sigma_c^2 + x_0^2\,\sigma_{ac} \\ &= \frac{\sigma^2}{N}\left(\frac{3N^2-7}{4(N-2)(N+2)} + \frac{x_0^2}{L^2}\frac{12(N-1)}{(N+1)}\right. \\ &\quad \left. + \frac{x_0^4}{L^4}\frac{180(N-1)^3}{(N-2)(N+1)(N+2)} + \frac{x_0^2}{L^2}\frac{30N^2}{(N-2)(N+2)}\right)\,.\end{aligned} \tag{F.19}$$

This equation is used for the discussion of vertex resolutions in section 9.4.6.

Appendix G
LPM effect

In the context of the development of electromagnetic showers the Landau–Pomeranchuk–Migdal (LPM) effect was introduced in section 15.2.2.3 (e.g. [610]). The effect suppresses the two most important processes of the shower development, bremsstrahlung and pair production, with increasing energy due to quantum mechanical interference. This leads to an extension of the longitudinal shower dimension and an increase of shower fluctuations.

In the quantum mechanical description of interactions an important feature is the *formation* or *coherence length* over which the interacting particle waves are coherent. This is due to the fact that quantum mechanically a particle interaction is described by a space-time integral over the product of the wavefunctions of the incoming and outgoing particles (the transition from initial to final state being determined by an interaction operator). Therefore, increasing decoherence between the wavefunctions leads to less contribution to the integral and thus to a smaller cross section. The relevant integration volume is given by the coherence length and coherence time (here we concentrate on the length). If the phase coherence is disturbed within this relevant volume the interaction probability is affected, mostly destructively. In the bremsstrahlung process, for example, the formation length is the range over which the relative phases of the incoming and outgoing electron and photon waves remain about the same. In the shower development, disturbances of the phases between the incoming and outgoing electrons and photons can occur through multiple scattering of the electrons if the formation length (for either bremsstrahlung or for pair production) extends over several atoms. Relevant is the phase evolution in the longitudinal direction, the dominant propagation direction of the involved electrons, positrons and photons.

The formation length of the bremsstrahlung process without disturbance, l_{f0}, corresponds in Fourier space to the longitudinal momentum transfer $q_\parallel$ onto the scattering centre (mostly a nucleus, see fig. 15.2):

$$l_{f0} = \frac{\hbar}{q_\parallel} \overset{\hbar=1}{=} \frac{1}{q_\parallel}\,. \tag{G.1}$$

Neglecting the emission angle of the photon, $q_\parallel$ results from the (longitudinal) momenta of the incoming and outgoing electron p_e and p'_e, respectively, as well as that of the photon p_γ:

$$q_\parallel = p_e - p'_e - p_\gamma = \sqrt{E^2 - m_e^2} - \sqrt{(E-k)^2 - m_e^2} - k\,. \tag{G.2}$$

Here E and $E-k$ are the energies of the incoming and outgoing electron, respectively, and $E_\gamma = p_\gamma = k$ is the photon energy. With that, $q_\parallel$ corresponds exactly to the momentum transfer calculated in (3.70). For $m_e/E \ll 1$ and $m_e/(E-k) \ll 1$ the resulting formation length is

$$l_{f0} = \frac{1}{q_{\parallel}} \approx \frac{2E(E-k)}{m_e^2\, k}\,. \tag{G.3}$$

If the primary electron energy E is large compared to the radiated energy k, l_{f0} becomes large compared to atomic distances. In amorphous media,[1] the suppression of bremsstrahlung sets in when l_{f0} becomes larger than the characteristic length between two phase-disturbing interactions.

In many cases multiple scattering is the dominant effect causing a phase disturbance. As a measure for the size of the phase shift between the in- and outgoing particles one uses as a criterion for a noticeable suppression the point when the multiple scattering angle becomes larger than the deflection by the radiation process:

$$\theta_{ms}^{space} > \theta_{rad}\,. \tag{G.4}$$

Following the literature (see e.g. [610]) we take for θ_{ms}^{space} the average (spatial) scattering angle at half the formation length and for the deflection by the emission process we take the characteristic radiation angle $1/\gamma$ for bremsstrahlung. Then we obtain from (G.4):

$$\theta_{ms}^{space} \approx \theta_{ms}^{space}(l_{f0}/2) = \frac{E_s}{E}\sqrt{\frac{l_{f0}}{2\,X_0}} > \frac{1}{\gamma} = \frac{m_e}{E}\,. \tag{G.5}$$

As average spatial scattering angle we take $\theta_{ms}^{space} = \sqrt{2}\theta_{ms}$ with the planar scattering angle θ_{ms} as defined in (3.100) with E_s as in (3.101). Inserting l_{f0} as in (G.3), we can rewrite (G.5) as a condition for the radiated photon energy at a fixed energy E below which the LPM suppression sets in [610]:

$$k < \frac{E^2}{E + \dfrac{m_e^4}{E_s^2}X_0} = \frac{E^2}{E + E_{\mathrm{LPM}}} = \frac{E}{1 + \dfrac{E_{\mathrm{LPM}}}{E}}\,. \tag{G.6}$$

With this formula E_{LPM} is introduced as a characteristic energy which defines the scale on which the suppressed region of photon energies increases as a function of the primary energy E (see fig. 15.6(a)). With increasing energy and $E \gg E_{\mathrm{LPM}}$ the radiated photons accumulate closer and closer to the maximal possible energy. The energy E_{LPM} depends on the medium only through the radiation length.[2] Starting from (G.6) we use for numerical calculations

$$E_{\mathrm{LPM}} = \frac{m_e^4}{E_s^2}X_0 = m_e^2\frac{\alpha}{4\pi}X_0 = 7.7\,\mathrm{TeV}\,\frac{X_0}{\mathrm{cm}}\,. \tag{G.7}$$

In the last term on the right-hand side the SI units are introduced by multiplying with a factor $c^4/(\hbar\, c)$. The energy E_s was defined in (3.101). Since here the physical distances matter, the radiation length is given in absolute length units, here in cm. For example, the characteristic energies of lead ($X_0 = 0.56\,\mathrm{cm}$) and water/ice ($X_0 = 36.1\,\mathrm{cm}$) are:

$$E_{\mathrm{LPM}} = \begin{cases} 4.3\,\mathrm{TeV} & \text{lead} \\ 305\,\mathrm{TeV} & \text{water or ice.} \end{cases} \tag{G.8}$$

[1] In crystals additional coherence effects can play a role.

[2] The definition of the characteristic energy E_{LPM}, which we take here from [610], is somewhat arbitrary (because suppression sets in quite slowly) and differs by factors of 2–8 in the literature.

The suppression also occurs for pair production, the other reaction responsible for the development of electromagnetic showers. Here the suppression is strongest for a relatively symmetric splitting of the energy between electrons and positrons, yielding strongly asymmetric energy distributions at high energies; see for example [610]. Therefore, with increasing energies the LPM effect causes the bremsstrahlung spectrum to be shifted to higher energies and causes the pair production process to transmit more energy to just one particle. At high energy both processes, bremsstrahlung and pair production, cause a stretching of showers, leading to fewer shower particles and stronger fluctuations of the energy deposition over the shower length. We quantitatively demonstrate this behaviour with the example in section 15.2.2.3.

Appendix H
Laplace transform

The Laplace transform is a powerful mathematical tool for the analysis of electrical circuits. Here we only introduce the essential characteristics and the mathematical method, together with some simple, but typical examples which are relevant for the treatment of electronic detector signals. Good introductions into the subject can be found for example in [78] or in [748].

The Laplace transform of a time-dependent function $f(t)$ which vanishes for negative times[1] ($f(t) = 0$ for $t < 0$) is defined by

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty f(t)\, \mathrm{e}^{-st} dt\,, \tag{H.1}$$

where $s = \sigma + i\omega$ is the complex Laplace variable, whose real part σ is a constant generating convergence for time-wise confined pulses, and whose imaginary part is the frequency $\omega = 2\pi f$.

For purely imaginary $s = i\omega$, (H.1) corresponds to the Fourier transform of the signal in the time domain:

$$F(i\omega) = \mathcal{F}[f(t)] = \int_{-\infty}^\infty f(t)\, \mathrm{e}^{-i\omega t} dt\,. \tag{H.2}$$

Both, Fourier and Laplace transform, are used in circuit calculation, each having properties that may be preferable depending on the application. The Fourier transform is generally preferred for boundary problems for which the solution of a differential equation vanishes at infinity, while the Laplace transform is optimal for differential equations with initial value condition (e.g. at a certain point in time t_0), as this is usually the case for electronic signals. Moreover, due to the $\mathrm{e}^{-\sigma t}$ term, the convergence behaviour for Laplace transforms is better than for Fourier transforms.

The signal in the time domain is retrieved by the inverse transformation:

$$\mathcal{L}^{-1}[F(s)] = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} F(s) \mathrm{e}^{st} ds = \begin{cases} f(t) & \text{for} \quad t \geq 0 \\ 0 & \text{for} \quad t < 0 \end{cases} \tag{H.3}$$

Table H.1 compiles some Laplace correspondences between $f(t)$ and $F(s)$ which are important for applications to electronic circuits.

We can demonstrate the usefulness of Laplace transforms employing a resonant circuit (fig. H.1). Since the derivative $f'(t) = df(t)/dt$ is mapped onto $sF(s)$ the Laplace transform of a linear differential equation usually is an easier algebraic equation to

[1] Besides the here defined 'unilateral' transformation there is also a 'bilateral' form for which the integral runs from $-\infty$ to $+\infty$. For $f(t < 0) = 0$ both definitions are equivalent.

Table H.1 Compilation of Laplace correspondences between functions in the time domain and in the frequency domain.

Operation	Time domain	Frequency domain
or function	$f(t) = \mathcal{L}^{-1}[F(s)]$	$F(s) = \mathcal{L}[f(t)]$
linearity	$a_1 f_1(t) + a_2 f_2(t)$	$a_1 F_1(s) + a_2 F_2(s)$
convolution	$\int_0^\infty f(t-t')g(t')dt'$	$F(s)G(s)$
nth derivative	$\frac{d^n}{dt^n} f(t)$	$s^n F(s)$
time integration	$\int_0^t f(t)dt$	$\frac{1}{s}F(s)$
scaling of t	$f(at)$	$\frac{1}{a}F(\frac{s}{a})$
time shift	$f(t-t_0)$	$\mathrm{e}^{-st_0}F(s)$
damping	$\mathrm{e}^{-s_0 t}f(t)$	$F(s+s_0)$
multiplication	$t^n f(t)$	$(-1)^n \frac{d^n}{ds^n}F(s)$
δ function	$\delta(t)$	1
derivative of the δ function	$\frac{d^n}{dt^n}\delta(t)$	s^n
step function	$\Theta(t)$	$\frac{1}{s}$
falling exponential	e^{-at}	$\frac{1}{s+a}$
rising exponential	$1-\mathrm{e}^{-at}$	$\frac{a}{s(s+a)}$
power function	t^n	$\frac{n!}{s^{n+1}}$

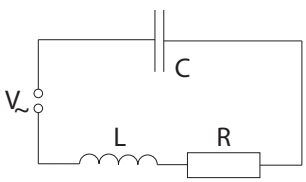


Fig. H.1 Series resonant circuit as a simple application example for the Laplace transform.

deal with. Kirchhoff's voltage law $\sum_i v_i = 0$ with impedances R, C and L leads to the well known second order differential equation

$$L\frac{d^2}{dt^2}i(t) + R\frac{d}{dt}i(t) + \frac{1}{C}i(t) = \frac{dV_\sim}{dt}\,, \tag{H.4}$$

to be solved. The equation of the Laplace-transformed functions is:

$$sLI(s) + RI(s) + \frac{1}{sC}I(s) = V(s)\,. \tag{H.5}$$

The total impedance of the circuit can immediately be computed as

$$Z(s) = R + \frac{1}{sC} + sL\,, \tag{H.6}$$

which, with the substitution $s \to i\omega$, yields the complex impedance of the resonant circuit.

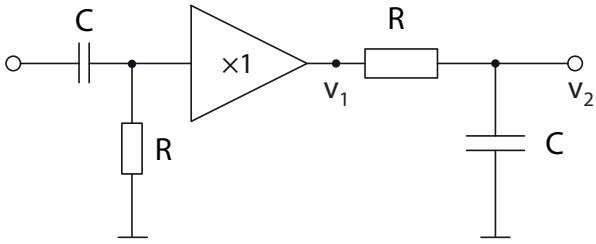


Fig. H.2 High pass–low pass sequence for pulse shaping. The (×1) amplifier serves as a buffer to prevent loading the high pass. It can remain disregarded in the Laplace transformation.

As an example which has a specific application in signal pulse shaping (see section 17.3) we consider a so-called CR-RC filter consisting of a connection of high- and low-pass filters in series (fig. H.2). In Laplace space the transfer function $H(s) = \frac{v_{out}(s)}{v_{in}(s)}$ of a series connection of the two circuit elements,

$$H_1(s) = \frac{sRC}{1+sRC}\,, \tag{H.7}$$

$$H_2(s) = \frac{1}{1+sRC}\,, \tag{H.8}$$

is simply the product of the individual transfer functions $H_1(s) \times H_2(s)$, such that the voltages in Laplace space are:

$$v_1(s) = H_1(s)\, v(s) = \frac{sRC}{1+sRC}\, v(s) = \frac{s\tau}{1+s\tau} v(s)\,, \tag{H.9}$$

$$v_2(s) = H_2(s)\, v_1(s) = \frac{1}{1+sRC}\, v_1(s) = \frac{s\tau}{(1+s\tau)^2} v(s)\,, \tag{H.10}$$

with $\tau = RC$.

Sending now a step function pulse into the input:

$$v(t) = V_0\ \Theta(t) = \begin{cases} 0 & ,\ t \le 0\,, \\ V_0 & ,\ t > 0\,, \end{cases} \qquad v(s) = V_0\, \frac{1}{s}\,, \tag{H.11}$$

we obtain

$$v_2(s) = \frac{V_0 \tau}{(1+s\tau)^2}\,. \tag{H.12}$$

Transforming back into the time domain we obtain, employing the rules given in table H.1,

$$v_2(t) = V_0 \frac{t}{\tau} \mathrm{e}^{-t/\tau}\,, \tag{H.13}$$

which corresponds to eq. (17.23) in section 17.3.

Appendix I
Physical noise sources

Tracing noise phenomena to their physical origins and their mathematical description belongs to the difficult themes of detector physics. Literature references are for example [799, 826, 905, 540, 472].

Since the mean value of noise fluctuations is zero, noise is generally quantified by the variance of the distribution of the fluctuating quantity.

There are two different approaches to describe noise in electronic circuits [826]. A view often used is the assumption that noise can be treated as some sort of signal if some rules are obeyed, for example that noise voltages or currents must be added in quadrature. The result of such a contemplation is for instance a statement like: 'The output signal of the amplifier is a sine function $V(t) = V_0 \sin\omega t$ with additional noise contributions having a mean (quadratic) intensity in the frequency interval df of $d\langle v_n^2\rangle$.' The second approach considers noise as statistical fluctuations, describing the same measurement as follows: 'The expectation value of the output signal is $V(t)$ with a mean quadratic fluctuation of size $\langle dv_n^2\rangle$.' Hence the equation

$$d\langle i_n^2\rangle = 2eI_0\,df \tag{I.1}$$

would mean in the first interpretation that a noise current (here shot noise) dissipates in a resistor R and frequency interval df a power of $R\,d\langle i_n^2\rangle$, whereas the second view would say that a measurement of the current with an instrument of bandwidth df has expectation value and standard deviation of $I_0 \pm \sqrt{d\langle i_n^2\rangle}$.

Both views are possible and are applied. Electronics people prefer the first, physicists usually the second point of view.

I.1 Thermal noise

Already in 1906 Einstein pointed out [371] that Brownian motion of charge carriers would lead to voltage fluctuations between the ends of any resistor. The effect was observed by Johnson in 1926 [577] and the power spectrum was computed by Nyquist [742] shortly afterwards. Thus thermal noise (also called Johnson noise or Nyqvist noise) results from velocity fluctuations of charge carriers.

To compute the thermal noise power spectrum [742] (see also [826]) we consider an open resistor R_1. Its thermal noise shall be given by a (quadratic) noise voltage $\langle v_1^2\rangle$ (fig. I.1). The noise voltage $\langle v_1^2\rangle$ over R_1 yields a noise power in R_2 when both resistors are short-circuited (fig. I.1):

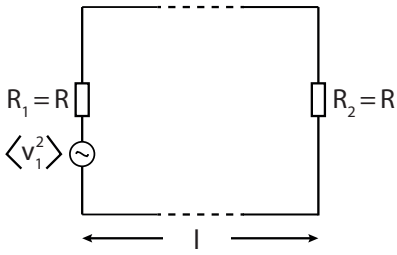


Fig. I.1 Illustration used for thermal noise description: a system of two equal resistors which can be connected (dashed lines).

$$P_{1\to 2} = \frac{v^2}{R_2} = \frac{\langle v_1^2\rangle}{R_2}\left(\frac{R_2}{R_1+R_2}\right)^2 = \frac{\langle v_1^2\rangle}{4R}\,, \tag{I.2}$$

where v is the voltage over R_2 caused by $\langle v_1^2\rangle$.

In thermal equilibrium R_2 transfers the same noise power to R_1

$$P_{1\to 2} = P_{2\to 1}$$

for every frequency portion of the noise fluctuation [826]. The power spectrum hence is a function of f, of R, and of the temperature T.

If $w(f)$ is the frequency spectrum of the voltage fluctuations over R for an open circuit,

$$d\,\langle v^2\rangle = w(f)\,df\,, \tag{I.3}$$

then the mean power transferred by a resistor R in the (closed) circuit of fig. I.1 in a frequency interval df is:

$$dP = \frac{1}{4R}w(f)\,df\,. \tag{I.4}$$

After a characteristic time l/u, where u is the propagation velocity of the fluctuation, thermal equilibrium is reached in the circuit of fig. I.1, where an average power dP according to (I.4) flows from left to right and from right to left with the corresponding average energy

$$dE = \frac{2l}{u}dP = \frac{l}{2uR}w(f)\,df\,. \tag{I.5}$$

In thermal equilibrium at a temperature T this energy is that of a standing wave in the circuit. Following the laws of statistical thermodynamics then every frequency mode can be assigned its energy according to Planck's law of radiation:

$$\epsilon(f) = \frac{hf}{\exp\left(\frac{hf}{kT}\right)-1} \simeq \frac{hf}{1+\frac{hf}{kT}-1} = kT\,, \tag{I.6}$$

for sufficiently small frequencies $hf \ll kT$ (i.e. $\lesssim$ THz).

The oscillation modes of the circuit appear at frequencies $f = \frac{n}{2}\frac{u}{l}$, with $n = 1, 2, 3, \ldots$ For sufficiently large l we have

$$dn = \frac{2l}{u}df\,,$$

and the thermal energy in df is

$$dE = \epsilon\, dn = \frac{2l}{u} kT\, df\,. \tag{I.7}$$

Equating (I.5) and (I.7) yields:

$$w(f) = 4kTR \qquad \text{and} \qquad \frac{dP}{df} = kT\,. \tag{I.8}$$

The thermal spectrum is 'white', meaning that it is independent of the frequency f. Between the ends of every resistor and in every frequency band df hence exists with (I.3) a mean quadratic noise voltage of

$$d\langle v_n^2\rangle = 4kTR\, df\,, \tag{I.9}$$

where the subscript n here just denotes 'noise'. For a resistor of 1 MΩ and a bandwidth of 1 MHz the thermal noise value $\sqrt{\langle v_n^2\rangle}$ is 127 μV at room temperature (293 K).

The thermal noise current in a resistor correspondingly is:

$$d\langle i_n^2\rangle = d\frac{\langle v_n^2\rangle}{R^2} = \frac{4kT}{R} df\,. \tag{I.10}$$

The factor 4 originates from the appearance of two resistors in the derivation (fig. I.1). The result, however, is general, as shown in Nyquist's original paper [742].

I.2 Shot noise

Shot noise emerges as a consequence of the discrete nature of electric charge and is given by statistical fluctuations in the number of charge carriers. The phenomenon is particularly distinctive when charge carriers must cross a barrier, such as in an electron tube or a semiconductor diode, where the charge quantisation plays a role for the resulting current. Shot noise was examined and explained first by Schottky in 1918 [867].

Consider injection of excess electrons in a system with a work function barrier, for example an electron tube (fig. I.2) or a pn boundary. Every emitted electron can be regarded as a current impulse:

$$i_k(t) = e\delta(t - t_k)\,. \tag{I.11}$$

The total current I_0 integrated over a longer time period T is the sum of individual current impulses:

$$I_0 = \frac{1}{T}\int_0^T \sum_{k=1}^{N} i_k(t)dt = \frac{eN}{T} = e\dot{N}\,. \tag{I.12}$$

The δ-function can be written as a Fourier series in time (see e.g. [117, 976]):

$$i_k(t) = e\delta(t - t_k) = \frac{2e}{T}\left(\sum_{m=1}^{\infty} \cos m\omega(t - t_k) + \frac{1}{2}\right)\,. \tag{I.13}$$

We are interested in the average fluctuation of the current I_0, averaged over all random t_k in the interval T, expressed by the variance $\langle I^2\rangle$. Note that in the variance the DC component of the current, accounted for by the 1/2 in (I.13), drops out. Hence,

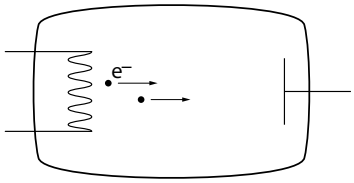


Fig. I.2 Electron tube as example for the generation of statistical fluctuations upon the emission of electrons from the cathode when surpassing the work function of the metal.

capitalising on the orthogonality of the Fourier components, every frequency mode m contributes to the variance with

$$d\langle I^2\rangle\big|_{m,k} = \langle (2e/T)^2\rangle \, \langle \cos^2 m\omega(t-t_k)\rangle \,, \tag{I.14}$$

where the explicit t dependence of the cosine does not appear in the average. Every component m contains the contributions of all electrons. Since the arrival times t_k of the electrons appear in an uncorrelated way somewhere in the interval $[0,T]$, their phases are randomly distributed and the $\cos^2$ contributions average to $1/2$. Hence every frequency component m for every electron k contributes

$$d\langle I^2\rangle\big|_{m,k} = \frac{2e^2}{T^2} \tag{I.15}$$

to the total variance. If $\dot{N}$ is the frequency of the current impulses then there are $m = \dot{N}\,T$ impulses within $[0,T]$ and we obtain:

$$d\langle I^2\rangle\big|_{m} = \frac{2e^2\,\dot{N}}{T} \,. \tag{I.16}$$

This is the average fluctuation at a given frequency m. A frequency band df contains $T\,df$ frequencies. Therefore the mean squared current fluctuation becomes:

$$d\langle I^2\rangle = \frac{2e^2\,\dot{N}}{T} T df = 2e^2 \dot{N} df = 2eI_0 df \tag{I.17}$$

with $\dot{N}e = I_0$ as introduced in (I.12).

It should be noted that shot noise requires the presence of a current and hence needs a power source, whereas thermal noise is present in devices even without current flow and without external power.

I.3 1/f noise

So-called 1/f noise phenomena are found in many physical and non-physical phenomena (see e.g. [795]). It is therefore highly improbable that the same mechanism can hold for its origin. In electronic systems 1/f spectral contributions are usually attributed to carrier trapping and releasing processes, for example and in particular at Si–SiO_2 interfaces as in MOS transistor channels. For a quantitative description [708] we assume that trapped electrons are released after a characteristic time τ following a relaxation law behaviour:

$$N(t) = N_0\,\mathrm{e}^{-t/\tau} \quad \text{for } t \geq 0\,, \qquad N(t) = 0 \quad \text{else}\,. \tag{I.18}$$

Fourier transform of a single exponential relaxation process into the frequency domain yields:

$$F(\omega) = \int_{-\infty}^{\infty} N(t) \mathrm{e}^{-i\omega t}\, dt = N_0 \int_0^{\infty} \mathrm{e}^{-(1/\tau + i\omega)t}\, dt = N_0 \frac{1}{1/\tau + i\omega} \quad . \tag{I.19}$$

For a sequence of such pulse processes occurring at different trapping times t_k we then have:

$$N(t, t_k) = N_0\, \mathrm{e}^{-\frac{t-t_k}{\tau}} \quad \text{for } t \geq t_k \ , \qquad N(t, t_k) = 0 \ \ \text{else}\,, \tag{I.20}$$

and the Fourier transform for many trapping processes becomes

$$F(\omega) = N_0 \sum_k \mathrm{e}^{-i\omega t_k} \int_0^{\infty} \mathrm{e}^{-(1/\tau + i\omega)t}\, dt = \frac{N_0}{1/\tau + i\omega} \sum_k \mathrm{e}^{i\omega t_k}\,. \tag{I.21}$$

The power spectrum then is given as

$$\begin{aligned} P(\omega) = \lim_{T\to\infty} \frac{1}{T} \left\langle |F(\omega)|^2 \right\rangle &= \frac{N_0^2}{(1/\tau)^2 + \omega^2} \lim_{T\to\infty} \frac{1}{T} \left\langle \left| \sum_k \mathrm{e}^{i\omega t_k} \right|^2 \right\rangle \\ &= \frac{N_0^2}{(1/\tau)^2 + \omega^2}\, n\,, \end{aligned} \tag{I.22}$$

where n is the average rate of the trapping/releasing processes and T is the interval over which the process is observed.

If we assume in addition that relaxation time constants can differ, and integrate over all τ_i between τ_1 and τ_2, assuming that the decay constants $\lambda_i = 1/\tau_i$ are evenly distributed, we obtain

$$\begin{aligned} P(\omega) &= \frac{1}{\frac{1}{\tau_1} - \frac{1}{\tau_2}} \int_{\frac{1}{\tau_2}}^{\frac{1}{\tau_1}} \frac{N_0^2\, n}{\left(\frac{1}{\tau}\right)^2 + \omega^2}\, d\left(\frac{1}{\tau}\right) = \frac{N_0^2\, n}{\omega\left(\frac{1}{\tau_1} - \frac{1}{\tau_2}\right)} \left[\arctan\frac{1}{\omega\tau_1} - \arctan\frac{1}{\omega\tau_2}\right] \\ &\approx \begin{cases} N_0^2\, n & \text{if } \ 0 < \omega \ll \frac{1}{\tau_1}, \frac{1}{\tau_2} \quad \rightarrow \quad \text{const.}\,, \\ \dfrac{N_0^2\, n\, \pi}{2\omega\left(\frac{1}{\tau_1} - \frac{1}{\tau_2}\right)} & \text{if } \ \frac{1}{\tau_2} \ll \omega \ll \frac{1}{\tau_1} \quad \rightarrow \quad \dfrac{1}{f}\,, \\ \dfrac{N_0^2\, n}{\omega^2} & \text{if } \ \frac{1}{\tau_1}, \frac{1}{\tau_2} \ll \omega \quad \rightarrow \quad \dfrac{1}{f^2}\,. \end{cases} \end{aligned} \tag{I.23}$$

This behaviour is shown in fig. 17.60(b) on page 788.

Bibliography

[1] Aab, A. et al. (Pierre Auger): Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. In: *JCAP* 2017 (2017), p. 038. doi: 10.1088/1475-7516/2017/04/038. – [Erratum in [2]]

[2] Aab, A. et al. (Pierre Auger): Erratum: Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. In: *JCAP* 2018 (2018), p. E02. doi: 10.1088/1475-7516/2018/03/e02. – [Erratum to [1]]

[3] Aad, G. et al. (ATLAS Collaboration): ATLAS pixel detector electronics and sensors. In: *JINST* 3 (2008), p. P07007. doi: 10.1088/1748-0221/3/07/P07007

[4] Aad, G. et al. (ATLAS Collaboration): The ATLAS Experiment at the CERN Large Hadron Collider. In: *JINST* 3 (2008), p. S08003. doi: 10.1088/1748-0221/3/08/S08003

[5] Aad, G. et al. (ATLAS Collaboration): Performance of the ATLAS b-tagging algorithms. ATL-PHYS-PUB-2009-018, ATL-COM-PHYS-2009-206. 2009. http://cds.cern.ch/record/1174257

[6] Aad, G. et al. (ATLAS Collaboration): Drift time measurement in the ATLAS liquid argon electromagnetic calorimeter using cosmic muons. In: *Eur. Phys. J.* C 70 (2010), p. 755. doi: 10.1140/epjc/s10052-010-1403-6

[7] Aad, G. et al. (ATLAS Collaboration): Electron performance measurements with the ATLAS detector using the 2010 LHC proton–proton collision data. In: *Eur. Phys. J.* C 72 (2012), p. 1909. doi: 10.1140/epjc/s10052-012-1909-1

[8] Aad, G. et al.: Searches for heavy long-lived sleptons and R-hadrons with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV. In: *Phys. Lett.* B 720 (2013), p. 277. doi: 10.1016/j.physletb.2013.02.015

[9] Aad, G. et al. (ATLAS Collaboration): Performance of b-jet identification in the ATLAS experiment. In: *JINST* 11 (2016), p. P04008. doi: 10.1088/1748-0221/11/04/P04008

[10] Aamodt, K. et al. (ALICE Collaboration): The ALICE experiment at the CERN LHC. In: *JINST* 3 (2008), p. S08002. doi: 10.1088/1748-0221/3/08/S08002

[11] Aarnio, P.A. et al. (DELPHI Collaboration): The DELPHI detector at LEP. In: *Nucl. Inst. Meth. A* 303 (1991), p. 233. doi: 10.1016/0168-9002(91)90793-P

[12] Aartsen, M. G. et al. (IceCube Collaboration): The IceCube Neutrino Observatory: instrumentation and online systems. In: *JINST* 12 (2017), p. P03012. doi: 10.1088/1748-0221/12/03/P03012

[13] Abachi, S. et al. (D0 Collaboration): The D0 Detector. In: *Nucl. Inst. Meth. A* 338 (1994), p. 185. doi: 10.1016/0168-9002(94)91312-9

[14] Abazov, V.M. et al. (D0 Collaboration): The upgraded D0 Detector. In: *Nucl. Inst. Meth. A* 565 (2006), p. 463. doi: 10.1016/j.nima.2006.05.248

[15] Abbasi, R. et al. (IceCube Collaboration): The IceCube data acquisition system: Signal capture, digitization, and timestamping. In: *Nucl. Inst. Meth. A* 601 (2009), p. 294. doi: 10.1016/j.nima.2009.01.001

[16] Abbasi, R. et al. (IceCube Collaboration): IceTop: The surface component of IceCube. In: *Nucl. Inst. Meth. A* 700 (2013), p. 188. doi: 10.1016/j.nima.2012.10.067

[17] Abbasi, R.U. et al. (Telescope Array): Measurement of the proton-air cross section with Telescope Array's Middle Drum detector and surface array in hybrid mode. In: *Phys. Rev. D* 92 (2015), p. 032007. doi: 10.1103/PhysRevD.92.032007

[18] Abbiendi, G. et al. (OPAL): Precision luminosity for Z0 line shape measurements with a silicon tungsten calorimeter. In: *Eur. Phys. J.* C 14 (2000), p. 373. doi: 10.1007/s100520000353

[19] Abbott, B. et al. (ATLAS IBL): Production and Integration of the ATLAS Insertable B-Layer. In: *JINST* 13 (2018), p. T05008. doi: 10.1088/1748-0221/13/05/t05008

[20] Abdurashitov et al. (SAGE Collaboration): Results from SAGE. In: *Phys. Lett.* B 328 (1994), p. 234. doi: 10.1016/0370-2693(94)90454-5

[21] Abe, F. et al. (CDF Collaboration): The CDF detector: an overview. In: *Nucl. Inst. Meth. A* 271 (1988), p. 387. doi: 10.1016/0168-9002(88)90298-7

[22] Abe, K. et al. (SLD Collaboration): Measurements of R(b) with impact parameters and displaced vertices. In: *Phys. Rev. D* 53 (1996), p. 1023. doi: 10.1103/PhysRevD.53.1023

[23] Abe, K. et al. (T2K): The T2K Experiment. In: *Nucl. Inst. Meth. A* 659 (2011), p. 106. doi: 10.1016/j.nima.2011.06.067

[24] Abe, T. et al.: R&D status of HAPD. Proc.: *International Workshop on New Photon Detectors (PD09),* Matsumoto, *Japan.* In: *PoS PD09* (2009), p. 014. doi: 10.22323/1.090.0014

[25] Abe, T. et al. (Belle-II Collaboration): Belle II Technical Design Report. In: *arXiv:1011.0352* (2010).

[26] Abelev, B. et al. (ALICE Collaboration): Centrality dependence of π, K, p production in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. In: *Phys. Rev.* C 88 (2013), p. 044910. doi: 10.1103/PhysRevC.88.044910

[27] Abelev, B. et al. (ALICE Collaboration): Performance of the ALICE Experiment at the CERN LHC. In: *Int. J. Mod. Phys.* A 29 (2014), p. 1430044. doi: 10.1142/S0217751X14300440

[28] Abgrall, N. et al. (T2K ND280 TPC collaboration): Time projection chambers for the T2K near detectors. In: *Nucl. Inst. Meth. A* 637 (2011), p. 25. doi: 10.1016/j.nima.2011.02.036

[29] Abraham, J. et al. (Pierre Auger Collaboration): Properties and performance of the prototype instrument for the Pierre Auger Observatory. In: *Nucl. Inst. Meth. A* 523 (2004), p. 50. doi: 10.1016/j.nima.2003.12.012

[30] Abraham, J. et al. (Pierre Auger Collaboration): The fluorescence detector of the Pierre Auger Observatory. In: *Nucl. Inst. Meth. A* 620 (2010), p. 227. doi: 10.1016/j.nima.2010.04.023

[31] Abramowicz, H. et al.: The response and resolution of an iron scintillator calorimeter for hadronic and electromagnetic showers between 10 GeV and 140 GeV. In: *Nucl. Inst. Meth.* 180 (1981), p. 429. doi: 10.1016/0029-554X(81)90083-5

[32] Abrams, G.S. et al.: The Mark-II Detector for the SLC. In: *Nucl. Inst. Meth. A* 281 (1989), p. 55. doi: 10.1016/0168-9002(89)91217-5

[33] Abreu, P. et al. (DELPHI Collaboration): Performance of the DELPHI detec-

tor. In: *Nucl. Inst. Meth. A* 378 (1996), p. 57. doi: 10.1016/0168-9002(96)00463-9

[34] Abt, I. et al. (H1 Collaboration): The H1 detector at HERA. In: *Nucl. Inst. Meth. A* 386 (1997), p. 310. doi: 10.1016/S0168-9002(96)00893-5

[35] Abt, I. et al. (H1 Collaboration): The tracking, calorimeter and muon detectors of the H1 experiment at HERA. In: *Nucl. Inst. Meth. A* 386 (1997), p. 348. doi: 10.1016/S0168-9002(96)00894-7

[36] Achterberg, A. et al. (IceCube Collaboration): Detection of atmospheric muon neutrinos with the IceCube 9-string detector. In: *Phys. Rev. D* 76 (2007), p. 027101. doi: 10.1103/PhysRevD.76.027101

[37] Acosta, D. et al.: Results of prototype studies for a spaghetti calorimeter. In: *Nucl. Inst. Meth. A* 294 (1990), p. 193. doi: 10.1016/0168-9002(90)91833-W

[38] Acosta, D. et al.: Detection of muons with a lead/scintillating-fiber calorimeter. In: *Nucl. Inst. Meth. A* 320 (1992), p. 128. doi: 10.1016/0168-9002(92)90772-V

[39] Acquafredda, R. et al.: The OPERA experiment in the CERN to Gran Sasso neutrino beam. In: *JINST* 4 (2009), p. P04018. doi: 10.1088/1748-0221/4/04/P04018

[40] Adachi, I. et al.: Study of 144-channel multi-anode hybrid avalanche photodetector for the Belle RICH counter. In: *Nucl. Inst. Meth. A* 623 (2010), p. 285. doi: 10.1016/j.nima.2010.02.223

[41] Adam, I. et al. (DIRC and BaBar Collaborations): The DIRC detector at BaBar. In: *Nucl. Inst. Meth. A* 433 (1999), p. 121. doi: 10.1016/S0168-9002(99)00352-6

[42] Adam, I. et al. (BaBar-DIRC Collaboration): The DIRC particle identification system for the BaBar experiment. In: *Nucl. Inst. Meth. A* 538 (2005), p. 281. doi: 10.1016/j.nima.2004.08.129

[43] Adam, W. et al.: The forward ring imaging Cherenkov detector of DELPHI. In: *Nucl. Inst. Meth. A* 338 (1994), p. 284. doi: 10.1016/0168-9002(94)91314-5

[44] Adam, W. et al.: The ring imaging Cherenkov detector of DELPHI. In: *Nucl. Inst. Meth. A* 343 (1994), p. 68. doi: 10.1016/0168-9002(94)90535-5

[45] Adam, W. et al.: The ring imaging Cherenkov detectors of DELPHI. In: *IEEE Trans. Nucl. Sci.* 42 (1995), p. 499. doi: 10.1109/23.467922

[46] Adam, W. et al.: Radiation hard diamond sensors for future tracking applications. In: *Nucl. Inst. Meth. A* 565 (2006), p. 278. doi: 10.1016/j.nima.2006.05.127

[47] Adamova, D. et al. (CERES Collaboration): The CERES/NA45 radial drift time projection chamber. In: *Nucl. Inst. Meth. A* 593 (2008), p. 203. doi: 10.1016/j.nima.2008.04.056

[48] Adamson, P. et al. (NOvA): First measurement of electron neutrino appearance in NOvA. In: *Phys. Rev. Lett.* 116 (2016). doi: 10.1103/PhysRevLett.116.151806

[49] Adamson, P. et al.: The NuMI neutrino beam. In: *Nucl. Inst. Meth. A* 806 (2016), p. 279. doi: 10.1016/j.nima.2015.08.063

[50] Adeva, B. et al.: The construction of the L3 experiment. In: *Nucl. Inst. Meth. A* 289 (1990), p. 35.

[51] Adinolfi, M. et al.: Performance of the LHCb RICH detector at the LHC. In: *Eur. Phys. J. C* 73 (2013), p. 2431. doi: 10.1140/epjc/s10052-013-2431-9

[52] Adloff, C. et al. (CALICE Collaboration): Tests of a particle flow algorithm with CALICE test beam data. In: *JINST* 6 (2011), p. P07005. doi: 10.1088/1748-0221/6/07/P07005

[53] Adloff, C. et al. (CALICE Collaboration): Calorimetry for lepton collider experiments – CALICE results and activities. In: *arXiv:1212.5127* (2012).

[54] Adragna, P. et al. (ATLAS Collaboration): Testbeam studies of production modules of the ATLAS Tile Calorimeter. In: *Nucl. Inst. Meth. A* 606 (2009), p. 362. doi: 10.1016/j.nima.2009.04.009

[55] Adriani, O. et al.: The Pamela experiment ready for flight. In: *Nucl. Inst. Meth. A* 572 (2007), p. 471. doi: 10.1016/j.nima.2006.10.316

[56] Adriani, O. et al. (PAMELA Collaboration): An anomalous positron abundance in cosmic rays with energies 1.5–100 GeV. In: *Nature* 458 (2009), p. 607. doi: 10.1038/nature07942

[57] Adriani, O. et al.: A statistical procedure for the identification of positrons in the PAMELA experiment. In: *Astropart. Phys.* 34 (2010), p. 1. doi: 10.1016/j.astropartphys.2010.04.007

[58] Affolder, A.A. et al. (CDF Collaboration): CDF central outer tracker. In: *Nucl. Inst. Meth. A* 526 (2004), p. 249. doi: 10.1016/j.nima.2004.02.020

[59] Agakichiev, G. et al.: Performance of the CERES electron spectrometer in the CERN SPS lead beam. In: *Nucl. Inst. Meth. A* 371 (1996), p. 16. doi: 10.1016/0168-9002(95)01135-8

[60] Agakishiev, G. et al. (CERES Collaboration): Performance of the CERES electron spectrometer in the CERN SPS lead beam. In: *Nucl. Inst. Meth. A* 371 (1996), p. 16. doi: 10.1016/0168-9002(95)01135-8

[61] Agnese, R. et al. (SuperCDMS Collaboration): Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS. In: *Phys. Rev. Lett.* 112 (2014), p. 241302. doi: 10.1103/PhysRevLett.112.241302

[62] Ahmed, S.N.: *Physics and Engineering of Radiation Detection.* Academic Press, 2007.

[63] Ahmet, K. et al.: The OPAL detector at LEP. In: *Nucl. Inst. Meth. A* 305 (1991), p. 275. doi: 10.1016/0168-9002(91)90547-4

[64] Ahn, H.S. et al.: The Cosmic Ray Energetics and Mass (CREAM) instrument. In: *Nucl. Inst. Meth. A* 579 (2007), p. 1034. doi: 10.1016/j.nima.2007.05.203

[65] Ahn, H.S. et al.: Measurements of cosmic-ray secondary nuclei at high energies with the first flight of the CREAM balloon-borne experiment. In: *Astropart. Phys.* 30 (2008), p. 133. doi: 10.1016/j.astropartphys.2008.07.010

[66] Aielli, G. et al.: Improving the RPC rate capability. In: *JINST* 11 (2016), p. P07014. doi: 10.1088/1748-0221/11/07/P07014

[67] Aker, E. et al. (Crystal Barrel Collaboration): The Crystal Barrel spectrometer at LEAR. In: *Nucl. Inst. Meth. A* 321 (1992), p. 69. doi: 10.1016/0168-9002(92)90379-I

[68] Akerib, D.S. et al. (CDMS Collaboration): First results from the cryogenic dark matter search in the Soudan Underground Lab. In: *Phys. Rev. Lett.* 93 (2004), p. 211301. doi: 10.1103/PhysRevLett.93.211301

[69] Akerib, D.S. et al. (LUX Collaboration): The Large Underground Xenon (LUX) experiment. In: *Nucl. Inst. Meth. A* 704 (2013), p. 111. doi: 10.1016/j.nima.2012.11.135

[70] Akesson, T. et al.: Properties of a fine-sampling uranium-copper scintillator

hadron calorimeter. In: *Nucl. Inst. Meth. A* 241 (1985), p. 17. doi: 10.1016/0168-9002(85)90513-3

[71] Akhmadaliev, S. et al.: Hadron energy reconstruction for the ATLAS calorimetry in the framework of the non-parametrical method. In: *Nucl. Inst. Meth. A* 480 (2002), p. 508. doi: 10.1016/S0168-9002(01)01229-3

[72] Albrecht, E. et al.: Operation, optimisation, and performance of the DELPHI RICH detectors. In: *Nucl. Inst. Meth. A* 433 (1999), p. 47. doi: 10.1016/S0168-9002(99)00320-4

[73] Albrecht, H. et al. (ARGUS Collaboration): ARGUS: A universal detector at DORIS-II. In: *Nucl. Inst. Meth. A* 275 (1989), p. 1. doi: 10.1016/0168-9002(89)90334-3

[74] Albrecht, H. et al. (ARGUS Collaboration): Physics with ARGUS. In: *Phys. Rep.* 276 (1996), p. 223. doi: 10.1016/S0370-1573(96)00008-7

[75] Albrecht, H. et al. (HERA-B Outer Tracker Group): Aging studies for the large honeycomb drift tube system of the Outer Tracker of HERA-B. In: *Nucl. Inst. Meth. A* 515 (2003), p. 155. doi: 10.1016/j.nima.2003.08.144

[76] Albrecht, H. et al. (HERA-B Outer Tracker Group): The outer tracker detector of the HERA-B Experiment. Part I: Detector. In: *Nucl. Inst. Meth. A* 555 (2005), p. 310. doi: 10.1016/j.nima.2005.09.043

[77] Albrecht, H. et al. (HERA-B Outer Tracker Group): The outer tracker detector of the HERA-B experiment. Part III. Operation and performance. In: *Nucl. Inst. Meth. A* 576 (2007), p. 312. doi: 10.1016/j.nima.2007.03.025

[78] Alexander, C.K.; Sadiku, M.N.O.: *Fundamentals of Electric Circuits*. 5th ed. McGraw-Hill, 2012.

[79] Alfonsi, M. et al.: High-rate particle triggering with triple-GEM detector. In: *Nucl. Inst. Meth. A* 518 (2004), p. 106. doi: 10.1016/j.nima.2003.10.035

[80] Alici, A. (ALICE Collaboration): The MRPC-based ALICE time-of-flight detector: status and performance. In: *Nucl. Inst. Meth. A* 706 (2013), p. 29. doi: 10.1016/j.nima.2012.05.004

[81] Alig, R.C.; Bloom, S.; Struck, C.W.: Scattering by ionization and phonon emission in semiconductors. In: *Phys. Rev. B* 22 (1980), p. 5565. doi: 10.1103/PhysRevB.22.5565

[82] Alikhanian, A.I.; Avakina, K.M; Garibian, G.M.; Lorikian, M.P.; Shikhlia, K.K.: Detection of X-ray transition radiation by means of a spark chamber. In: *Phys. Rev. Lett.* 25 (1970), p. 635. doi: 10.1103/PhysRevLett.25.635

[83] Allaire, C. (ATLAS Collaboration): A high-granularity timing detector (HGTD) in ATLAS: Performance at the HL-LHC / CERN. ATL-LARG-PROC-2018-003. Feb 2018. https://cds.cern.ch/record/2302827

[84] Allison, W.W.M.; Cobb, J.H.: Relativistic charged particle identification by energy loss. In: *Annu. Rev. Nucl. Part. Sci.* 30 (1980), p. 253. doi: 10.1146/annurev.ns.30.120180.001345

[85] Alme, J. et al.: The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events. In: *Nucl. Inst. Meth. A* 622 (2010), p. 316. doi: 10.1016/j.nima.2010.04.042

[86] Alvarez, L.W. et al.: Search for hidden chambers in the pyramids. In: *Science* 167 (1970), p. 832. doi: 10.1126/science.167.3919.832

[87] Alves Jr., A.A. et al. (LHCb Collaboration): The LHCb detector at the LHC.

In: *JINST* 3 (2008), p. S08005. doi: 10.1088/1748-0221/3/08/S08005

[88] Amaldi, U.: Fluctuations in calorimetry measurements. In: *Phys. Scripta* 23 (1981), p. 409. doi: 10.1088/0031-8949/23/4A/012

[89] Amaldi, U.; Kraft, G.: Recent applications of synchrotrons in cancer therapy with carbon ions. In: *Europhys. News* 36 (2005), p. 114. doi: 10.1051/epn:2005402

[90] Ambats, I. et al. (MINOS Collaboration): The MINOS Detectors Technical Design Report. NUMI-L-337, FERMILAB-DESIGN-1998-02. 1998. http://inspirehep.net/record/492611/files/fermilab-design-1998-02.pdf

[91] Amerio, S. et al. (ICARUS Collaboration): Design, construction and tests of the ICARUS T600 detector. In: *Nucl. Inst. Meth. A* 527 (2004), p. 329. doi: 10.1016/j.nima.2004.02.044

[92] Amole, C. et al. (PICO Collaboration): Dark Matter search results from the PICO-60 C_3F_8 bubble chamber. In: *Phys. Rev. Lett.* 118 (2017), p. 251301. doi: 10.1103/PhysRevLett.118.251301

[93] AMS-02 Collaboration: AMS-02 – The Alpha Magnetic Spectrometer. http://www.ams02.org/

[94] Anderson, C.D.: The positive electron. In: *Phys. Rev.* 43 (1933), p. 491. doi: 10.1103/PhysRev.43.491

[95] Andivahis, L. et al.: A precise calibration of the SLAC 8-GeV spectrometer using the floating wire technique. SLAC-PUB-5753. 1992. http://slac.stanford.edu/pubs/slacpubs/5750/slac-pub-5753.pdf

[96] Andresen, A. et al. (ZEUS Calorimeter Group): Construction and beam test of the ZEUS forward and rear calorimeter. In: *Nucl. Inst. Meth. A* 309 (1991), p. 101. doi: 10.1016/0168-9002(91)90095-8

[97] Andricek, L.; Lutz, G.; Richter, R.; Reiche, M.: Processing of ultra-thin silicon sensors for future e^+e^- linear collider experiments. In: *IEEE Trans. Nucl. Sci.* 51 (2004), p. 1117. doi: 10.1109/TNS.2004.829531

[98] Andrieu, B. et al. (H1 Calorimeter Group): The H1 liquid argon calorimeter system,. In: *Nucl. Inst. Meth. A* 336 (1993), p. 460. doi: 10.1016/0168-9002(93)91257-N

[99] Andrieu, B. et al. (H1 Calorimeter Group): Electron/pion separation with the H1 LAr calorimeters. In: *Nucl. Inst. Meth. A* 344 (1994), p. 492. doi: 10.1016/0168-9002(94)90870-2

[100] Angloher, G. et al.: Results from 730 kg days of the CRESST-II Dark Matter search. In: *Eur. Phys. J.* C 72 (2012), p. 1971. doi: 10.1140/epjc/s10052-012-1971-8

[101] Angloher, G. et al. (CRESST-II): Results on low mass WIMPs using an upgraded CRESST-II detector. In: *Eur. Phys. J.* C 74 (2014), p. 3184. doi: 10.1140/epjc/s10052-014-3184-9

[102] Annenkov, A. A.; Korzhik, M. V.; Lecoq, P.: Lead tungstate scintillation material. In: *Nucl. Inst. Meth. A* 490 (2002), p. 30. doi: 10.1016/S0168-9002(02)00916-6

[103] Annis, P. et al. (CHORUS): Performance and calibration of the CHORUS scintillating fiber tracker and optoelectronics readout system. Proc.: *7th Int. Wire Chamber Conference,* Vienna. In: *Nucl. Inst. Meth. A* 367 (1995), p. 367. doi: 10.1016/0168-9002(95)00734-2

[104] Annis, P. et al. (RD46): Tracking with capillaries and liquid scintillator. Proc.:

5th Int. Conference on Advanced Technology and Particle Physics (ICATPP-5), Villa Olmo,Como, Italy, *1996*. In: *Nucl. Phys. B – Proc. Suppl.* 61 (1998), p. 390. doi: 10.1016/S0920-5632(97)00592-6

[105] Anthony, P. L. et al.: Measurement of dielectric suppression of Bremsstrahlung. In: *Phys. Rev. Lett.* 76 (1996), p. 3550. doi: 10.1103/PhysRevLett.76.3550

[106] Antokhonov, B.A. et al.: A new 1-km^2 EAS Cherenkov array in the Tunka valley. In: *Nucl. Inst. Meth. A* 639 (2011), p. 42. doi: 10.1016/j.nima.2010.09.142

[107] Antoni, T. et al. (KASCADE Collaboration): The cosmic-ray experiment KASCADE. In: *Nucl. Inst. Meth. A* 513 (2003), p. 490. doi: 10.1016/S0168-9002(03)02076-X

[108] Antonioli, P. et al.: The AQUA-RICH atmospheric neutrino experiment. In: *Nucl. Inst. Meth. A* 433 (1999), p. 104. doi: 10.1016/S0168-9002(99)00542-2

[109] Aoki, S. et al.: The fully automated emulsion analysis system. In: *Nucl. Inst. Meth.* B 51 (1990), p. 466. doi: 10.1016/0168-583X(90)90569-G

[110] Apel, W.D. et al.: KASCADE-Grande – Contributions to the 32nd International Cosmic Ray Conference, Beijing, August, 2011. In: *arXiv:1111.5436* (2011).

[111] Appuhn, R.D. et al. (H1 SPACAL Group): The H1 lead/scintillating-fibre calorimeter. In: *Nucl. Inst. Meth. A* 386 (1997), p. 397. doi: 10.1016/S0168-9002(96)01171-0

[112] Aprile, E. et al. (XENON100 Collaboration): The XENON100 dark matter experiment. In: *Astropart. Phys.* 35 (2012), p. 573. doi: 10.1016/j.astropartphys.2012.01.003

[113] Aprile, E. et al. (XENON Collaboration): First dark matter search results from the XENON1T experiment. In: *Phys. Rev. Lett.* 119 (2017), p. 181301. doi: 10.1103/PhysRevLett.119.181301

[114] Aprile, E. et al. (XENON Collaboration): Search for WIMP inelastic scattering off xenon nuclei with XENON100. In: *Phys. Rev. D* 96 (2017), p. 022008. doi: 10.1103/PhysRevD.96.022008

[115] Aprile, E.; Bolotnikov, A.E.; Bolozdynya, A.L.; Doke, T.: *Noble Gas Detectors*. Wiley, 2008. doi: 10.1002/9783527610020

[116] Arai, Y. et al.: Developments of SOI monolithic pixel detectors. In: *Nucl. Inst. Meth. A* 623 (2010), p. 186. doi: 10.1016/j.nima.2010.02.190

[117] Arfken, G.B.; Weber, H.J.: *Mathematical Methods for Physicists*. Academic Press, 2012.

[118] Arrabito, L. et al.: Track reconstruction in the emulsion-lead target of the OPERA experiment using the ESS microscope. In: *JINST* 2 (2007), p. P05004. doi: 10.1088/1748-0221/2/05/P05004

[119] Artru, X.; Yodh, G.B.; Mennessier, G.: Practical theory of the multilayered transition radiation detector. In: *Phys. Rev.* D12 (1975), p. 1289. doi: 10.1103/PhysRevD.12.1289

[120] Ashcroft, N.W.; Mermin, N.D.: *Solid State Physics*. Saunders, 1976.

[121] Askariyan, G. A.: Excess negative charge of an electron–photon shower and its coherent radio emission. In: *Sov. Phys. JETP* 14 (1962), p. 441.

[122] Assamagan, K.A. et al. (ATLAS Collaboration): Muons in the calorimeters: Energy loss corrections and muon tagging. ATL-PHYS-PUB-2009-009. 2009. http://cds.cern.ch/record/1169055

[123] Assran, Y.; Sharma, A.: Transport properties of operational gas mixtures

used at LHC. In: *arXiv:1110.6761* (2011).
[124] ATLAS Collaboration: ATLAS liquid argon calorimeter: Technical design report. Version: 1996. http://cds.cern.ch/record/331061
[125] ATLAS Collaboration: Particle identification performance of the ATLAS transition radiation tracker. ATLAS-CONF-2011-128. 2011. http://cds.cern.ch/record/1383793
[126] Atwood, W. et al.: Performance of the ALEPH time projection chamber. In: *Nucl. Inst. Meth. A* 306 (1991), p. 446.
[127] Atwood, W.B. et al. (LAT Collaboration): The Large Area Telescope on the Fermi Gamma-ray Space Telescope Mission. In: *Astropphys. J.* 697 (2009), p. 1071. doi: 10.1088/0004-637X/697/2/1071
[128] Aubert, B. et al. (BABAR Collaboration): The BaBar detector. In: *Nucl. Inst. Meth. A* 479 (2002), p. 1. doi: 10.1016/S0168-9002(01)02012-5
[129] Aubert, Bernard et al. (ATLAS Electromagnetic Barrel Liquid Argon Calorimeter Group): Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter. In: *nima* 558 (2006), p. 388. doi: 10.1016/j.nima.2005.11.212
[130] Aubin, F. et al.: Discrimination of nuclear recoils from alpha particles with superheated liquids. In: *New J. Phys.* 10 (2008), p. 103017. doi: 10.1088/1367-2630/10/10/103017
[131] Auger, P.; Ehrenfest, P.; Maze, R.; Daudin, J.; Fréon, R.Ã.: Extensive Cosmic-Ray Showers. In: *Rev. Mod. Phys.* 11 (1939), p. 288. doi: 10.1103/RevModPhys.11.288
[132] Avdeichikov, V; Fomichev, A.S.; Jakobsson, B; Rodin, A.M; Ter-Akopian, G.M.: Range–energy relation, range straggling and response function of CsI(Tl), BGO and GSO(Ce) scintillators for light ions. In: *Nucl. Inst. Meth. A* 439 (2000), p. 158. doi: 10.1016/S0168-9002(99)00944-4
[133] Avdeichikov, V.; Jakobsson, B.; Nikitin, V.A.; Nomokonov, P.V.; Wegner, A.: Systematics in the light response of BGO, CsI(Tl) and GSO(Ce) scintillators to charged particles. In: *Nucl. Inst. Meth. A* 484 (2002), p. 251. doi: 10.1016/S0168-9002(01)01963-5
[134] Avery, Paul: Fitting Theory Writeups and References. http://www.phys.ufl.edu/~avery/fitting.html
[135] Avoni, G. et al.: The electromagnetic calorimeter of the HERA-B experiment. In: *Nucl. Inst. Meth. A* 580 (2007), p. 1209. doi: 10.1016/j.nima.2007.06.030
[136] Axen, D. et al.: The lead–liquid argon sampling calorimeter of the SLD detector. In: *Nucl. Inst. Meth. A* 328 (1993), p. 472. doi: 10.1016/0168-9002(93)90664-4
[137] Ayzenshtat, A.I.; Budnitsky, D.L.; Koretskaya, O.B.; Okaevich, L.S.; Novikov, V.A. et al.: GaAs as a material for particle detectors. In: *Nucl. Inst. Meth. A* 494 (2002), p. 120. doi: 10.1016/S0168-9002(02)01455-9
[138] Bacci, C. et al.: Results from the analysis of data collected with a 50 m^2 RPC carpet at YangBaJing. In: *Nucl. Inst. Meth. A* 456 (2000), p. 121. doi: 10.1016/S0168-9002(00)00976-1
[139] Bachmann, S. et al.: Charge amplification and transfer processes in the gas electron multiplier. In: *Nucl. Inst. Meth. A* 438 (1999), p. 376. doi: 10.1016/S0168-9002(99)00820-7
[140] Bagaturia, Y. et al. (HERA-B Inner Tracker Collaboration): Studies of aging

and HV break down problems during development and operation of MSGC and GEM detectors for the inner tracking system of HERA-B. In: *Nucl. Inst. Meth. A* 490 (2002), p. 223.

[141] Bähr, A.; Kluck, H.; Ninkovic, J.; Schieck, J.; Treis, J.: DEPFET detectors for direct detection of MeV Dark Matter particles. In: *Eur. Phys. J.* C 77 (2017), p. 905. doi: 10.1140/epjc/s10052-017-5474-5

[142] Bailey, R. et al.: First measurement of efficiency and precision of CCD detectors for high-energy physics. In: *Nucl. Inst. Meth.* 213 (1983), p. 201. doi: 10.1016/0167-5087(83)90413-1

[143] Baird, A. et al.: A fast high resolution track trigger for the H1 experiment. In: *IEEE Trans. Nucl. Sci.* 48 (2001), p. 1276. doi: 10.1109/23.958765

[144] Balagura, V. et al.: The first-level trigger of the HERA-B experiment: Performance and expectations. In: *Nucl. Inst. Meth. A* 494 (2002), p. 526. doi: 10.1016/S0168-9002(02)01544-9

[145] Ballabriga, R.; Campbell, M.; Heijne, E.H.M.; Llopart, X.; Tlustos, L.: The Medipix3 prototype, a pixel readout chip working in single photon counting mode with improved spectrometric performance. In: *IEEE Trans. Nucl. Sci.* NS-54 (2007), p. 1824. doi: 10.1109/TNS.2007.906163

[146] Ballin, J.A. et al.: Monolithic active pixel sensors (MAPS) in a quadruple well technology for nearly 100 per cent fill factor and full CMOS pixels. In: *Sensors* 8 (2008), p. 5336. doi: 10.3390/s8095336

[147] Bäni, L. et al. (RD42): Diamond detectors for high energy physics experiments. Proc.: *11th International Conference on Position Sensitive Detectors (PSD11):* Milton Keynes, UK, *2017*. In: *JINST* 13 (2018), p. C01029. doi: 10.1088/1748-0221/13/01/C01029

[148] Barbarino, G.C.; Cerrito, L.; Paternoster, G.; Patricelli, S.: Measurement of the second coordinate in a drift chamber using the charge division method. In: *Nucl. Inst. Meth.* 179 (1981), p. 353. doi: 10.1016/0029-554X(81)90060-4

[149] Barbero, M. et al. (RD42 Collaboration): Development of diamond tracking detectors for high luminosity experiments at LHC. CERN-LHCC-2007-002, CERN-LHCC-RD-012. 2007. http://cds.cern.ch/record/1009654

[150] Barbero, M. et al. (RD42 Collaboration): Development of Diamond Tracking Detectors for High Luminosity Experiments at the LHC. CERN/LHCC 2008-005, LHCC-RD-016. 2008. http://cds.cern.ch/record/1098155

[151] Barbosa Marinho, P.R. et al. (LHCb Collaboration): LHCb: Outer tracker technical design report. CERN-LHCC-2001-024. 2001. http://cds.cern.ch/record/519146

[152] Bardeen, J.; Shockley, W.: Deformation potentials and mobilities in non-polar crystals. In: *Phys. Rev.* 80 (1950), p. 72. doi: 10.1103/PhysRev.80.72

[153] Barkas, W.H.; Birnbaum, W.; Smith, F.M.: Mass-ratio method applied to the measurement of L-meson masses and the energy balance in pion decay. In: *Phys. Rev.* 101 (1956), p. 778. doi: 10.1103/PhysRev.101.778

[154] Barlow, A. et al.: Instrumentation and beam diagnostics in the ISR. Proc.: *8th International Conference on High-Energy Accelerators (HEACC 71)*. In: *eConf* C710920 (1971), p. 426. http://cds.cern.ch/record/309851

[155] Barlow, R.J.: *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Wiley, 2013 (Manchester Physics Series).

[156] Barnes, V.E. et al.: Observation of a hyperon with strangeness minus three. In: *Phys. Rev. Lett.* 12 (1964), p. 204. doi: 10.1103/PhysRevLett.12.204
[157] Barnett, R.M. et al. (Particle Data Group): Review of particle physics. In: *Phys. Rev.* D 54 (1996), p. 1. doi: 10.1103/PhysRevD.54.1
[158] Barrelet, E. et al.: A two-dimensional, single-photoelectron drift detector for Cherenkov ring imaging. In: *Nucl. Inst. Meth.* 200 (1982), p. 219. doi: 10.1016/0167-5087(82)90434-3
[159] Barszczak, T.: Images of Super-Kamiokande events from tscan. http://www.ps.uci.edu/~tomba/sk/tscan/pictures.html
[160] Basolo, S. et al.: A 20 kpixels CdTe photon-counting imager using XPAD chip. In: *Nucl. Inst. Meth. A* 589 (2008), p. 268. doi: 10.1016/j.nima.2008.02.042
[161] Bauer, C. et al.: The HERA-B vertex detector system. In: *Nucl. Inst. Meth. A* 453 (2000), p. 103. doi: 10.1016/S0168-9002(00)00614-8
[162] Baur, R. et al.: The CERES RICH detector system. In: *Nucl. Inst. Meth. A* 343 (1994), p. 87. doi: 10.1016/0168-9002(94)90537-1
[163] Bazylev, V.A. et al.: X-ray Cherenkov radiation: Theory and experiment. In: *Sov. Phys. JETP* 54 (1981), p. 884. http://jetp.ac.ru/cgi-bin/e/index/e/54/5/p884?a=list
[164] Beatty, J.J.; Matthews, J.; Wakely, S.P.: Cosmic rays. In: *[765]* (2016), p. 421.
[165] Becker, C. et al.: Gate-controlled diodes for characterization of the Si–SiO_2 interface with respect to surface effects of silicon detectors. In: *Nucl. Inst. Meth. A* 444 (2000), p. 605. doi: 10.1016/S0168-9002(99)01177-8
[166] Becquerel, H.: Nobel Lecture: On Radioactivity, a New Property of Matter. http://www.nobelprize.org/nobel_prizes/physics/laureates/1903/becquerel-lecture.html
[167] Behnke, E. et al. (COUPP Collaboration): First dark matter search results from a 4-kg CF_3I bubble chamber operated in a deep underground site. In: *Phys. Rev.* D 86 (2012), p. 052001. doi: 10.1103/PhysRevD.86.052001
[168] Behnke, T. et al.: The International Linear Collider Technical Design Report – Volume 1: Executive Summary. In: *arXiv:1306.6327* (2013).
[169] Behnke, T. et al.: The International Linear Collider Technical Design Report – Volume 4: Detectors. In: *arXiv:1306.6329* (2013).
[170] Beitzel, V. et al.: The transition radiation detector for ZEUS. In: *Nucl. Inst. Meth. A* 323 (1992), p. 135. doi: 10.1016/0168-9002(92)90279-D
[171] Belau, E. et al.: Charge collection in silicon strip detectors. In: *Nucl. Inst. Meth.* 214 (1983), p. 253. doi: 10.1016/0167-5087(83)90591-4
[172] Bell, Kenneth W. et al.: The development of vacuum phototriodes for the CMS electromagnetic calorimeter. In: *Nucl. Inst. Meth. A* 469 (2001), p. 29. doi: 10.1016/S0168-9002(01)00700-8
[173] Bell, K.W. et al.: Vacuum phototriodes for the CMS electromagnetic calorimeter endcap. In: *IEEE Trans. Nucl. Sci.* 51 (2004), p. 2284. doi: 10.1109/TNS.2004.836053
[174] Bellamy, E. H. et al.: Absolute calibration and monitoring of a spectrometric channel using a photomultiplier. In: *Nucl. Inst. Meth. A* 339 (1994), p. 468. doi: 10.1016/0168-9002(94)90183-X
[175] Bencivenni, G. et al.: Performance of a test prototype for MONOLITH. In: *Nucl. Inst. Meth. A* 461 (2001), p. 319. doi: 10.1016/S0168-9002(00)01232-8

[176] Bencivenni, G. et al.: A triple GEM detector with pad readout for high rate charged particle triggering. In: *Nucl. Inst. Meth. A* 488 (2002), p. 493. doi: 10.1016/S0168-9002(02)00515-6

[177] Benitez, J. et al.: Status of the Fast Focusing DIRC (fDIRC). In: *Nucl. Inst. Meth. A* 595 (2008), p. 104. doi: 10.1016/j.nima.2008.07.042

[178] Benoit, M. et al.: Testbeam results of irradiated ams H18 HV-CMOS pixel sensor prototypes. In: *JINST* 13 (2018), p. P02011. doi: 10.1088/1748-0221/13/02/P02011

[179] Berdalovic, I. et al.: Monolithic pixel development in TowerJazz 180 nm CMOS for the outer pixel layers in the ATLAS experiment. In: *JINST* 13 (2018), p. C01023. doi: 10.1088/1748-0221/13/01/C01023

[180] Berestetskii, V.B.; Lifshitz, E.M.; Pitaevskii, L.P.: *Quantum Electrodynamics*. 2nd ed. Pergamon, 1982 (Course of theoretical physics Vol. 4).

[181] Berger, M. J. et al.: XCOM: Photon Cross Sections Database (interactive). http://physics.nist.gov/PhysRefData/Xcom/Text/XCOM.html

[182] Berger, M.J.; Coursey, J.S.; Zucker, M.A.; Chang, J.: ESTAR, PSTAR, and ASTAR: computer programs for calculating stopping-power and range tables for electrons, protons, and helium ions. ICRU Report 49, 37 (version 1.2.3). 2005. – Tables and graphical presentations are interactively available: http://www.nist.gov/pml/data/star.

[183] Bernabei, R. et al.: Final model independent result of DAMA/LIBRA–phase1. In: *Eur. Phys. J.* C 73 (2013), p. 2648. doi: 10.1140/epjc/s10052-013-2648-7

[184] Bernlöhr, K.: Simulation of imaging atmospheric Cherenkov telescopes with CORSIKA and sim_telarray. In: *Astropart. Phys.* 30 (2008), p. 149. doi: 10.1016/j.astropartphys.2008.07.009

[185] Bernstein, D. et al.: The MARK-III spectrometer. In: *Nucl. Inst. Meth. A* 226 (1984), p. 301. doi: 10.1016/0168-9002(84)90043-3

[186] Bertolin, A. et al.: The RPC system of the OPERA experiment. In: *Nucl. Inst. Meth. A* 602 (2009), p. 631. doi: 10.1016/j.nima.2008.12.071

[187] Bethe, H.: Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. In: *Ann. Phys.* 5 (1930), p. 325. doi: 10.1002/andp.19303970303

[188] Bethe, H.: Bremsformel für Elektronen relativistischer Geschwindigkeit. In: *Z. Phys.* 76 (1932), p. 293. doi: 10.1007/BF01342532

[189] Bethe, H.; Heitler, W.: On the stopping of fast particles and on the creation of positive electrons. In: *Proc. Roy. Soc. A* 146 (1934), p. 83. doi: 10.1098/rspa.1934.0140

[190] Bettini, A.: *Introduction to Elementary Particle Physics*. Cambridge University Press, 2008.

[191] Betts, W.; Gong, W.; Hjort, E.; Wieman, H.: Studies of Several Wire and Pad Configurations for the STAR TPC. STAR Note 263. 1996. http://drupal.star.bnl.gov/STAR/starnotes/public/sn0263

[192] Beynon, J.D.E.; Lamb, D.R.: *Charge-Coupled Devices and their Applications*. McGraw-Hill, 1980.

[193] Bhat, M.; Pattison, J.; Bibbo, G.; Caon, M.: Diagnostic x-ray spectra: A comparison of spectra generated by different computational methods with a measured spectrum. In: *Med. Phys.* 25 (1998), p. 114. doi: 10.1118/1.598170

[194] Bheesette, S.: *Design and characterisation studies of resistive plate chambers*, Indian Inst. Tech., Mumbai, India, Diss., 2009. http://www.ino.tifr.res.in/in

o/theses/2009/satya_thesis.pdf
[195] Biagi, S.: Magboltz – transport of electrons in gas mixtures. http://consult.cern.ch/writeup/magboltz/
[196] Biagi, S.F.: Monte Carlo simulation of electron drift and diffusion in counting gases under the influence of electric and magnetic fields. In: *Nucl. Inst. Meth. A* 421 (1999), p. 234. doi: 10.1016/S0168-9002(98)01233-9
[197] Bibber, K. van: Scaling up the search for dark matter. In: *Physics* 2 (2009), p. 2. doi: 10.1103/Physics.2.2
[198] Bichsel, H.: Multiple scattering of protons. In: *Phys. Rev.* 112 (1958), p. 182. doi: 10.1103/PhysRev.112.182
[199] Bichsel, H.: Straggling in thin silicon detectors. In: *Rev. Mod. Phys.* 60 (1988), p. 663. doi: 10.1103/RevModPhys.60.663
[200] Bichsel, H.: Charged-particle-matter interactions. In: Drake, G. (ed.): *Springer Handbook of Atomic, Molecular, and Optical Physics*. Springer, 2006, p. 1373. doi: 10.1007/978-0-387-26308-3_91
[201] Bichsel, H.; Groom, D.E.; Klein, S.R.: Passage of particles through matter. In: *[746]* (2014), p. 398.
[202] Biebel, O. et al.: Performance of the OPAL jet chamber. In: *Nucl. Inst. Meth. A* 323 (1992), p. 169. doi: 10.1016/0168-9002(92)90284-B
[203] Binder, E.: *Test eines Flüssig-Argon-Kalorimeters für den H1-Detektor mit Untersuchungen zur Kompensation durch Softwaremethoden*, Universität Hamburg, diploma thesis, 1990. http://www-library.desy.de/preparch/desy/int_rep/f21-90-02.pdf
[204] Binkley, M.E. et al.: Aging in large CDF tracking chambers. In: *Nucl. Inst. Meth. A* 515 (2003), p. 53. doi: 10.1016/j.nima.2003.08.130
[205] Bionta, R.M. et al.: Observation of a neutrino burst in coincidence with supernova SN 1987a in the Large Magellanic Cloud. In: *Phys. Rev. Lett.* 58 (1987), p. 1494. doi: 10.1103/PhysRevLett.58.1494
[206] Birks, J.B.: *The Theory and Practice of Scintillation Counting*. Pergamon Press, 1964.
[207] Blackett, P.M.S.: Nobel Lecture: Cloud Chamber Researches in Nuclear Physics and Cosmic Radiation. http://www.nobelprize.org/nobel_prizes/physics/laureates/1948/blackett-lecture.html
[208] Blackett, P.M.S.; Lees, D.S.: Investigations with a Wilson chamber. I. On the photography of artificial disintegration collisions. In: *Proc. Roy. Soc. A* 136 (1932), p. 325. doi: 10.1098/rspa.1932.0084
[209] Blackett, P.M.S.; Occhialini, G.: Photography of penetrating corpuscular radiation. In: *Nature* 130 (1932), p. 363. doi: 10.1038/130363a0
[210] Blackett, P.M.S.; Occhialini, G.P.S.: Some Photographs of the tracks of penetrating radiation. In: *Proc. Roy. Soc. A* 139 (1933), p. 699. doi: 10.1098/rspa.1933.0048
[211] Bláha, B. et al.: Electrical properties of semiconducting glass. In: *Nucl. Inst. Meth. A* 416 (1998), p. 345. doi: 10.1016/S0168-9002(98)00701-3
[212] Blanc, A.: Recherches sur les mobilites des ions dans les gaz. In: *J. Phys. Theor. Appl.* 7 (1908), p. 825. doi: 10.1051/jphystap:019080070082501
[213] Bloch, F.: Bremsvermögen von Atomen mit mehreren Elektronen. In: *Z. Phys.* 81 (1933), p. 363. doi: 10.1007/BF01344553
[214] Bloch, F.: Zur Bremsung rasch bewegter Teilchen beim Durchgang durch

Materie. In: *Ann. Phys.* 16 (1933), p. 285. doi: 10.1002/andp.19334080303

[215] Bloom, E.D.; Feldman, G.J.: Quarkonium. In: *Sci. Am.* 246 (1982), p. 66. doi: 10.1038/scientificamerican0582-66

[216] Bloom, E.D.; Peck, C.: Physics with the Crystal Ball detector. In: *Annu. Rev. Nucl. Part. Sci.* 33 (1983), p. 143. doi: 10.1146/annurev.ns.33.120183.001043

[217] Blum, W.; Rolandi, L.; Riegler, W.: *Particle Detection with Drift Chambers.* Springer, 2008.

[218] Blümer, H.: *Messungen zur Kalorimetrie von Elektronen und Hadronen*, Universiät Dortmund, diploma thesis, 1982.

[219] Blümer, J.; Engel, R.; Hörandel, J.R.: Cosmic rays from the knee to the highest energies. In: *Prog. Part. Nucl. Phys.* 63 (2009), p. 293. doi: 10.1016/j.ppnp.2009.05.002

[220] Bock, P. et al.: Signal propagation in long wire chambers. In: *JINST* 7 (2012), p. P09003. doi: 10.1088/1748-0221/7/09/P09003

[221] Bode, H.W.: *Network Analysis and Feedback Amplifier Design.* Van Nostrand, 1945.

[222] Boerner, H. et al.: The large cylindrical drift chamber of TASSO. In: *Nucl. Inst. Meth.* 176 (1980), p. 151. doi: 10.1016/0029-554X(80)90695-3

[223] Boger, J. et al. (SNO Collaboration): The Sudbury neutrino observatory. In: *Nucl. Inst. Meth. A* 449 (2000), p. 172. doi: 10.1016/S0168-9002(99)01469-2

[224] Bohm, G.; Zech, G.: *Introduction to Statistics and Data Analysis for Physicists.* DESY, 2010. http://www-library.desy.de/preparch/books/vstatmp_engl.pdf

[225] Bohm, J. et al.: High rate operation and lifetime studies with microstrip gas chambers. In: *Nucl. Inst. Meth. A* 360 (1995), p. 34. doi: 10.1016/0168-9002(94)01218-0

[226] Bohr, N.: II. On the theory of the decrease of velocity of moving electrified particles on passing through matter. In: *Phil. Mag.* 25 (1913), p. 10. doi: 10.1080/14786440108634305

[227] Bohr, N.: *The Penetration of Atomic Particles through Matter.* Copenhagen: Munksgaard, 1948 (Det kgl. Danske videnskabernes selskab. Mathematisk-fysiske meddelelser. Vol. 18, no. 8).

[228] Boie, R.A.; Hrisoho, A.T.; Rehak, P.: Signal shaping and tail cancellation for gas proportional detectors at high counting rates. In: *IEEE Trans. Nucl. Sci.* 28 (1981), p. 603. doi: 10.1016/0029-554X(82)90846-1

[229] Bollinger, L.M.; Thomas, G.E.: Measurement of time dependence of scintillation intensity by a delayed-coincidence method. In: *Rev. Sci. Inst.* 32 (1961), p. 1044. doi: 10.1063/1.1717610

[230] Bolmont, J.; Voigt, B.; Nahnhauer, R. (for the IceCube Collaboration): Very high energy electromagnetic cascades in the LPM regime with IceCube. Proc.: *30th Int. Cosmic Ray Conference,* Merida, Mexico, vol. 3 (2007), p. 1245. http://indico.nucleares.unam.mx/event/4/contribution/843. *arXiv:0711.0353*

[231] Bolotovskii, B.M. (ed.); Frenkel, V.Ya. (ed.); Peierls, R. (ed.): *I. E. Tamm – Selected Papers.* Springer, 1991. doi: 10.1007/978-3-642-74626-0

[232] Bolozdynya, A.I.: Two-phase emission detectors and their applications. In: *Nucl. Inst. Meth. A* 422 (1999), p. 314. doi: 10.1016/S0168-9002(98)00965-6

[233] Bondarenko, G. et al.: Limited Geiger-mode microcell silicon photodiode:

New results. In: *Nucl. Inst. Meth. A* 442 (2000), p. 187. doi: 10.1016/S0168-9002(99)01219-X

[234] Börner, H.: *Die zylindrische Driftkammer des TASSO-Experiments am e^+e^--Speicherring PETRA*, Universität Bonn, Dissertation, 1981.

[235] Bornheim, A.: On the usage of precision timing detectors in high rate and high pileup environments. In: *PoS* Vertex2016 (2017), p. 044. doi: 10.22323/1.287.0044

[236] Bortfeldt, J. et al.: PICOSEC: Charged particle timing at sub-25 picosecond precision with a Micromegas based detector. In: *Nucl. Inst. Meth. A* 903 (2018), p. 317. doi: 10.1016/j.nima.2018.04.033

[237] Bothe, W.: Nobel Lecture: The Coincidence Method. http://www.nobelprize.org/nobel_prizes/physics/laureates/1954/bothe-lecture.html

[238] Bothe, W.: Zur Vereinfachung von Koinzidenzzählungen. In: *Z. Phys.* 59 (1930), p. 1. doi: 10.1007/BF01337830

[239] Bothe, W.; Kolhörster, W.: Das Wesen der Höhenstrahlung. In: *Z. Phys.* 56 (1929), p. 751. doi: 10.1007/BF01340137

[240] Bouhova-Thacker, E. et al.: Vertex reconstruction in the ATLAS experiment at the LHC. ATL-INDET-PUB-2009-001, ATL-COM-INDET-2009-011. 2009. http://cds.cern.ch/record/1176154

[241] Boyle, P.J.: The elemental composition of high-energy cosmic rays: measurements with TRACER. In: *Mod. Phys. Lett.* A 23 (2008), p. 2031. doi: 10.1142/S0217732308028260

[242] Boyle, W.S.; Smith, G.E.: Charge coupled semiconductor deadapted devices. In: *Bell Syst. Tech. J.* 49 (1970), p. 587. doi: 10.1002/j.1538-7305.1970.tb01790.x

[243] Braccini, S. et al.: First results on proton radiography with nuclear emulsion detectors. In: *JINST* 5 (2010), p. P09001. doi: 10.1088/1748-0221/5/09/P09001

[244] Braem, A.; Joram, C.; Piuz, F.; Schyns, E.; Seguinot, J.: Technology of photocathode production. In: *Nucl. Inst. Meth. A* 502 (2003), p. 205. doi: 10.1016/S0168-9002(03)00275-4

[245] Brandelik, R. et al. (TASSO Collaboration): Properties of hadron final states in e^+e^- annihilation at 13-GeV and 17-GeV center-of-mass energies. In: *Phys. Lett.* B 83 (1979), p. 261. doi: 10.1016/0370-2693(79)90699-3

[246] Brandt, S.: *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*. Springer, 2014. doi: 10.1007/978-3-319-03762-2

[247] Brau, J.E.; Jaros, J.A.; Ma, H.: Advances in calorimetry. In: *Annu. Rev. Nucl. Part. Sci.* 60 (2010), p. 615. doi: 10.1146/annurev.nucl.012809.104449

[248] Brennan, K.F.: *The Physics of Semiconductors*. Cambridge University Press, 1999.

[249] Breskin, A.; Charpak, G.; Sauli, F.; Atkinson, M.; Schultz, G.: Recent observations and measurements with high accuracy drift chambers. In: *Nucl. Inst. Meth.* 124 (1975), p. 189. doi: 10.1016/0029-554X(75)90403-6

[250] Bretz, T. et al.: An integrated general purpose SiPM based optical module with a high dynamic range. In: *JINST* 13 (2018), p. P06001. doi: 10.1088/1748-0221/13/06/P06001

[251] Breuker, H. et al.: Particle identification with the OPAL jet chamber in the region of the relativistic rise. In: *Nucl. Inst. Meth. A* 260 (1987), p. 329. doi: 10.1016/0168-9002(87)90097-0

[252] Brigham Young University: Cleanroom Home, Optical Absorption Coefficient Calculator. https://cleanroom.byu.edu/OpticalCalc
[253] Brock, I.C. et al.: Luminosity measurement in the L3 detector at LEP. In: *Nucl. Inst. Meth. A* 381 (1996), p. 236. doi: 10.1016/S0168-9002(96)00734-6
[254] Brockmann, R. et al.: Development of a time projection chamber with high two track resolution capability for collider experiments: RD-32 status report. CERN-DRDC-94-10, CERN-RD-32-STATUS-REPORT. 1994. http://cds.cern.ch/record/290990
[255] Bronshtein, I.N.; Semendyayew, K.A.; Musiol, G.; Mühlig, H.: *Handbook of Mathematics*. Springer, 2015. doi: 10.1007/978-3-662-46221-8
[256] Bross, A. et al.: The D0 scintillating fiber tracker. In: *AIP Conf. Proc.* 450 (1998), p. 221. doi: 10.1063/1.56972
[257] Brown, J.S. et al.: The Mark-III time-of-flight system. In: *Nucl. Inst. Meth. A* 221 (1984), p. 503. doi: 10.1016/0167-5087(84)90058-9
[258] Brückmann, H. et al.: Hadron sampling calorimetry, a puzzle of physics. In: *Nucl. Inst. Meth. A* 263 (1988), p. 136. doi: 10.1016/0168-9002(88)91026-1
[259] Budnev, N. et al.: Tunka-25 Air Shower Cherenkov array: the main results. In: *Astropart. Phys.* 50–52 (2013), p. 18. doi: 10.1016/j.astropartphys.2013.09.006
[260] Bureau International des Poids et Mesures (BIPM): The International System of Units (SI). Version: 9th edition, 2019. https://www.bipm.org/en/publications/si-brochure/
[261] Burger, J. et al.: The central jet chamber of the H1 experiment. In: *Nucl. Inst. Meth. A* 279 (1989), p. 217. doi: 10.1016/0168-9002(89)91084-X
[262] Buridon, V. et al. (CALICE): First results of the CALICE SDHCAL technological prototype. In: *JINST* 11 (2016), p. P04001. doi: 10.1088/1748-0221/11/04/P04001
[263] Burkhardt, H. et al.: The Tasso gas and aerogel Cherenkov counters. In: *Nucl. Inst. Meth.* 184 (1981), p. 319. doi: 10.1016/0029-554X(81)90732-1
[264] Buskulic, D. et al. (ALEPH Collaboration): Performance of the ALEPH detector at LEP. In: *Nucl. Inst. Meth. A* 360 (1995), p. 481. doi: 10.1016/0168-9002(95)00138-7
[265] Buton, C. et al.: Comparison of three types of XPAD3.2/CdTe single chip hybrids for hard X-ray applications in material science and biomedical imaging. In: *Nucl. Inst. Meth. A* 758 (2014), p. 44. doi: 10.1016/j.nima.2014.04.067
[266] Buzhan, P. et al.: An advanced study of silicon photomultiplier. In: *ICFA Instrum. Bull.* 23 (2001), p. 28. http://www.slac.stanford.edu/pubs/icfa/fall01/paper3/paper3a.html
[267] Buzhan, P. et al.: Silicon photomultiplier and its possible applications. In: *Nucl. Inst. Meth. A* 504 (2003), p. 48. doi: 10.1016/S0168-9002(03)00749-6
[268] C. Davoisne et al.: Chemical and morphological evolution of a silicate surface under low-energy ion irradiation. In: *A&A* 482 (2008), p. 541. doi: 10.1051/0004-6361:20078964
[269] CADENCE PSPICE. http://www.cadence.com/products/orcad/pspice_simulation/Pages/default.aspx
[270] Campbell, M.: MEDIPIX webpage. https://medipix.web.cern.ch/
[271] Campbell, M. et al.: A readout chip for a 64×64 pixel matrix with 15 bit single photon counting. In: *IEEE Trans. Nucl. Sci.* 45 (1998), p. 751. doi: 10.1109/23.682629

[272] Cardarelli, R.; Santonico, R.; Biagio, A. D.; Lucci, A.: Progress in resistive plate counters. In: *Nucl. Inst. Meth. A* 263 (1988), p. 20. doi: 10.1016/0168-9002(88)91011-X

[273] Carter, J.R. et al.: The OPAL vertex drift chamber. In: *Nucl. Inst. Meth. A* 286 (1990), p. 99. doi: 10.1016/0168-9002(90)90211-N

[274] Cartiglia, N. et al.: Design optimization of ultra-fast silicon detectors. In: *Nucl. Inst. Meth.* A796 (2015), p. 141. doi: 10.1016/j.nima.2015.04.025

[275] Cartiglia, N. et al.: The 4D pixel challenge. Proc.: *8th International Workshop on Semiconductor Pixel Detectors for Particles and Imaging (PIXEL 2016),* Sestri Levante, Italy. In: *JINST* 11 (2016), p. C12016. doi: 10.1088/1748-0221/11/12/C12016

[276] Cartiglia, N. et al.: Beam test results of a 16 ps timing system based on ultra-fast silicon detectors. In: *Nucl. Inst. Meth. A* 850 (2017), p. 83. doi: 10.1016/j.nima.2017.01.021

[277] Cartiglia, N. et al.: Tracking in 4 dimensions. Proc.: *14th Vienna Conference on Instrumentation (VCI 2016),* Vienna, Austria. In: *Nucl. Inst. Meth. A* 845 (2017), p. 47. doi: 10.1016/j.nima.2016.05.078

[278] Carusone, T.; Johns, D.A.; Martin, K.: *Analog Integrated Circuit Design.* Wiley, 2011.

[279] Casse, G.: Radiation hardness of p-type silicon detectors. In: *Nucl. Inst. Meth. A* 612 (2010), p. 464. doi: 10.1016/j.nima.2009.08.050

[280] Cassel, D.G. et al.: Design and construction of the CLEO-II drift chamber. In: *Nucl. Inst. Meth. A* 252 (1986), p. 325. doi: 10.1016/0168-9002(86)91201-5

[281] Caughey, D.M.; Thomas, R.E.: Carrier mobilities in silicon empirically related to doping and field. In: *Proc. IEEE* 55 (1967), p. 2192. doi: 10.1109/PROC.1967.6123

[282] Cavalleri, G.; Gatti, E.; Fabri, G.; Svelto, V.: Extension of Ramo's theorem as applied to induced charge in semiconductor detectors. In: *Nucl. Inst. Meth.* 92 (1971), p. 137. doi: 10.1016/0029-554X(71)90235-7

[283] Cenna, F. et al.: Weightfield2: A fast simulator for silicon and diamond solid state detectors. In: *Nucl. Inst. Meth. A* 796 (2015), p. 149. doi: 10.1016/j.nima.2015.04.015

[284] CERN: High School Teachers at CERN – Bubble Chambers HST2001. Version: 2001. http://teachers.web.cern.ch/teachers/materials/bubblechambers.htm

[285] CERN: The High-Luminosity LHC. Version: 2015. http://cds.cern.ch/record/2114693

[286] Cerron Zeballos, E. et al.: A New type of resistive plate chamber: the multigap RPC. In: *Nucl. Inst. Meth. A* 374 (1996), p. 132. doi: 10.1016/0168-9002(96)00158-1

[287] Chamanina, J. et al.: Si-pixel transition radiation detector with separation of TR-photon and particle track by B-field. LC-DET-2000-083. 2000. https://flc.desy.de/lcnotes/notes/LC-DET-2000-038.pdf

[288] Chamberlain, O.; Segre, E.; Wiegand, C.; Ypsilantis, T.: Observation of antiprotons. In: *Phys. Rev.* 100 (1955), p. 947. doi: 10.1103/PhysRev.100.947

[289] Chantler, C.T. et al.: NIST X-ray form factor and scattering tables. http://physics.nist.gov/PhysRefData/FFast/html/form.html

[290] Charpak, G.: Nobel Lecture: Electronic Imaging of Ionizing Radiation with

Limited Avalanches in Gases. http://www.nobelprize.org/nobel_prizes/physics/laureates/1992/charpak-lecture.html

[291] Charpak, G.: Evolution of the automatic spark chambers. In: *Annu. Rev. Nucl. Part. Sci.* 20 (1970), p. 195. doi: 10.1146/annurev.ns.20.120170.001211

[292] Charpak, G. et al.: First beam test results with Micromegas, a high rate, high resolution detector. In: *Nucl. Inst. Meth. A* 412 (1998), p. 47. doi: 10.1016/S0168-9002(98)00311-8

[293] Charpak, G.; Bouclier, R.; Bressani, T.; Favier, J.; Zupancic, C.: The Use of Multiwire Proportional Counters to Select and Localize Charged Particles. In: *Nucl. Inst. Meth.* 62 (1968), p. 262. doi: 10.1016/0029-554X(68)90371-6

[294] Charpak, G.; Derre, J.; Giomataris, Y.; Rebourgeard, P.: MICROMEGAS, a multipurpose gaseous detector. In: *Nucl. Inst. Meth. A* 478 (2002), p. 26. doi: 10.1016/S0168-9002(01)01713-2

[295] Charpak, G.; Petersen, G.; Policarpo, A.; Sauli, F.: High accuracy measurements in proportional chambers and clarification of the avalanche mechanisms around wires. In: *IEEE Trans. Nucl. Sci.* 25 (1978), p. 122. doi: 10.1109/TNS.1978.4329289

[296] Charpak, G.; Sauli, F.: High-resolution electronic particle detectors. In: *Annu. Rev. Nucl. Part. Sci.* 34 (1984), p. 285. doi: 10.1146/annurev.ns.34.120184.001441

[297] Chase, R. L.; La Taille, C. de; Seguin, N.; Rescia, S.: Transmission line connections between detector and front end electronics in liquid argon calorimetry. In: *Nucl. Inst. Meth. A* 330 (1993), p. 228. doi: 10.1016/0168-9002(93)91327-J

[298] Chatrchyan, S. et al. (CMS Collaboration): The CMS experiment at the CERN LHC. In: *JINST* 3 (2008), p. S08004. doi: 10.1088/1748-0221/3/08/S08004

[299] Chatrchyan, S. et al. (CMS Collaboration): Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. In: *JINST* 7 (2012), p. P10002. doi: 10.1088/1748-0221/7/10/P10002

[300] Chatrchyan, S. et al. (CMS Collaboration): The performance of the CMS muon detector in proton–proton collisions at sqrt(s) = 7 TeV at the LHC. In: *JINST* 8 (2013), p. P11002. doi: 10.1088/1748-0221/8/11/P11002

[301] Chen, W. et al.: Performance of the multianode cylindrical silicon drift detector in the CERES NA45 experiment: first results. In: *Nucl. Inst. Meth. A* 326 (1993), p. 273. doi: 10.1016/0168-9002(93)90363-M

[302] Cherenkov, P.A.: Nobel Lecture: Radiation of particles moving at a velocity exceeding that of light, and some of the possibilities for their use in experimental physics. http://www.nobelprize.org/nobel_prizes/physics/laureates/1958/cerenkov-lecture.html

[303] Cherenkov, P.A.: Visible emission of clean liquids by action of γ radiation. In: *Dokl. Akad. Nauk SSSR* 2 (1934), p. 451.

[304] Cherenkov, P.A.: Visible radiation produced by electron moving in a medium with velocities exceeding that of light. In: *Phys. Rev.* 52 (1937), p. 87. doi: 10.1103/PhysRev.52.378

[305] Cherry, M.L.: Measurements of the spectrum and energy dependence of x-ray transition radiation. In: *Phys. Rev. D* 17 (1978), p. 2245. doi: 10.1103/PhysRevD.17.2245

[306] Cherry, M.L.; Hartmann, G.; Müller, D.; Prince, T.A.: Transition radiation from relativistic electrons in periodic radiators. In: *Phys. Rev. D* 10 (1974), p. 3594. doi: 10.1103/PhysRevD.10.3594

[307] Cherry, M.L.; Mueller, D.; Prince, T.A.: The efficient identification of relativistic particles by transition radiation. In: *Nucl. Inst. Meth.* 115 (1974), p. 141. doi: 10.1016/0029-554X(74)90439-X

[308] Chilingarov, A.: Temperature dependence of the current generated in Si bulk. In: *JINST* 8 (2013), p. P10003. http://stacks.iop.org/1748-0221/8/i=10/a=P10003

[309] Chirkin, D.; Rhode, W.: Muon Monte Carlo: a high-precision tool for muon propagation through matter. In: *arXiv:hep-ph/0407075* (2004).

[310] Christophorou, L.G.; Olthoff, J.K.; Rao, M.V.V.S.: Electron interactions with CF_4. In: *J. Phys. Chem. Ref. Data* 25 (1996), p. 1341. doi: 10.1063/1.555986

[311] Clemen, M.; Humanic, T.; Kraus, D.; Vilkelis, G.; Rehak, P. et al.: Double particle resolution measured in a silicon drift chamber. In: *Nucl. Inst. Meth. A* 316 (1992), p. 283. doi: 10.1016/0168-9002(92)90911-M

[312] Cleveland, B.T. et al.: Measurement of the solar electron neutrino flux with the Homestake chlorine detector. In: *Astropphys. J.* 496 (1998), p. 505. http://stacks.iop.org/0004-637X/496/i=1/a=505

[313] Clevers, R.H.M.: Volume and temperature dependence of the 1/f noise parameter α in Si. In: *Phys. Rev.* B 154 (1988), p. 214. doi: 10.1016/0921-4526(89)90071-9

[314] CLIC Collaboration: The Compact Linear Collider (CLIC) Study. http://clic-study.web.cern.ch/CLIC-Study

[315] CMS Collaboration: CMS Crystal Endcap webpage (Crystal Palace). http://cms-ecal-ee-rc.web.cern.ch/cms-ecal-ee-rc/Index.htm

[316] Cobb, J. et al.: Transition radiators for electron identification at the CERN ISR. In: *Nucl. Inst. Meth.* 140 (1977), p. 413. doi: 10.1016/0029-554X(77)90355-X

[317] Cohen, R.E.; Lide, D.; Trigg, G.: *Physicist's Desk Reference*. Springer, 2003.

[318] Contin, G. et al.: The STAR MAPS-based PiXeL detector. In: *Nucl. Inst. Meth. A* 907 (2018), p. 60. doi: 10.1016/j.nima.2018.03.003

[319] Conwell, E.; Weisskopf, V.F.: Theory of impurity scattering in semiconductors. In: *Phys. Rev.* 77 (1950), p. 388. doi: 10.1103/PhysRev.77.388

[320] Cormack, A.M.: Dead-time losses with pulsed beams. In: *Nucl. Inst. Meth.* 15 (1962), p. 268. doi: 10.1016/0029-554X(62)90086-1

[321] Cowan, G.: Statistics. In: *published in [765], chapter 39* (2016).

[322] Coyle, P. et al.: The DIRC counter: A new type of particle identification device for B factories. In: *Nucl. Inst. Meth. A* 343 (1994), p. 292. doi: 10.1016/0168-9002(94)90568-1

[323] Cozza, D. et al.: The CSI-based RICH detector array for the identification of high momentum particles in ALICE. In: *Nucl. Inst. Meth. A* 502 (2003), p. 101. doi: 10.1016/S0168-9002(02)02163-0

[324] Cramér, H.: *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press, 1945 (Princeton Landmarks in Mathematics and Physics Series).

[325] Crank, J.: *The Mathematics of Diffusion*. 2nd ed., Oxford University Press, 1975.

[326] Craun, R.L.; Smith, D.L.: Analysis of response data for several organic scintillators. In: *Nucl. Inst. Meth.* 80 (1970), p. 239. doi: 10.1016/0029-554X(70)90768-8

[327] Cravens, J.P. et al. (Super-Kamiokande Collaboration): Solar neutrino measurements in Super-Kamiokande-II. In: *Phys. Rev. D* 78 (2008), p. 032002. doi: 10.1103/PhysRevD.78.032002

[328] Crawley, H.B. et al.: Characterization and radiation testing of the Harris HS9008RH flash analog-to-digital converter. In: *Nucl. Inst. Meth. A* 345 (1994), p. 329. doi: 10.1016/0168-9002(94)91010-3

[329] Creusot, A. (ANTARES Collaboration): The Antares detector. In: *Nucl. Inst. Meth. A* 718 (2013), p. 489. doi: 10.1016/j.nima.2012.11.071

[330] Creutz, E. (ed.): *Encyclopedia of Physics*. vol. 45: *Nuclear Instrumentation II*. Springer, 1958. doi: 10.1007/978-3-642-45903-0

[331] Crittenden, J.A. (ZEUS Calorimeter Group): The performance of the ZEUS calorimeter. Proc.: *5th International Conference on Calorimetry in High Energy Physics,* Brookhaven, NY. World Scientific, (1995), p. 58. http://www-library.desy.de/cgi-bin/showprep.pl?desy94-234

[332] Cronin, G.; Haisty, R.: The preparation of semi-insulating gallium arsenide by chromium doping. In: *J. Electrochem. Soc.* 111 (1964), p. 874. doi: 10.1149/1.2426273

[333] CTA Collaboration: CTA – Cherenkov Telescope Array. http://www.cta-observatory.org/

[334] Da Via, C. et al.: Radiation hardness properties of full-3D active edge silicon sensors. In: *Nucl. Inst. Meth. A* 587 (2008), p. 243. doi: 10.1016/j.nima.2007.12.027

[335] Da Via, C. et al.: 3D active edge silicon sensors: Device processing, yield and QA for the ATLAS-IBL production. In: *Nucl. Inst. Meth. A* 699 (2013), p. 18. doi: 10.1016/j.nima.2012.05.070

[336] Dalal, V.L.: Hole Velocity in p-GaAs. In: *Appl. Phys. Lett.* 16 (1970), p. 489. doi: 10.1063/1.1653077

[337] Damerell, C.J.S.: CCD vertex detectors in particle physics. In: *Nucl. Inst. Meth. A* 342 (1994), p. 78. doi: 10.1016/0168-9002(94)91412-5

[338] Damerell, C.J.S. et al.: A CCD based vertex detector for SLD. In: *Nucl. Inst. Meth. A* 288 (1990), p. 236. doi: 10.1016/0168-9002(90)90491-N

[339] Danilov, M. et al.: The ARGUS drift chamber. In: *Nucl. Inst. Meth.* 217 (1983), p. 153. doi: 10.1016/0167-5087(83)90124-2

[340] Danysza, M.; Pniewskia, J.: Delayed disintegration of a heavy nuclear fragment: I. In: *Phil. Mag.* 44 (1953), p. 348. doi: 10.1080/14786440308520318

[341] Davidek, T.; Leitner, R.: Parametrization of the Muon Response in the Tile Calorimeter. CERN-ATL-TILECAL-97-114. 1997. http://cdsweb.cern.ch/record/683578

[342] Davies, H.; Bethe, H.A.; Maximon, L.C.: Theory of bremsstrahlung and pair production. 2. Integral cross section for pair production. In: *Phys. Rev.* 93 (1954), p. 788. doi: 10.1103/PhysRev.93.788

[343] Daya Bay Collaboration: Daya Bay Reactor Neutrino Experiment. http://dayabay.ihep.ac.cn

[344] De Lellis, G. et al.: Measurement of the fragmentation of carbon nuclei used in hadron-therapy. In: *Nucl. Phys. A* 853 (2011), p. 124. doi:

10.1016/j.nuclphysa.2011.01.019
[345] De Lellis, G.; Ereditato, A.; Niwa, K.: Nuclear emulsions. In: *[387]*, p. 262. doi: 10.1007/978-3-642-03606-4_9
[346] Decamp, D. et al. (ALEPH Collaboration): ALEPH: a detector for electron–positron annnihilations at LEP. In: *Nucl. Inst. Meth. A* 294 (1990), p. 121. doi: 10.1016/0168-9002(90)91831-U
[347] DELHI Collaboration: DELPHI. http://delphiwww.cern.ch/delfigs/export/pubdet4.html
[348] Dell'Orso, M. (CDF Collaboration): The CDF silicon vertex trigger. In: *Nucl. Phys. Proc. Suppl.* 156 (2006), p. 139. doi: 10.1016/j.nuclphysbps.2006.02.135
[349] Dell'Orso, M.; Ristori, L.: VLSI structures for track finding. In: *Nucl. Inst. Meth. A* 278 (1989), p. 436. doi: 10.1016/0168-9002(89)90862-0
[350] Derrick, M. et al.: Design and construction of the ZEUS barrel calorimeter. In: *Nucl. Inst. Meth. A* 309 (1991), p. 77. doi: 10.1016/0168-9002(91)90094-7
[351] Dhawan, S.; Majka, R.: Development status of microchannel plate photomultipliers. In: *IEEE Trans. Nucl. Sci.* 24 (1977), p. 270. doi: 10.1109/TNS.1977.4328688
[352] Diehl, E.: Calibration and performance of the precision chambers of the ATLAS muon spectrometer. Proc.: *2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011)*. In: *Physics Procedia* 37 (2012), p. 543. doi: https://doi.org/10.1016/j.phpro.2012.02.404
[353] Dierlamm, A.: Hardness factor kappa. 2018. Private communication.
[354] Diethorn, W.: *A methane proportional counter system for natural radiocarbon measurements*, Carnegie Institute of Technology, Pittsburgh, PA, Diss., 1956. USAEC Report NYO-6628. http://www.osti.gov/scitech/biblio/4345702
[355] Dinu, N.; Nagai, A.; Para, A.: Studies of MPPC detectors down to cryogenic temperatures. In: *Nucl. Inst. Meth. A* 787 (2015), p. 275. doi: 10.1016/j.nima.2014.12.061
[356] Dissertori, G. et al.: A visualization of the damage in lead tungstate calorimeter crystals after exposure to high-energy hadrons. In: *Nucl. Inst. Meth. A* 684 (2012), p. 57. doi: 10.1016/j.nima.2012.04.069
[357] Dolgoshein, B.: Transition radiation detectors. In: *Nucl. Inst. Meth. A* 326 (1993), p. 434. doi: 10.1016/0168-9002(93)90846-A
[358] Dolgoshein, B.A.; Rodionov, B.U.; Luchkov, B.I.: Streamer chamber. In: *Nucl. Inst. Meth.* 29 (1964), p. 270. doi: 10.1016/0029-554X(64)90379-9
[359] Doll, D. et al.: A counting silicon microstrip detector for precision Compton polarimetry. In: *Nucl. Inst. Meth. A* 492 (2002), p. 356. doi: 10.1016/S0168-9002(02)01396-7
[360] Donegani, E.: *Energy-dependent proton damage in silicon*, University of Hamburg, Diss., 2017. http://www.iexp.uni-hamburg.de/groups/pd/?q=theses/energy-dependent-proton-damage-silicon
[361] Dorenbos, P.; Haas, J.T.M. de; Eijk, C.W.E. van: Non-proportionality in the scintillation and the energy resolution with scintillation crystals. In: *IEEE Trans. Nucl. Sci.* 42(6) (1995), p. 2190. doi: 10.1109/23.489415
[362] Double Chooz Collaboration: Double Chooz. http://doublechooz.in2p3.fr
[363] Drasal, Z.; Riegler, W.: An extension of the Gluckstern formulae for multiple scattering: Analytic expressions for track parameter resolution us-

ing optimum weights. In: *Nucl. Inst. Meth. A* 910 (2018), p. 127. doi: 10.1016/j.nima.2018.08.078

[364] Drews, G. et al.: Experimental determination of sampling fluctuations in uranium and lead hadronic calorimeters. In: *Nucl. Inst. Meth. A* 290 (1990), p. 335. doi: 10.1016/0168-9002(90)90549-L

[365] Driscoll, T.: The Schwarz-Christoffel Toolbox for MATLAB. http://www.math.udel.edu/~driscoll/SC/

[366] Drude, P.: Zur Elektronentheorie der Metalle. In: *Ann. Phys.* 306 (1900), p. 566. doi: 10.1002/andp.19003060312

[367] Drumm, H. et al.: Experience with the jet-chamber of the JADE detector at PETRA. In: *Nucl. Inst. Meth.* 176 (1980), p. 333. doi: 10.1016/0029-554X(80)90725-9

[368] Eberhardt, J.E.: Fano factor in silicon at 90 K. In: *Nucl. Inst. Meth.* 80 (1970), p. 291. doi: 10.1016/0029-554X(70)90774-3

[369] Egger, J.; Hildebrandt, M.; Petitjean, C. (MuCap Collaboration): The 10 bar hydrogen time projection chamber of the MuCap experiment. In: *Nucl. Inst. Meth. A* 628 (2011), p. 199. doi: 10.1016/j.nima.2010.06.316

[370] Eidelman, S. et al. (Particle Data Group): Review of particle physics. In: *Phys. Lett.* B 592 (2004), p. 1. doi: 10.1016/j.physletb.2004.06.001

[371] Einstein, A.: Zur Theorie der Brownschen Bewegung. In: *Ann. Phys.* 19 (1906), p. 371.

[372] Ekelof, T.: The experimental method of ring imaging Cherenkov (RICH) counters. Proc.: *12th SLAC Summer Institute on Particle Physics*. Stanford: SLAC. In: *eConf* C840723 (1984), p. 244. http://www.slac.stanford.edu/econf/C840723/pdf/ssi84-006.pdf

[373] EMI Electronics Ltd: EMI photomultiplier tubes. Version: 1970. https://frank.pocnet.net/other/EMI/EMI_Photomultiplier_Tubes_1970.pdf

[374] Engler, J. et al.: A warm-liquid calorimeter for cosmic-ray hadrons. In: *Nucl. Inst. Meth. A* 427 (1999), p. 528. doi: 10.1016/S0168-9002(99)00051-0

[375] Engstrom, R.W.: *Photomultiplier Handbook*. Burle Industries, Inc., 1989. http://psec.uchicago.edu/links.php

[376] Erazo, F.; Lallena, A.M.: Calculation of beam quality correction factors for various thimble ionization chambers using the Monte Carlo code PENELOPE. In: *Physica Medica* 29 (2013), p. 163. doi: 10.1016/j.ejmp.2012.01.001

[377] Ereditato, A.: The study of neutrino oscillations with emulsion detectors. In: *Adv. High Energy Phys.* 2013 (2013), p. 382172. doi: 10.1155/2013/382172

[378] Eremin, V.; Li, Z.; Ilyashenko, I.: Trapping induced N(eff) and electrical field transformation at different temperatures in neutron irradiated high resistivity silicon detectors. In: *Nucl. Inst. Meth. A* 360 (1995), p. 458. doi: 10.1016/0168-9002(95)00112-3

[379] Erginsoy, C: Neutral impurity scattering in semiconductors. In: *Phys. Rev.* 79 (1950), p. 1013. doi: 10.1103/PhysRev.79.1013

[380] Eskut, E. et al. (CHORUS Collaboration): The CHORUS experiment to search for $\nu_\mu \to \nu_\tau$ oscillation. In: *Nucl. Inst. Meth. A* 401 (1997), p. 7. doi: 10.1016/S0168-9002(97)00931-5

[381] European Atomic Energy Community: CAMAC – a modular instumentation system for data handling – description and specification / Euratom, Luxemburg. Euratombericht EUR 4100 e. 1969.

[382] Evans, R.D.: *The Atomic Nucleus*. Krieger, 1982.
[383] Fabjan, C. W.: Calorimetry in high-energy physics. Proc.: *Techniques and Concepts of High-Energy Physics: 3rd NATO Advanced Study Institute,* St. Croix, USA*, 1984*. In: *NATO Sci. Ser. B* 128 (1985), p. 281. doi: 10.1007/978-1-4684-5018-7_5
[384] Fabjan, C. W. et al.: Iron liquid–argon and uranium liquid–argon calorimeters for hadron energy measurement. In: *Nucl. Inst. Meth.* 141 (1977), p. 61. doi: 10.1016/0029-554X(77)90747-9
[385] Fabjan, C.W.; Gianotti, F.: Calorimetry for Particle Physics. In: *Rev. Mod. Phys.* 75 (2003), p. 1243. doi: 10.1103/RevModPhys.75.1243
[386] Fabjan, C.W.; Ludlam, T.: Calorimetry in high-energy physics. In: *Annu. Rev. Nucl. Part. Sci.* 32 (1982), p. 335. doi: 10.1146/annurev.ns.32.120182.002003
[387] Fabjan, C.W. (ed.); Schopper, H. (ed.): *Landolt-Börnstein – Group I Elementary Particles, Nuclei and Atoms*. vol. 21B1: *Detectors for Particles and Radiation. Part 1: Principles and Methods*. Springer, 2011. doi: 10.1007/978-3-642-03606-4
[388] Fabjan, C.W.; Struczinski, W.: Coherent emission of transition radiation in periodic radiators. In: *Phys. Lett.* B 57 (1975), p. 483. doi: 10.1016/0370-2693(75)90274-9
[389] Faccio, F. et al.: Total ionizing dose effects in shallow trench isolation oxides. In: *Microelectronics Reliability* 48 (2008), p. 1000. doi: 10.1016/j.microrel.2008.04.004
[390] Fairchild Semiconductor: LVDS Fundamentals. Application note AN-5017. 2005. http://www.onsemi.com/pub/Collateral/AN-5017.pdf.pdf
[391] Fairstein, E.: Bipolar pulse shaping revisited. In: *IEEE Trans. Nucl. Sci.* 44 (1997), p. 424. doi: 10.1109/23.603684
[392] Fano, U.: Penetration of protons, alpha particles, and mesons. In: *Annu. Rev. Nucl. Part. Sci.* 13 (1963), p. 1. doi: 10.1146/annurev.ns.13.120163.000245
[393] Fanti, V. et al. (NA48 Collaboration): A new measurement of direct CP violation in two pion decays of the neutral kaon. In: *Phys. Lett.* B 465 (1999), p. 335. doi: 10.1016/S0370-2693(99)01030-8
[394] Farr, W.; Heuer, R.D.; Wagner, A.: Readout of drift chambers with a 100-MHz flash ADC system. In: *IEEE Trans. Nucl. Sci.* 30 (1983), p. 95. doi: 10.1109/TNS.1983.4332227
[395] Faxén, H.; Holtsmark, J.: Beitrag zur Theorie des Durchganges langsamer Elektronen durch Gase. In: *Z. Phys.* 45 (1927), p. 307. doi: 10.1007/BF01343053
[396] Fayard, L.: Transition radiation. Proc.: *7ème Ecole Joliot Curie 1988 – Instrumentation in Nuclear Physics and Particle Physics,* Carcans, France, (1988). http://inspirehep.net/record/268533
[397] Ferbel, T. (ed.): *Experimental Techniques in High-energy Nuclear and Particle Physics*. World Scientific, 1991 (Frontiers in Physics). doi: 10.1142/1571
[398] Fermi, E.: The ionization loss of energy in gases and in condensed materials. In: *Phys. Rev.* 57 (1940), p. 485. doi: 10.1103/PhysRev.57.485
[399] Fermi Collaboration: The Fermi Large Area Telescope. http://www-glast.stanford.edu/
[400] Fermi Laboratory: LBNF/DUNE: An international flagship neutrino experiment. Version: 2019. https://www.fnal.gov/pub/science/lbnf-dune

[401] Ferrari, A.; Sala, P.R.; Fasso, A.; Ranft, J.: FLUKA: A multi-particle transport code. CERN-2005-010. 2005. doi: 10.5170/CERN-2005-010

[402] Ferry, D.: *Semiconductor Transport*. Taylor & Francis, 2000. doi: 10.4324/9781315267548

[403] Feusels, T.: *Measurement of cosmic ray composition and energy spectrum between 1 PeV and 1 EeV with IceTop and IceCube*, Ghent University, Diss., 2013. http://hdl.handle.net/1854/LU-4337238

[404] Fidecaro, G.: The high frequency properties of a coaxial cable and the distortion of fast pulses. In: *Nuovo Cim. Suppl.* 15 (1960), p. 254. doi: 10.1007/BF02724869

[405] Field, C. et al.: Timing and detection efficiency properties of multianode PMTs for a focusing DIRC. SLAC-PUB-10231. 2003. https://www.slac.stanford.edu/pubs/slacpubs/10000/slac-pub-10231.pdf

[406] Field, C. et al.: Development of Photon Detectors for a Fast Focusing DIRC. In: *Nucl. Inst. Meth. A* 553 (2005), p. 96. doi: 10.1016/j.nima.2005.08.046

[407] Fink, J.: *Characterization of the Imaging Performance of the Simultaneously Counting and Integrating X-ray Detector CIX*, Universität Bonn, Diss., 2009. http://hss.ulb.uni-bonn.de/2010/1987/1987.htm

[408] Fink, J.; Kraft, E.; Krueger, H.; Wermes, N.; Engel, K.J.; Herrmann, C.: Comparison of pixelated CdZnTe, CdTe and Si sensors with the simultaneously counting and integrating CIX chip. In: *IEEE Trans. Nucl. Sci.* 56 (2009), p. 3819. doi: 10.1109/TNS.2009.2033111

[409] Fink, J.; Lodomez, P.; Kruger, H.; Pernegger, H.; Weilhammer, P.; Wermes, N.: TCT characterization of different semiconductor materials for particle detection. In: *Nucl. Inst. Meth. A* 565 (2006), p. 227. doi: 10.1016/j.nima.2006.05.003

[410] Finkelnburg, W.: *Einführung in die Atomphysik*. Springer, 1967. doi: 10.1007/978-3-642-64980-6

[411] Fischer, H.G.: Multiwire proportional quantameters. In: *Nucl. Inst. Meth.* 156 (1978), p. 81. doi: 10.1016/0029-554X(78)90695-X

[412] Fischer, J.; Iwata, S.; Radeka, V.; Wang, C.L.; Willis, W.J.: Lithium transition radiator and xenon detector systems for particle identification at high-energies. In: *Nucl. Inst. Meth.* 127 (1975), p. 525. doi: 10.1016/0029-554X(75)90655-2

[413] Fischer, P.: Comments on the reconstruction of hit positions in segmented detectors. 2015. Private communication and unpublished presentations at workshops and in lectures.

[414] Fischer, P.; Hausmann, J.; Overdick, M.; Raith, B.; Wermes, N. et al.: A counting pixel readout chip for imaging applications. In: *Nucl. Inst. Meth. A* 405 (1998), p. 53. doi: 10.1016/S0168-9002(97)01146-7

[415] FlexPDE software: http://www.pdesolutions.com/ .

[416] FLUKA team: official FLUKA site. Version: 2019. http://www.fluka.org

[417] Fontanelli, F. (LHCb RICH Collaboration): The pixel hybrid photon detector of the LHCb RICH. In: *Nucl. Phys. Proc. Suppl.* 197 (2009), p. 292. doi: 10.1016/j.nuclphysbps.2009.10.088

[418] Fonte, P.: Applications and new developments in resistive plate chambers. In: *IEEE Trans. Nucl. Sci.* 49 (2002), p. 881. doi: 10.1109/TNS.2002.1039583

[419] Förster, T.: Zwischenmolekulare Energiewanderung und Fluoreszenz. In:

Ann. Phys. 437 (1948), p. 55. doi: 10.1002/andp.19484370105

[420] Foster, B. et al. (ZEUS Collaboration): The design and construction of the ZEUS central tracking detector. In: *Nucl. Inst. Meth. A* 338 (1994), p. 254. doi: 10.1016/0168-9002(94)91313-7

[421] Frach, T.: Optimization of the digital silicon photomultiplier for Cherenkov light detection. In: *JINST* 7 (2012), p. C01112. doi: 10.1088/1748-0221/7/01/C01112

[422] Frach, T.; Prescher, G.; Degenhart, C.; Zwaans, B.: The digital silicon photomultiplier – a novel sensor for the detection of scintillation light. Proc.: *2009 IEEE Nuclear Science Symposium,* Orlando, FL, (2009), p. 2386. doi: 10.1109/NSSMIC.2009.5402190

[423] Frach, T.; Prescher, G.; Degenhart, C; Zwaans, B.: The digital silicon photomultiplier – principle of operation and intrinsic detector performance. Proc.: *2009 IEEE Nuclear Science Symposium,* Orlando, FL, (2009), p. 1959. doi: 10.1109/NSSMIC.2009.5402143

[424] Frank, I.M.: Transition radiation and the Cerenkov effect. In: *Sov. Phys. Usp.* 4 (1962), p. 740. http://stacks.iop.org/0038-5670/4/i=5/a=A08

[425] Frank, I.M.; Ginsburg, V.L.: Radiation of a uniformly moving electron due to its transition from one medium into another. In: *J. Phys. (USSR) (Zh. Eksp. Teor. Fiz. 16 (1946) 15)* 9 (1945), p. 353.

[426] Frank, I.M.; Tamm, I.: Coherent visible radiation of fast electrons passing through matter. In: *Dokl. Akad. Nauk SSSR* 14 (1937), p. 109. – reprinted in [231].

[427] Freeman, J. et al.: Introduction to HOBIT, a b-jet identification tagger at the CDF experiment optimized for light Higgs boson searches. In: *Nucl. Inst. Meth. A* 697 (2013), p. 64. doi: 10.1016/j.nima.2012.09.021

[428] Friedl, M. et al.: The Belle II silicon vertex detector. In: *Nucl. Inst. Meth. A* 732 (2013), p. 83. doi: 10.1016/j.nima.2013.05.171

[429] Fruhwirth, R.: Application of Kalman filtering to track and vertex fitting. In: *Nucl. Inst. Meth. A* 262 (1987), p. 444. doi: 10.1016/0168-9002(87)90887-4

[430] Fujii, K. et al.: Automated monitoring and calibrating system of gas gain and electron drift velocity for:Prototype system and accumulation of reference data. In: *Nucl. Inst. Meth. A* 245 (1986), p. 35. doi: 10.1016/0168-9002(86)90255-X

[431] Fukuda, Y. et al. (Super-Kamiokande Collaboration): Evidence for oscillation of atmospheric neutrinos. In: *Phys. Rev. Lett.* 81 (1998), p. 1562. doi: 10.1103/PhysRevLett.81.1562

[432] Fukuda, Y. et al. (Super-Kamiokande Collaboration): The Super-Kamiokande detector. In: *Nucl. Inst. Meth. A* 501 (2003), p. 418. doi: 10.1016/S0168-9002(03)00425-X

[433] Fulbright, H.W.: Ionization chambers in nuclear physics. In: *[330]*, p. 1. doi: 10.1007/978-3-642-45903-0_1

[434] Fülöp, L.; Biro, T.: Cherenkov radiation spectrum. In: *Int. J. Theor. Phys.* 31 (1992), p. 61. doi: 10.1007/BF00674341

[435] Furletova, J.; Furletov, S.: New transition radiation detection technique based on DEPFET silicon pixel matrices. In: *Nucl. Inst. Meth. A* 628 (2011), p. 309. doi: 10.1016/j.nima.2010.06.342

[436] Gabriel, T.A.; Groom, D.E.; Job, P.K.; Mokhov, N.V.; Stevenson, G.R.: Energy dependence of hadronic activity. In: *Nucl. Inst. Meth. A* 338 (1994),

p. 336. doi: 10.1016/0168-9002(94)91317-X
[437] Gaillard, J.M. (UA2 Collaboration): The UA2 scintillating fibre detector. In: *Nucl. Phys. Proc. Suppl.* 16 (1990), p. 509. doi: 10.1016/0920-5632(90)90457-6
[438] Gaisser, T.K.; Engel, R.; Resconi, E.: *Cosmic Rays and Particle Physics*. 2nd ed. Cambridge University Press, 2016. doi: 10.1017/CBO9781139192194
[439] Gaisser, T.K.; Hillas, A.M.: Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers. Proc.: *International Cosmic Ray Conference (ICRC1977)*, vol. 8 (1977), p. 353. http://adsabs.harvard.edu/abs/1977ICRC....8..353G
[440] Gandhi, R.; Quigg, C.; Reno, M.H.; Sarcevic, I.: Ultrahigh-energy neutrino interactions. In: *Astropart.Phys.* 5 (1996), p. 81. doi: 10.1016/0927-6505(96)00008-4
[441] Garabatos, C.: The ALICE TPC. In: *Nucl. Inst. Meth. A* 535 (2004), p. 197. doi: 10.1016/j.nima.2004.07.127
[442] Garcia-Sciveres, M.; Wermes, N.: A review of advances in pixel detectors for experiments with high rate and radiation. In: *Rep. Prog. Phys.* 81 (2018), p. 066101. doi: 10.1088/1361-6633/aab064
[443] Garibian, G.M.: Contribution to the theory of transition radiation. In: *Sov. Phys. JETP* 6 (1958), p. 1079. http://www.jetp.ac.ru/cgi-bin/dn/e_006_06_1079.pdf
[444] Garibyan, G.M.: Transition radiation under inclined incidence of the charge. In: *Sov. Phys. JETP* 11 (1960), p. 1306. http://www.jetp.ac.ru/cgi-bin/dn/e_011_06_1306.pdf
[445] Garwin, R.L.: The design of liquid scintillation cells. In: *Rev. Sci. Inst.* 23 (1952), p. 755. doi: 10.1063/1.1746152
[446] Gas detectors development group at CERN: MSGC webpage. https://gdd.web.cern.ch/GDD/msgc.htmL
[447] Gatti, E.; Geraci, A.: Considerations about Ramo's theorem extension to conductor media with variable dielectric constant. In: *Nucl. Inst. Meth. A* 525 (2004), p. 623. doi: 10.1016/j.nima.2004.02.002
[448] Gatti, E.; Manfredi, P.F.: Processing the signals from solid state detectors in elementary particle physics. In: *Riv. Nuovo Cim.* 9N1 (1986), p. 1. doi: 10.1007/BF02822156
[449] Gatti, E.; Padovini, G.; Radeka, V.: Signal evaluation in multielectrode radiation detectors by means of a time dependent weighting vector. In: *Nucl. Inst. Meth.* 193 (1982), p. 651. doi: 10.1016/0029-554X(82)90265-8
[450] Gatti, E.; Rehak, P.: Semiconductor drift chamber – an application of a novel charge transport scheme. In: *Nucl. Inst. Meth. A* 225 (1984), p. 608.
[451] Gavrila, M.: Relativistic K-shell photoeffect. In: *Phys. Rev.* 113 (1959), p. 514. doi: 10.1103/PhysRev.113.514
[452] Geant4 Collaboration: Physics Reference Manual. http://www.geant4.org/geant4/support/index.shtml
[453] Geiger, H.; Marsden, E.: LXI. The laws of deflexion of a particles through large angles. In: *Phil. Mag.* 25 (1913), p. 604. doi: 10.1080/14786440408634197
[454] Geiger, H.; Müller, W.: Elektronenzählrohr zur Messung schwächster Aktivitäten. In: *Naturwissenschaften* 16 (1928), p. 617. doi: 10.1007/BF01494093
[455] Gessler, P.: Next generation electronics based on μTCA for beam-diagnostics at FLASH and XFEL. Proc.: *10th European Workshop on Beam Diagnostics*

and Instrumentation for Particle Accelerators, Hamburg, (2011), p. 294. http://adweb.desy.de/mpy/DIPAC2011/papers/tuob03.pdf

[456] Giesch, M. et al.: Status of magnetic horn and neutrino beam. In: *Nucl. Inst. Meth.* 20 (1963), p. 58. doi: 10.1016/0029-554X(63)90391-4

[457] Ginsburg, D.; Ginzburg, V.L.; Glebov, O.: *Applications of Electrodynamics in Theoretical Physics and Astrophysics.* CRC Press, 1989.

[458] Ginzburg, V.L.: Transition radiation and transition scattering. Proc.: *16th International Cosmic Ray Conference,* Kyoto, vol. 14 (1979), p. 42. https://ui.adsabs.harvard.edu/link_gateway/1979ICRC...14...42G/ADS_PDF

[459] Ginzburg, V.L.: Transition radiation and transition scattering. In: *Physica Scripta* T2A, 1982 (1982), p. 182. doi: 10.1088/0031-8949/1982/t2a/024

[460] Giomataris, Y.; Rebourgeard, P.; Robert, J.P.; Charpak, G.: MICROMEGAS: a high granularity position sensitive gaseous detector for high particle flux environments. In: *Nucl. Inst. Meth. A* 376 (1996), p. 29. doi: 10.1016/0168-9002(96)00175-1

[461] Giunti, C.; Kim, C.W.: *Fundamentals of Neutrino Physics and Astrophysics.* Oxford University Press, 2007. doi: 10.1093/acprof:oso/9780198508717.003.0001

[462] Glaser, D.A.: Nobel Lecture: Elementary Particles and Bubble Chambers. http://www.nobelprize.org/nobel_prizes/physics/laureates/1960/glaser-lecture.html

[463] Glaser, D.A.; Rahm, D.C.: Characteristics of bubble chambers. In: *Phys. Rev.* 97 (1955), p. 474. doi: 10.1103/PhysRev.97.474

[464] Glassel, P.: The limits of the ring image Cherenkov technique. In: *Nucl. Inst. Meth. A* 433 (1999), p. 17. doi: 10.1016/S0168-9002(99)00315-0

[465] Gluckstern, R.L.: Uncertainties in track momentum and direction, due to multiple scattering and measurement errors. In: *Nucl. Inst. Meth.* 24 (1963), p. 381. doi: 10.1016/0029-554X(63)90347-1

[466] Goldsmith, P.; Jelley, J.V.: Optical transition radiation from protons entering metal surfaces. In: *Phil. Mag.* 4 (1959), p. 836. doi: 10.1080/14786435908238241

[467] Goldstein, H.; Poole, C.P.; Safko, J.L.: *Classical Mechanics.* Addison Wesley, 2002.

[468] Gordon, H. et al. (Brookhaven-CERN-Copenhagen-Lund-Rutherford-Tel Aviv Collaboration): The axial field spectrometer at the CERN ISR. In: *Nucl. Inst. Meth.* 196 (1982), p. 303. doi: 10.1016/0029-554X(82)90660-7

[469] Gorelov, I. et al.: Electrical characteristics of silicon pixel detectors. In: *Nucl. Inst. Meth. A* 489 (2002), p. 202. doi: 10.1016/S0168-9002(02)00557-0

[470] Graaf, H. van der: New developments in gaseous tracking and imaging detectors. In: *Nucl. Inst. Meth. A* 607 (2009), p. 78. doi: 10.1016/j.nima.2009.03.137

[471] Graaf, H. van der: Single electron sensitive GridPix TPCs and their application in Dark Matter search and ν-less double beta decay experiments. In: *J. Phys. Conf. Ser.* 309 (2011), p. 012016. doi: 10.1088/1742-6596/309/1/012016

[472] Gray, P.R.; Hurst, P.J.; Lewis, S.R.; Meyer, R.G.: *Analog Integrated Circuits.* 5th ed. John Wiley and Sons, 2009.

[473] Green, D.: *The Physics of Particle Detectors.* Cambridge University Press, 2005.

[474] Green, M. A.: Intrinsic concentration, effective density of states, and effective mass in silicon. In: *J. Appl. Phys.* 67 (1990), p. 2944. doi: 10.1063/1.345414
[475] Greisen, K.: Cosmic ray showers. In: *Annu. Rev. Nucl. Part. Sci.* 10 (1960), p. 63. doi: 10.1146/annurev.ns.10.120160.000431
[476] Greisen, K.: End to the cosmic ray spectrum? In: *Phys. Rev. Lett.* 16 (1966), p. 748. doi: 10.1103/PhysRevLett.16.748
[477] Grimm, O.: *Driftgeschwindigkeits- und Signalverstärkungsmessungen in Gasen für das äußere Spurkammersystem des HERA-B Detektors*, Universität Hamburg, diploma thesis, 1998. DESY-THESIS-1998-023. http://www-library.desy.de/preparch/desy/thesis/desy-thesis-98-023.pdf
[478] Groom, D.E.: Silicon photodiode detection of bismuth germanate scintillation light. In: *Nucl. Inst. Meth. A* 219 (1984), p. 141. doi: 10.1016/0167-5087(84)90146-7
[479] Groom, D.E.: Energy loss in matter by heavy particles. PDG-93-06. 1993. http://pdg.lbl.gov/rpp/encoders/pdg_note_9306.pdf
[480] Groom, D.E.: Explanation of some entries in 'Atomic and Nuclear Properties of Materials'. Version: 2018. http://pdg.lbl.gov/2018/AtomicNuclearProperties/explain_elem.html
[481] Groom, D.E.: Gas densities and optical properties. Version: 2018. http://pdg.lbl.gov/2018/AtomicNuclearProperties/explain_gas.html
[482] Groom, D.E.; Mokhov, N.V.; Striganov, S.I.: Muon stopping power and range tables 10-MeV to 100-TeV. In: *Atom. Data Nucl. Data Tabl.* 78 (2001), p. 183. doi: 10.1006/adnd.2001.0861
[483] Groom, D.E.: Temperature Dependence of Mean Number of e-h Pairs per eV of X-ray Energy Deposit. Version: 2004. http://www-ccd.lbl.gov/w_Si.pdf
[484] Grosse-Knetter, J.: *Vertex Measurement at a Hadron Collider – The ATLAS Pixel Detector*, Universität Bonn, habilitation thesis, 2008. BONN-IR-2008-04. http://hep1.physik.uni-bonn.de/fileadmin/Publications/ATLAS_Analysis/habil_grosse-knetter.pdf
[485] Grote, H.: Review of pattern recognition in high-energy physics. In: *Rep. Prog. Phys.* 50 (1987), p. 473. doi: 10.1088/0034-4885/50/4/002
[486] Grove, A.S.: *Physics and Technology of Semiconductor Devices*. John Wiley and Sons, 1967.
[487] Grupen, C.: *Astroparticle Physics*. Springer, 2005. doi: 10.1007/3-540-27670-X
[488] Grupen, C. (ed.); Buvat, I. (ed.): *Handbook of Particle Detection and Imaging, Vol. 1 and Vol. 2*. Springer, 2012. doi: 10.1007/978-3-642-13271-1
[489] Grupen, C.; Shwartz, B.: *Particle Detectors*. 2nd ed. Cambridge University Press, 2008. doi: 10.1017/CBO9780511534966
[490] GSI Darmstadt: Heavy Ion Therapy. https://www.gsi.de/en/researchaccelerators/research_an_overview/ion_beam_radiotherapy_in_the_fight_against_cancer.htm
[491] Gutierrez, E.A.; Deen, M.J.; C., Claeys: *Low Tempereture Electronics: Physics, Devices, Circuits and Applications*. Academic Press, 2001.
[492] Haidt, D.: Discovery of weak neutral currents in Gargamelle. In: *AIP Conf. Proc.* 300 (1994), p. 187. doi: 10.1063/1.45429
[493] Hall, R. N.: Electron–hole recombination in germanium. In: *Phys. Rev.* 87 (1952), p. 387. doi: 10.1103/PhysRev.87.387
[494] Hallgren, B.; Verweij, H.: New developments in time and pulse

height digitizers. In: *IEEE Trans. Nucl. Sci.* 27 (1980), p. 333. doi: 10.1109/TNS.1980.4330850

[495] HAMAMATSU: Photonics Handbook: Detectors. https://www.photonics.com/EDU/Handbook.aspx?AID=25535

[496] HAMAMATSU: Photonics Online. http://www.photonicsonline.com/doc/photomulitiplier-tube-series-uba-sba-0002

[497] HAMAMATSU PHOTONICS: *Photomultiplier Tubes – Basics and Appications*. ed. 3a. 2007. https://www.hamamatsu.com/resources/pdf/etd/PMT_handbook_v4E.pdf

[498] HAMANN, E.: *Characterization of High Resistivity GaAs as Sensor Material for Photon Counting Semiconductor Pixel Detectors*, Universität Freiburg, Dissertation, 2013. https://nbn-resolving.org/urn:nbn:de:bsz:25-opus-93550

[499] HAMANN, M.: *Studies for a Linear Collider Drift Chamber and Search for Heavy Stable Charged Particles in e^+e^- Collisions up to $\sqrt{s} = 209$ GeV*, Universität Hamburg, Diss., 2003. doi: 10.3204/DESY-THESIS-2003-046

[500] HAMEL, L.-A.; JULIEN, M.: Generalized demonstration of Ramo's theorem with space charge and polarization effects. In: *Nucl. Inst. Meth. A* 597 (2008), p. 207. doi: 10.1016/j.nima.2008.09.008

[501] HAMPEL, W. et al. (GALLEX Collaboration): GALLEX solar neutrino observations: results for GALLEX IV. In: *Phys. Lett. B* 447 (1999), p. 127. doi: 10.1016/S0370-2693(98)01579-2

[502] HAMS, T. et al.: Measurement of the abundance of radioactive Be-10 and other light isotopes in cosmic radiation up to 2-GeV/nucleon with the balloon-borne instrument ISOMAX. In: *Astropphys. J.* 611 (2004), p. 892. doi: 10.1086/422384

[503] HARIGEL, G.: Die Große Europäische Blasenkammer im CERN (Teil I). In: *Physikalische Blätter* 31 (1975), p. 13. doi: 10.1002/phbl.19750310105

[504] HARIGEL, G. G. (ed.); COLLEY, C. D. (ed.); CUNDY, C. D. (ed.): *Proceedings of the Conference on the Bubble Chamber and its Contributions to Particle Physics (Marking the 40th Anniversary of the Bubble Chamber)*, Geneva, 1993. In: *Nucl. Phys. Proc. Suppl.* 36, 1994. https://www.sciencedirect.com/science/journal/09205632/36/supp/C

[505] HARKONEN, J. et al.: Radiation hardness of Czochralski silicon, float zone silicon and oxygenated float zone silicon studied by low energy protons. In: *Nucl. Inst. Meth. A* 518 (2004), p. 346. doi: 10.1016/j.nima.2003.11.018

[506] HARTMANN, F.: Evolution of silicon sensor technology in particle physics. In: *Springer Tracts Mod. Phys.* 275 (2017). doi: 10.1007/978-3-319-64436-3

[507] HAUNGS, A.: KASCADE-Grande: Luftschauer über Karlsruhe. Version: 2007. http://www.weltderphysik.de/gebiet/astro/kosmische-strahlung/detektoren/kascade-grande

[508] HAUSCHILD, M.: Progress in dE/dx techniques used for particle identification. In: *Nucl. Inst. Meth. A* 379 (1996), p. 436. doi: 10.1016/0168-9002(96)00607-9

[509] HAUSCHILD, M. et al.: Particle identification with the OPAL jet chamber. In: *Nucl. Inst. Meth. A* 314 (1992), p. 74. doi: 10.1016/0168-9002(92)90501-T

[510] HAVRÁNEK, M. et al.: DMAPS: a fully depleted monolithic active pixel sensor – analog performance characterization. In: *arXiv:1407.0641* (2014).

[511] HAYNES, W.M. (ed.); LIDE, D.R. (ed.); BRUNO, T.J. (ed.): *CRC Handbook of Chemistry and Physics 2015-2016*. 96th edition. CRC Press, 2015. https://hbcp.chemnetbase.com/

[512] He, Z.: Review of the Shockley–Ramo theorem and its application in semiconductor gamma-ray detectors. In: *Nucl. Inst. Meth. A* 463 (2001), p. 250. doi: 10.1016/S0168-9002(01)00223-6

[513] Hebbeker, T.; Hoepfner, K.: Muon spectrometers. In: *[488]*, p. 473. doi: 10.1007/978-3-642-13271-1_19

[514] Hecht, K.: Zum Mechanismus des lichtelektrischen Primärstromes in isolierenden Kristallen. In: *Z. Phys.* 77 (1932), p. 235.

[515] Heck, D. et al.: CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers. Report FZKA 6019, Forschungszentrum Karlsruhe. 1998. https://web.ikp.kit.edu/corsika

[516] Heck, D.; Pierog, T.: Extensive Air Shower Simulation with CORSIKA: A User's Guide. https://web.ikp.kit.edu/corsika/usersguide/usersguide.pdf

[517] Heitler, W.: *The Quantum Theory of Radiation.* 3rd ed. Clarendon Press, 1954.

[518] Hellmig, J. et al.: The CDMS II Z-sensitive ionization and phonon germanium detector. In: *Nucl. Inst. Meth. A* 444 (2000), p. 308. doi: 10.1016/S0168-9002(99)01403-5

[519] Hemperek, T.; Kishishita, T.; Krüger, H.; Wermes, N.: A monolithic active pixel sensor for ionizing radiation using a 180 nm HV-SOI process. In: *Nucl. Inst. Meth. A* 796 (2015), p. 8. doi: 10.1016/j.nima.2015.02.052

[520] Hess, V.F.: Nobel Lecture: Unsolved Problems in Physics: Tasks for the Immediate Future in Cosmic Ray Studies. http://www.nobelprize.org/nobel_prizes/physics/laureates/1936/hess-lecture.html

[521] H.E.S.S. Collaboration: H.E.S.S. High Energy Stereoscopic System. http://www.mpi-hd.mpg.de/hfm/HESS

[522] Highland, V.L.: Some practical remarks on multiple scattering. In: *Nucl. Inst. Meth.* 129 (1975), p. 497. doi: 10.1016/0029-554X(75)90743-0

[523] Hilke, H.J.: Time projection chambers. In: *Rep. Prog. Phys.* 73 (2010), p. 116201. doi: 10.1088/0034-4885/73/11/116201

[524] Hillas, A.M.: Cerenkov light images of EAS produced by primary gamma. Proc.: *International Cosmic Ray Conference (ICRC 1985)*, vol. 3 (1985), p. 445. http://adsabs.harvard.edu/abs/1985ICRC....3..445H

[525] Hirao, Y. et al.: Heavy ion synchrotron for medical use – HIMAC project at NIRS-Japan. In: *Nucl. Phys. A* 538 (1992), p. 541. doi: 10.1016/0375-9474(92)90803-R

[526] Hirata, K. et al. (KAMIOKANDE-II Collaboration): Observation of a neutrino burst from the supernova SN 1987a. In: *Phys. Rev. Lett.* 58 (1987), p. 1490. doi: 10.1103/PhysRevLett.58.1490

[527] Hirono, T. et al.: Depleted fully monolithic active CMOS pixel sensors (DMAPS) in high resistivity 150 nm technology for LHC. In: *Nucl. Inst. Meth. A* 924 (2019), p. 87. doi: 10.1016/j.nima.2018.10.059

[528] HIT: Heavy Ion Therapy Center, Heidelberg. https://www.klinikum.uni-heidelberg.de/Heidelberger-Ionenstrahl-Therapie-HIT.112189.0.html

[529] Hoddeson, L. (ed.); Brown, L. (ed.); Riordan, M. (ed.); Dresden, M. (ed.): *The Rise of the Standard Model – A History of Particle Physics from 1964 to 1979.* Cambridge University Press, 1997. doi: 10.1017/CBO9780511471094

[530] Hofstadter, R. et al.: The Crystal Ball Experiment. Proc.: *Particle Physics in GeV Region,* Tokyo, Japan. Univ. Tokyo, (1979), p. 559.

[531] Hohlmann, M. (ed.); Padilla, C. (ed.); Tesch, N. (ed.); Titov, M. (ed.): *Proc. International Workshop on Aging Phenomena in Gaseous Detectors 2001,* Hamburg, Germany. In: *Nucl. Inst. and Meth. A* 515, 2003. p. 1. doi: 10.1016/j.nima.2003.08.120

[532] Holder, M. et al.: A detector for high-energy neutrino interactions. In: *Nucl. Inst. Meth.* 148 (1978), p. 235. doi: 10.1016/0029-554X(70)90173-4

[533] Holroyd, R.A.; Anderson, D.F.: The physics and chemistry of room-temperature liquid-filled ionization chambers. In: *Nucl. Inst. Meth. A* 236 (1985), p. 294. doi: 10.1016/0168-9002(85)90164-0

[534] Holst, G.C.: *CCD Arrays, Cameras and Displays.* JCD Publishing, 1998. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.1243&rep=rep1&type=pdf

[535] Holstein, T.: Energy distribution of electrons in high frequency gas discharges. In: *Phys. Rev.* 70 (1946), p. 367. doi: 10.1103/PhysRev.70.367

[536] Honsberg, C.; Bowden, S.: Optical Properties of Silicon. http://www.pveducation.org/pvcdrom/materials/optical-properties-of-silicon

[537] Hook, J.R.; Hall, H.E.: *Solid State Physics.* John Wiley & Sons, 1995.

[538] Hoppe, M.: *Differential energy analysis of electron beam: A study of photoemission from NEA-GaAs*, Universität Heidelberg, Diss., 2001. doi: 10.11588/heidok.00001452

[539] Höppner, C.: *The Composition of Cosmic Rays at High Energies*, Technische Universität München, diploma thesis, 2006.

[540] Horowitz, P.; Hill, W.: *The Art of Electronics.* Cambridge University Press, 2001.

[541] Hosaka, J. et al. (Super-Kamiokande Collaboration): Solar neutrino measurements in super-Kamiokande-I. In: *Phys. Rev. D* 73 (2006), p. 112001. doi: 10.1103/PhysRevD.73.112001

[542] Hott, T.: Aging problems of the inner tracker of HERA-B: an example for new detectors and new effects. In: *Nucl. Inst. Meth. A* 515 (2003), p. 242. doi: 10.1016/j.nima.2003.09.005

[543] Hough, P.V.C.: The angular distribution of pair-produced electrons and bremsstrahlung. In: *Phys. Rev.* 74 (1948), p. 80. doi: 10.1103/PhysRev.74.80

[544] Huang, K.: *Statistical Mechanics.* Wiley, 1987.

[545] Hubbell, J.H.; Seltzer, S.M.: NIST Data Base. http://physics.nist.gov/PhysRefData/XrayMassCoef/cover.html

[546] Huege, T.; Falcke, H.: Radio emission from cosmic ray air showers: coherent geosynchrotron radiation. In: *Astron. Astrophys.* 412 (2003), p. 19. doi: 10.1051/0004-6361:20031422

[547] Huffman, R.E.; Larrabee, J.C.; Tanaka, Y.: Rare gas continuum light sources for photoelectric scanning in the vacuum ultraviolet. In: *Appl. Opt.* 4 (1965), p. 1581. doi: 10.1364/AO.4.001581

[548] Huhtinen, M.: Simulation of non-ionising energy loss and defect formation in silicon. In: *Nucl. Inst. Meth. A* 491 (2002), p. 194. doi: 10.1016/S0168-9002(02)01227-5

[549] Huq, S.B. and Goldie, J.: An Overview of LVDS Technology. National Semiconductor Application Note 971. 1998. http://www.ti.com/lit/an/snla165/snla165.pdf

[550] Huxley, L.G.H.; Crompton, R.W.: *The Diffusion and Drift of Electrons in*

Gases. Wiley, 1974 (Wiley series in plasma physics).

[551] Iarocci, E.: Plastic streamer tubes and their applications in high energy physics. In: *Nucl. Inst. Meth.* 217 (1983), p. 30. doi: 10.1016/0167-5087(83)90107-2

[552] IEEE Standard: 1014-1987 – IEEE Standard for a Versatile Backplane Bus: VMEbus. Version: 1987. https://standards.ieee.org/findstds/standard/1014-1987.html

[553] IEEE Standard: IEEE 802.3 – IEEE Standard for Ethernet. Version: 2012. http://standards.ieee.org/about/get/802/802.3.html

[554] Iijima, T. et al.: A novel type of proximity focusing RICH counter with multiple refractive index Aerogel radiator. In: *Nucl. Inst. Meth. A* 548 (2005), p. 383. doi: 10.1016/j.nima.2005.05.030

[555] ILC Collaboration: International Linear Collider (ILC). http://www.linearcollider.org

[556] Ilford Photo: Fact Sheet Ilford Nuclear Emulsions. Version: 2011. http://www.ilfordphoto.com/products/page.asp?n=136

[557] Ing, H.; Noulty, R.A.; McLean, T.D.: Bubble detectors – a maturing technology. In: *Radiat. Meas.* 27 (1997), p. 1. doi: 10.1016/S1350-4487(96)00156-4

[558] Insepov, Z.: PSEC – Large-Area Picosecond Photo-Detectors Project. http://psec.uchicago.edu/library/photocathodes/

[559] International commission on radiation units and measurements (ICRU): Stopping Powers for Electrons and Positrons. ICRU Report 37. 1984. – Tables and graphical presentations are interactively available at http://physics.nist.gov/PhysRefData/Star/Text/contents.html.

[560] Ioffe Data Base: New Semiconductor Materials. Characteristics and Properties. http://www.ioffe.ru/SVA

[561] Irwin, K.D. et al.: A quasiparticle-trap-assisted transition-edge sensor for phonon-mediated particle detection. In: *Rev. Sci. Inst.* 66 (1995), p. 5322. doi: 10.1063/1.1146105

[562] Isberg, J. et al.: High carrier mobility in single-crystal plasma-deposited diamond. In: *Science* 297 (2002), p. 1670. doi: 10.1126/science.1074374

[563] Isberg, J.; Lindblom, A.; Tajani, A.; Twitchen, D.: Temperature dependence of hole drift mobility in high-purity single-crystal CVD diamond. In: *Phys. Stat. Sol. (a)* 202 (2005), p. 2194. doi: 10.1002/pssa.200561915

[564] Jackson, J.D.: *Classical Electrodynamics*. 3rd ed. Wiley, 1998.

[565] Jacoboni, C.; Canali, C.; Ottaviani, G.; Alberigi Quaranta, A.: A review of some charge transport properties of Silicon. In: *Solid State Electron.* 20 (1977), p. 77. doi: 10.1016/0038-1101(77)90054-5

[566] James, F.: *Statistical Methods in Experimental Physics*. Hackensack World Scientific, 2006.

[567] Jansen, H.: *CVD Diamond: Charge Carrier Movement at Low Temperatures and use in critical timing applications*, Universität Bonn, Dissertation, 2013. BONN-IR-2013-13. http://hep1.physik.uni-bonn.de/fileadmin/Publications/ATLAS_Pixels/Diss/jansenhendrik-klein.pdf

[568] Jansons, J.L.; Krumins, V.J.; Rachko, Z.A.; Valbis, J.A.: Luminescence due to radiative transitions between valence band and upper core band in ionic-crystals (crossluminescence). In: *Phys. Stat. Sol. (b)* 144 (1987), p. 835. doi: 10.1002/pssb.2221440244

[569] Jansweijer, P.P.M.; Peek, H.Z.; De Wolf, E.: White Rabbit: sub-nanosecond timing over Ethernet. In: *Nucl. Inst. Meth. A* 725 (2013), p. 187. doi: 10.1016/j.nima.2012.12.096

[570] Jean-Marie, B.; Lepeltier, V.; L'Hote, D.: Systematic measurement of electron drift velocity and study of some properties of four gas mixtures: A–CH_4, A–C_2H_4, A–C_2H_6, A–C_3H_8. In: *Nucl. Inst. Meth.* 159 (1979), p. 213. doi: 10.1016/0029-554X(79)90349-5

[571] Jelley, J.V.: *Cherenkov Radiation and its Applications.* Pergamon Press, 1958.

[572] Jen, C.K.: On the induced current and energy balance in electronics. In: *Proc. I.R.E.* (1941), p. 345. doi: 10.1109/JRPROC.1941.230316

[573] Jenkins, D.A.: Large angle bremsstrahlung. In: *J. Phys. G* 17 (1991), p. 155. doi: 10.1088/0954-3899/17/2/009

[574] Jessop, C.: Photograph of the BaBar electromagnetic calorimeter during installation. Version: retrieved in 2015. https://www3.nd.edu/~cjessop/research/images/jb.jpg

[575] Jiang, X. et al.: Coalescence and overgrowth of diamond grains for improved heteroepitaxy on silicon(001). In: *J. Appl. Phys.* 83 (1998), p. 2511. doi: 10.1063/1.367012

[576] Johnson, J.B.: The Schottky effect in low frequency circuits. In: *Phys. Rev.* 26 (1925), p. 71. doi: 10.1103/PhysRev.26.71

[577] Johnson, J.B.: Thermal agitation of electricity in conductors. In: *Phys. Rev.* 32 (1928), p. 97. doi: 10.1103/PhysRev.32.97

[578] Joram, C.; Haefeli, G.; Leverington, B.: Scintillating fibre tracking at high luminosity colliders. In: *JINST* 10 (2015), p. C08005. doi: 10.1088/1748-0221/10/08/C08005

[579] Junkes, A.: *Influence of Radiation Induced Defect Clusters on Silicon Particle Detectors*, Universität Hamburg, Diss., 2011. DESY-THESIS-2011-031. http://www-library.desy.de/cgi-bin/showprep.pl?desy-thesis-11-031

[580] Junkes, A.; Donegani, E.M.; Neubüser, C.: Energy dependence of proton radiation damage in Si-sensors. Proc.: *21st Symposium on Room-Temperature Semiconductor X-ray and Gamma-ray Detectors (RTSD 2014),* Seattle, W, (2016), p. 7431260. doi: 10.1109/NSSMIC.2014.7431260

[581] Kadyk, J.A. (ed.): *Workshop on Radiation Damage to Wire Chambers.* Lawrence Berkeley Nat. Lab. LBL-21170, 1986. http://www.escholarship.org/uc/item/0zx777vz

[582] Kagan, H.: Recent advances in diamond detector development. In: *Nucl. Inst. Meth. A* 541 (2005), p. 221. doi: 10.1016/j.nima.2005.01.060

[583] Kajita, T.: Atmospheric neutrinos. In: *New J. Phys.* 6 (2004), p. 194. doi: 10.1088/1367-2630/6/1/194

[584] Kamata, K.; Nishimura, J.: The lateral and angular structure functions of electron showers. In: *Prog. Theor. Phys. Suppl.* 6 (1958), p. 93. doi: 10.1143/PTPS.6.93

[585] Kampert, K.-H.; Unger, M.: Measurements of the cosmic ray composition with air shower experiments. In: *Astropart. Phys.* 35 (2012), p. 660. doi: 10.1016/j.astropartphys.2012.02.004

[586] Kanaya, N. et al.: Test results on hybrid photodiodes. In: *Nucl. Inst. Meth. A* 421 (1999), p. 512. doi: 10.1016/S0168-9002(98)01256-X

[587] Kaplon, M.; Peters, B.; Ritson, D. M.: Emulsion cloud-chamber study of a high energy interaction in the cosmic radiation. In: *Phys. Rev.* 85 (1952), Mar, p. 900. doi: 10.1103/PhysRev.85.900

[588] Kase, M.; Akioka, T.; Mamyoda, H.; Kikuchi, J.; Doke, T.: Fano factor in pure argon. In: *Nucl. Inst. Meth.* 227 (1984), p. 311. doi: 10.1016/0168-9002(84)90139-6

[589] Kastli, H.C. et al.: CMS barrel pixel detector overview. In: *Nucl. Inst. Meth. A* 582 (2007), p. 724. doi: 10.1016/j.nima.2007.07.058

[590] Katsura, T. et al.: Energy resolution of a multiwire proportional quantameter. In: *Nucl. Inst. Meth.* 105 (1972), p. 245. doi: 10.1016/0029-554X(72)90565-4

[591] Katz, U.F.; Spiering, C.: High-energy neutrino astrophysics: status and perspectives. In: *Prog. Part. Nucl. Phys.* 67 (2012), p. 651. doi: 10.1016/j.ppnp.2011.12.001

[592] Kawrakow, I.; Rogers, D.W.O.: The EGSnrc code system: Monte Carlo simulation of electron and photon transport. NRCC Report PIRS-701. 2006. https://nrc-cnrc.github.io/EGSnrc/

[593] Keil, G.: Design principles of fluorescence radiation converters. In: *Nucl. Inst. Meth.* 89 (1970), p. 111. doi: 10.1016/0029-554X(70)90813-X

[594] Kemmer, J. et al.: Experimental confirmation of a new semiconductor detector principle. Proc.: *Fifth European Symposium on Semiconductors Detectors*. In: *Nucl. Inst. Meth. A* 288 (1990), p. 92. doi: 10.1016/0168-9002(90)90470-Q

[595] Kemmer, J.; Belau, E.; Prechtel, U.; Welser, W.; Lutz, G.: Low capacity drift diode. In: *Nucl. Inst. Meth. A* 253 (1987), p. 378. doi: 10.1016/0168-9002(87)90519-5

[596] Kempainen, S.: Low-Voltage Differential Signaling (LVDS). National Semiconductor Application Note 1382-6. 2002. http://literature.cdn.keysight.com/litweb/pdf/5988-4797EN.pdf

[597] Kenney, C.; Parker, S.; Segal, J.; Storment, C.: Silicon detectors with 3-D electrode arrays: fabrication and initial test results. In: *IEEE Trans. Nucl. Sci.* 48 (1999), p. 1224. doi: 10.1109/23.785737

[598] Kenney, C.J.; Parker, S.I.; Peterson, V.Z.; Snoeys, W.J.; Plummer, J.D.; Aw, C. H.: A prototype monolithic pixel detector. In: *Nucl. Inst. Meth. A* 342 (1994), p. 59. doi: 10.1016/0168-9002(94)91411-7

[599] Kester, W.: *Data Conversion Handbook*. Elsevier, 2005.

[600] Ketek GmbH: SiPM Technology. Version: 2015. http://www.ketek.net/products/sipm-technology/

[601] Khachatryan, V. et al. (CMS Collaboration): Strange particle production in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV. In: *JHEP* 1105 (2011), p. 064. doi: 10.1007/JHEP05(2011)064

[602] Kharzheev, Yu.N.: Use of silica aerogels in Cherenkov counters. In: *Phys. Part. Nucl.* 39 (2008), p. 107. doi: 10.1007/s11496-008-1008-3

[603] Kharzheev, Yu.N.: Scintillation counters in modern high-energy physics experiments (review). In: *Phys. Part. Nucl.* 46 (2015), p. 678. doi: 10.1134/S1063779615040048

[604] Kharzheev, Yu.N.: Radiation hardness of scintillation detectors based on organic plastic scintillators and optical fibers. In: *Phys. Part. Nucl.* 50 (2019), p. 42. doi: 10.1134/S1063779619010027

[605] Kimura, M. et al.: Development of nuclear emulsions with 1 μm spatial res-

olution for the AEgIS experiment. In: *Nucl. Inst. Meth. A* 732 (2013), p. 325. doi: 10.1016/j.nima.2013.04.082

[606] Kirsten, T.A.: Solar neutrino experiments: results and implications. In: *Rev. Mod. Phys.* 71 (1999), p. 1213. doi: 10.1103/RevModPhys.71.1213

[607] Kishishita, T.; Krüger, H.; Hemperek, T.; Lemarenko, M.; Koch, M.; Gronewald, M.; Wermes, N.: Prototype of a gigabit data transmitter in 65 nm CMOS for DEPFET pixel detectors at Belle-II. In: *Nucl. Inst. Meth. A* 718 (2013), p. 168. doi: 10.1016/j.nima.2012.11.013

[608] Klages, C.P.: Chemical vapour deposition of diamond. In: *Appl. Phys. A* 56 (1993), p. 513. doi: 10.1007/BF00331401

[609] Klein, O.; Nishina, Y.: Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. In: *Z. Phys.* 52 (1929), p. 853. doi: 10.1007/BF01366453

[610] Klein, S.: Suppression of bremsstrahlung and pair production due to environmental factors. In: *Rev. Mod. Phys.* 71 (1999), p. 1501. doi: 10.1103/RevModPhys.71.1501

[611] Klein, S.: The time projection chamber turns 25. In: *CERN Cour.* 44N1 (2004), p. 40. http://cerncourier.com/cws/article/cern/29014

[612] Kleinknecht, K.: *Detectors for Particle Radiation.* Cambridge University Press, 1998.

[613] Kleinknecht, K.: *Detektoren für Teilchenstrahlung.* Teubner Verlag, 4. Auflage, 2005.

[614] Klempt, W.: Review of particle identification by time-of-flight techniques. In: *Nucl. Inst. Meth.* A433 (1999), p. 542. doi: 10.1016/S0168-9002(99)00323-X

[615] Klepser, S.: *Reconstruction of Extensive Air Showers and Measurement of the Cosmic Ray Energy Spectrum in the Range of 1–80 PeV at the South Pole*, Humboldt-Universität zu Berlin, Diss., 2008. doi: 10.18452/15790

[616] Knoll, G.F.: *Radiation Detection and Measurement.* 4th ed. John Wiley and Sons, 2010.

[617] Knopf, A.-C.; Lomax, A.: In vivo proton range verification: a review. In: *Phys. in Med. Biol.* 58 (2013), p. R131. doi: 10.1088/0031-9155/58/15/R131

[618] Knulst, W.; Luiten, O.J.; Wiel, M.J. van d.; Verhoeven, J.: Observation of narrow band Si L-edge Cerenkov radiation generated by 5 MeV electrons. In: *Appl. Phys. Lett.* 79 (2001), p. 2999. doi: 10.1063/1.1415049

[619] Koch, H.W.; Motz, J.W.: Bremsstrahlung cross-section formulas and related data. In: *Rev. Mod. Phys.* 31 (1959), p. 920. doi: 10.1103/RevModPhys.31.920

[620] Kodama, K. et al. (DONUT Collaboration): Observation of tau neutrino interactions. In: *Phys. Lett.* B 504 (2001), p. 218. doi: 10.1016/S0370-2693(01)00307-0

[621] Kodama, K. et al.: Identification of neutrino interactions using the DONUT spectrometer. In: *Nucl. Inst. Meth. A* 516 (2004), p. 21. doi: 10.1016/j.nima.2003.07.035

[622] Kohler, M. et al.: Beam test measurements with 3D-DDTC silicon strip detectors on n-type substrate. In: *IEEE Trans. Nucl. Sci.* 57 (2010), p. 2987. doi: 10.1109/TNS.2010.2058863

[623] Koike, J.; Parkin, D.M.; Mitchell, T.E.: Displacement threshold energy for type IIa diamond. In: *Appl. Phys. Lett.* 60 (1992), p. 1450. doi: 10.1063/1.107267

[624] Kolanoski, H.; Wermes, N.: *Teilchendetektoren – Grundlagen und Anwendungen*. Springer, 2016. doi: 10.1007/978-3-662-45350-6

[625] Kölbig, K.S.; Schorr, B.: A program package for the Landau distribution. In: *Comput. Phys. Commun.* 31 (1984), p. 97. doi: 10.1016/0010-4655(84)90085-7. – (Erratum: *Comput. Phys. Commun.* 178 (2008), p. 972)

[626] Komin, N.: *Detection of Gamma Rays from the Supernova Remnant RX J0852.0-4622 with H.E.S.S.*, Humboldt Universität zu Berlin, Diss., 2005. doi: 10.18452/15551

[627] Kopitzki, K.; Herzog, P.: *Einführung in die Festkörperphysik*. Springer, 2007. http://www.springer.com/de/book/9783835101449

[628] Kopp, G.; Lean, J.L.: A new, lower value of total solar irradiance: evidence and climate significance. In: *Geophys. Res. Lett.* 38 (2011), p. L01706. doi: 10.1029/2010GL045777

[629] Kopp, S. E.: Accelerator-based neutrino beams. In: *Phys. Rep.* 439 (2007), p. 101. doi: 10.1016/j.physrep.2006.11.004

[630] Korff, S.A.: *Electron and Nuclear Counters*. Van Nostrand, 1946.

[631] Kramberger, G. et al.: Modeling of electric field in silicon micro-strip detectors irradiated with neutrons and pions. In: *JINST* 9 (2014), p. P10016. doi: 10.1088/1748-0221/9/10/P10016

[632] Kramberger, G.; Cindro, V.; Mandic, I.; Mikuz, M.; Milovanovic, M.; Zavrtanik, M.; Zagar, K.: Investigation of irradiated silicon detectors by edge-TCT. In: *IEEE Trans. Nucl. Sci.* 57 (2010), p. 2294. doi: 10.1109/TNS.2010.2051957

[633] Kramberger, G.; Cindro, V.; Mandic, I.; Mikuz, M.; Zavrtanik, M.: Effective trapping time of electrons and holes in different silicon materials irradiated with neutrons, protons and pions. In: *Nucl. Inst. Meth. A* 481 (2002), p. 297. doi: 10.1016/S0168-9002(01)01263-3

[634] Kraus, J.D.: *Electromagnetics*. 3rd ed. McGraw-Hill, 1984.

[635] Krautscheid, T. et al.: Gridpix: Production and application of integrated pixel readouts. Proc.: *12th Pisa Meeting on Advanced Detectors*. In: *Nucl. Inst. Meth. A* 718 (2013), p. 391. doi: 10.1016/j.nima.2012.10.055

[636] Krizan, P.; Korpar, S.; Iijima, T.: Study of a nonhomogeneous aerogel radiator in a proximity focusing RICH detector. In: *Nucl. Inst. Meth. A* 565 (2006), p. 457. doi: 10.1016/j.nima.2006.05.233

[637] Krüger, H.; Fink, J.; Kraft, E.; Wermes, N.; Fischer, P. et al.: CIX – a detector for spectrally enhanced x-ray imaging by simultaneous counting and integrating. Proc.: *Medical Imaging 2008: Physics of Medical Imaging*. In: *Proc. SPIE* 6913 (2008), p. 245. doi: 10.1117/12.771706

[638] Lachnit, W.: *Studien zur Auslese von CsI-, BGO und BaF_2-Szintillationskristallen*, Universität Bonn, diploma thesis, 1993. BONN-IR-94-40.

[639] Laker, K.R; Sansen, W.M.C.: *Design of Analog Integrated Circuits and Systems*. 5th ed. McGraw-Hill, 1994.

[640] Landau, L.D.: On the energy loss of fast particles by ionization. In: *J. Phys. (USSR)* 8 (1944), p. 201. – in [641].

[641] Landau, L.D.: On the energy loss of fast particles by ionization. In: Ter Haar, D. T. (ed.): *Collected Papers of L.D. Landau*. Pergamon, 1965, p. 417. doi: 10.1016/B978-0-08-010586-4.50061-4

[642] Lang, R.F.; Seidel, W.: Search for Dark Matter with CRESST. In: *New J.*

Phys. 11 (2009), p. 105017. doi: 10.1088/1367-2630/11/10/105017

[643] Langner, J.: *Event-Driven Motion Compensation in Positron Emission Tomography: Development of a Clinically Applicable Method*, TU Dresden, Diss., 2008. http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-23509

[644] Lari, T. et al.: Characterization and modeling of non-uniform charge collection in CVD diamond pixel detectors. In: *Nucl. Inst. Meth. A* 537 (2005), p. 581. doi: 10.1016/j.nima.2004.08.069

[645] Lazo, M.S.; Woodall, D.M.; McDaniel, P.J.: Silicon and silicon dioxide neutron damage functions. Proc.: *Fast Burt. React. Workshop, 1986.* Sandia National Laboratories (SAND87-0098), vol. 1 (1987), p. 85.

[646] Lechner, P. et al.: Silicon drift detectors for high resolution room temperature X-ray spectroscopy. In: *Nucl. Inst. Meth. A* 377 (1996), p. 346. doi: 10.1016/0168-9002(96)00210-0

[647] Lecoq, P. et al.: *Inorganic Scintillators for Detector Systems: Physical Principles and Crystal Engineering.* Springer, 2006. doi: 10.1007/3-540-27768-4

[648] Lee, Y.H.; Corbett, J.W.: EPR studies of defects in electron-irradiated silicon – triplet-state of vacancy-oxygen complexes. In: *Phys. Rev. B* 13 (1976), p. 2653. doi: 10.1103/PhysRevB.13.2653

[649] Lehaut, G.: Scintillation properties of N_2 and CF_4 and performances of a scintillating ionization chamber. In: *Nucl. Inst. Meth. A* 797 (2015), p. 57. doi: 10.1016/j.nima.2015.05.050

[650] Lemarenko, M. et al.: Test results of the data handling processor for the DEPFET pixel vertex detector. In: *JINST* 8 (2013), p. C01032. doi: 10.1088/1748-0221/8/01/C01032

[651] Lenzen, G.; Schyns, E.; Thadome, J.; Werner, J.: The Use of fluorocarbon radiators in the DELPHI RICH detectors. In: *Nucl. Inst. Meth. A* 343 (1994), p. 268. doi: 10.1016/0168-9002(94)90562-2

[652] Leo, R.W.: *Techniques for Nuclear and Particle Physics Experiments.* 2nd ed. Springer, 1994.

[653] Leroy, C.; Rancoita, P.-G.: *Radiation Interaction in Matter and Detection.* World Scientific, 2004.

[654] Leroy, C.; Rancoita, P.-G.: *Silicon Solid State Devices and Radiation Detection.* World Scientific, 2012.

[655] Leroy Davis: Interface buses. Version: 2014. http://www.interfacebus.com

[656] LHCb Collaboration: LHCb-Public Home Page. http://lhcb-public.web.cern.ch/lhcb-public/Welcome_270811.html

[657] LHCb Collaboration: The LHCb RICH pixel hybrid photon detectors. http://cern.ch/~gys/LHCb/PixelHPDs.htm

[658] Li, S.S.: *Semiconductor Physical Electronics.* 2nd ed. Springer, 2006.

[659] Lin, J. F.; Li, S.S.; Linares, L.C; Teng, K.W.: Theoretical analysis of Hall factor and Hall mobility in p-type silicon. In: *Solid State Electron.* 24 (1981), p. 827. doi: 10.1016/0038-1101(81)90098-8

[660] Lindstrom, G. et al. (RD48 (ROSE) Collaboration): Radiation hard silicon detectors – developments by the RD48 (ROSE) Collaboration. In: *Nucl. Inst. Meth. A* 466 (2001), p. 308. doi: 10.1016/S0168-9002(01)00560-5

[661] Lindstrom, G.; Moll, M.; Fretwurst, E.: Radiation hardness of silicon detectors: a challenge from high-energy physics. In: *Nucl. Inst. Meth. A* 426 (1999), p. 1. doi: 10.1016/S0168-9002(98)01462-4

[662] Lint, V.A.J. van et al.: *Mechanisms of Radiation Effects in Electronic Materials*. John Wiley and Sons, 1980.

[663] Lipari, P.; Stanev, T.: Propagation of multi-TeV muons. In: *Phys. Rev. D* 44 (1991), p. 3543. doi: 10.1103/PhysRevD.44.3543

[664] Lippmann, C.: *Detector Physics of Resistive Plate Chambers*, Universität Frankfurt, Dissertation, 2003. http://cds.cern.ch/record/1303626/

[665] Lippmann, C.: Particle identification. In: *Nucl. Inst. Meth. A* 666 (2012), p. 148. doi: 10.1016/j.nima.2011.03.009

[666] Lippmann, C.; Riegler, W.: Space charge effects in resistive plate chambers. In: *Nucl. Inst. Meth. A* 517 (2004), p. 54. doi: 10.1016/j.nima.2003.08.174

[667] Lippmann, C.; Riegler, W.; Kalweit, A.: Rate effects in resistive plate chambers. In: *Nucl. Phys. Proc. Suppl.* 158 (2006), p. 127. doi: 10.1016/j.nuclphysbps.2006.07.037

[668] Löcker, M. et al.: Single photon counting x-ray imaging with Si and CdTe single chip pixel detectors and multichip pixel modules. In: *IEEE Trans. Nucl. Sci.* 51 (2004), p. 1717. doi: 10.1109/TNS.2004.832610

[669] Loef, E.V.D. van; Dorenbos, P.; Eijk, C.W.E. van; Krämer, K.; Güdel, H.U.: High-energy-resolution scintillator: Ce^{3+} activated $LaCl_3$. In: *Appl. Phys. Lett.* 77 (2000), p. 1467. doi: 10.1063/1.1308053

[670] Lohrmann, E.; Söding, P.: *Von schnellen Teilchen und hellem Licht – 50 Jahre Deutsches Elektronen-Synchrotron DESY*. Wiley-VCH, 2009. http://www.wiley-vch.de/publish/dt/books/bySubjectPH00/ISBN3-527-40990-4

[671] Lohse, T.; Witzeling, W.: The time projection chamber. In: *[850]*, p. 81. doi: 10.1142/9789814360333_0002

[672] Longo, E.; Sestili, I.: Monte Carlo calculation of photon initiated electromagnetic showers in lead glass. In: *Nucl. Inst. Meth.* 128 (1975), p. 283. doi: 10.1016/0029-554X(75)90679-5

[673] Lubelsmeyer, K. et al.: Upgrade of the Alpha Magnetic Spectrometer (AMS-02) for long term operation on the International Space Station (ISS). In: *Nucl. Inst. Meth. A* 654 (2011), p. 639. doi: 10.1016/j.nima.2011.06.051

[674] Lund-Jensen, B.: *Single-Photon Detectors for Cherenkov Ring Imaging*, Uppsala University, Diss., 1988. http://cds.cern.ch/record/193058/

[675] Lupberger, M. et al.: InGrid: Pixelated Micromegas detectors for a pixel-TPC. Proc.: *3rd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2014)*. In: *PoS* TIPP2014 (2014), p. 225. doi: 10.22323/1.213.0225

[676] Lutz, G.: Correlated noise in silicon strip detector readout. In: *Nucl. Inst. Meth. A* 309 (1991), p. 545. doi: 10.1016/0168-9002(91)90260-W

[677] Lutz, G.: *Semiconductor Radiation Detectors*. Springer, 1999. doi: 10.1007/978-3-540-71679-2

[678] Lutz, G. et al.: DEPFET-detectors: new developments. In: *Nucl. Inst. Meth. A* 572 (2007), p. 311. doi: 10.1016/j.nima.2006.10.339

[679] Lynch, G.R.; Dahl, O.I.: Approximations to multiple Coulomb scattering. In: *Nucl. Inst. Meth.* B 58 (1991), p. 6.

[680] Macchiolo, A. et al.: Thin n-in-p pixel sensors and the SLID-ICV vertical integration technology for the ATLAS upgrade at the HL-LHC. In: *Nucl. Inst. Meth. A* 731 (2013), p. 210. doi: 10.1016/j.nima.2013.04.077

[681] Madelung, O.: *Festkörpertheorie III*. Springer-Verlag, 2013.

[682] Mager, M. (ALICE): ALPIDE, the monolithic active pixel sensor for the ALICE ITS upgrade. In: *Nucl. Inst. Meth. A* 824 (2016), p. 434. doi: 10.1016/j.nima.2015.09.057

[683] Magic Collaboration: The MAGIC Telescopes. http://magic.mppmu.mpg.de

[684] Maire, M.: Electromagnetic interactions of particles with matter. Proc.: *IX Int. Conf. on Calorimetry in High Energy Physics,* Annecy. In: *Frascati Phys. Ser.* 21 (2000), p. 3.

[685] Makarenko, L.F.; Lastovskii, S.B.; Yakushevich, H.S.; Moll, M.; Pintilie, I.: Effect of electron injection on defect reactions in irradiated silicon containing boron, carbon, and oxygen. In: *J. Appl. Phys.* 123 (2018), p. 161578. doi: 10.1063/1.5010965

[686] Maloberti, F.: *Data Converters.* 1st. Springer, 2010. doi: 10.1007/978-0-387-32486-9

[687] Malter, L.: Thin film field emission. In: *Phys. Rev.* 50 (1936), p. 48. doi: 10.1103/PhysRev.50.48

[688] Mandya, D.: Advantages of 65-nm Technology over 90-nm for Sync/NoBL SRAMs. Version: 2016. http://www.cypress.com/file/157776/download. Cypress Note AN94077, Doc. No. 001-94077 Rev. *C

[689] Manghisioni, M.; Ratti, L.; Re, V.; Speziali, V.: Submicron CMOS technologies for low-noise analog front-end circuits. In: *IEEE Trans. Nucl. Sci.* 49(4) (2002), p. 1783. doi: 10.1109/TNS.2002.801540

[690] Mangiarotti, A.; Gobbi, A.: On the physical origin of tails in the time response of spark counters. In: *Nucl. Inst. Meth. A* 482 (2002), p. 192. doi: 10.1016/S0168-9002(01)01623-0

[691] Mankel, R.: Pattern recognition and event reconstruction in particle physics experiments. In: *Rep. Prog. Phys.* 67 (2004), p. 553. doi: 10.1088/0034-4885/67/4/R03

[692] Mankel, R.; Spiridonov, A.: The concurrent track evolution algorithm: extension for track finding in the inhomogeneous magnetic field of the HERA-B spectrometer. In: *Nucl. Inst. Meth. A* 426 (1999), p. 268. doi: 10.1016/S0168-9002(99)00013-3

[693] Mantaro Impedance Calculator. http://www.mantaro.com/resources/impedance_calculator.htm#microstrip_impedance

[694] Mapelli, A. (ATLAS ALFA Collaboration): ALFA: Absolute Luminosity For ATLAS: development of a scintillating fibre tracker to determine the absolute LHC luminosity at ATLAS. In: *Nucl. Phys. Proc. Suppl.* 197 (2009), p. 387. doi: 10.1016/j.nuclphysbps.2009.10.110

[695] Marcuse, D.: Compression of a bundle of light rays. In: *Appl. Opt.* 10 (1971), p. 494. doi: 10.1364/AO.10.000494

[696] Markov, M.A.: On high energy neutrino physics. Proc.: *1960 Annual International Conference on High-Energy Physics.* University of Rochester, (1960), p. 578. http://inspirehep.net/record/1341439/files/C60-08-25-p578.pdf

[697] Marmier, P.; Sheldon, E.: *Physics of Nuclei and Particles.* Academic Press, 1969.

[698] Marrodán Undagoitia, T.; Rauch, L.: Dark matter direct-detection experiments. In: *J. Phys.* G 43 (2016), p. 013001. doi: 10.1088/0954-3899/43/1/013001

[699] Mathes, M. et al.: Test beam characterizations of 3D silicon pixel detectors. In: *IEEE Trans. Nucl. Sci.* 55 (2008), p. 3731. doi: 10.1109/TNS.2008.2005630

[700] Mathews, J.H.; Howell, R.W.: *Complex Analysis for Mathematics and Engineering*. Jones and Bartlett Learning, 2006.

[701] Maxim Integrated: ADC Tutorials. http://www.maximintegrated.com/en/design/techdocs/tutorials

[702] May, P.W.: CVD Diamond – a New Technology for the Future? http://www.chm.bris.ac.uk/pt/diamond/end.htm

[703] McPeak, J.: Radiation detection devices: various means to quantify various types of radiation emitted. Version: 2010. http://magnusslayde.wordpress.com/article/radiation-detection-devices-33qgvqgci3cqt-7/.

[704] Meek, J.M.; Craggs, J.D.: *Electrical Breakdown of Gases*. Clarendon Press, 1954. http://www.archive.org/details/electricalbreakd031039mbp

[705] Meidinger, N. et al.: pnCCD for photon detection from near-infrared to X-rays. In: *Nucl. Inst. Meth. A* 565 (2006), p. 251. doi: 10.1016/j.nima.2006.05.006

[706] Meynants, G.; Dierickx, B.; Scheffer, D.: CMOS active pixel image sensor with CCD performance. Proc.: *Advanced Focal Plane Arrays and Electronic Cameras II*. SPIE, vol. 3410 (1998), p. 68. doi: 10.1117/12.323997

[707] Migdal, A.B.: Bremsstrahlung and pair production in condensed media at high-energies. In: *Phys. Rev.* 103 (1956), p. 1811. doi: 10.1103/PhysRev.103.1811

[708] Milotti, E.: 1/f noise: a pedagogical review. In: *arXiv:physics/0204033* (2002).

[709] Miyoshi, T. et al.: Recent progress of pixel detector R&D based on SOI technology. Proc.: *2nd International Conference on Technology and Instrumentation in Particle Physics 2011*. In: *Phys. Procedia* 37 (2012), p. 1039. doi: 10.1016/j.phpro.2012.02.450

[710] Moliere, G.: Theorie der Streuung schneller geladener Teilchen I. Einzelstreuung am abgeschirmten Coulomb-Feld. In: *Z. Naturforsch.* A2 (1947), p. 133. http://zfn.mpdl.mpg.de/data/Reihe_A/2/ZNA-1947-2a-0133.pdf

[711] Moliere, G.: Theorie der Streuung schneller geladener Teilchen II. Mehrfach- und Vielfachstreuung. In: *Z. Naturforsch.* A3 (1948), p. 78. http://zfn.mpdl.mpg.de/data/Reihe_A/3/ZNA-1948-3a-0078.pdf

[712] Moll, M.: *Radiation Damage in Silicon Particle Detectors - Microscopic Defects and Macroscopic Properties*, University of Hamburg, Diss., 1999. http://cds.cern.ch/record/425274

[713] Moll, M. (CERN-RD50 Collaboration): Recent progress of the RD50 Collaboration – development of radiation tolerant tracking detectors. Proc.: *22nd International Workshop on Vertex Detectors (Vertex 2013)*. In: *PoS* Vertex2013 (2013), p. 026.

[714] Moll, M.: Displacement damage in silicon detectors for high energy physics. Proc.: *Radecs 2017*, Geneva. In: *IEEE Trans. Nucl. Sci.* 65 (2018), p. 1561. doi: 10.1109/TNS.2018.2819506

[715] Moll, M.; Fretwurst, E.; Lindstrom, G. (RD48(ROSE) Collaboration): Investigation on the improved radiation hardness of silicon detectors with high oxygen concentration. In: *Nucl. Inst. Meth. A* 439 (2000), p. 282. doi: 10.1016/S0168-9002(99)00842-6

[716] Montgomery, C.G.; Montgomery, D.D.: Geiger–Mueller counters. In: *J.*

of the Franklin Inst. 231 (1941), p. 447. doi: 10.1016/S0016-0032(41)90498-2

[717] Morishima, K. et al.: Development of an automated nuclear emulsion analyzing system. In: *Radiat. Meas.* 50 (2013), p. 237. doi: 10.1016/j.radmeas.2012.06.016

[718] Morse, P.M.; Feshbach, H.: *Methods of Theoretical Physics, Part I and II.* McGraw-Hill, 1953.

[719] Motz, J.W.; Olsen, H.A.; Koch, H.W.: Pair production by photons. In: *Rev. Mod. Phys.* 41 (1969), p. 581. doi: 10.1103/RevModPhys.41.581

[720] NA35 Collaboration: NA35: sulphur-gold collision. Version: Jul. 1991. http://cds.cern.ch/record/39453

[721] Nakamura, H.; Kitamura, H.; Hazama, R.: Radiation measurements with heat-proof polyethylene terephthalate bottles. In: *Proc. Roy. Soc. A* 466 (2010), p. 2847. doi: 10.1098/rspa.2010.0118

[722] Nakamura, K.: Hyper-Kamiokande: a next generation water Cherenkov detector. In: *Int. J. Mod. Phys.* A18 (2003), p. 4053. doi: 10.1142/S0217751X03017361

[723] Nakamura, K.: Solar neutrinos. In: *published in [370]* (2004).

[724] Nakamura, K. et al. (Particle Data Group): Review of particle physics. In: *J. Phys.* G37 (2010), p. 075021. doi: 10.1088/0954-3899/37/7A/075021

[725] Nastase, A.: Mastering electronics design. https://masteringelectronicsdesign.com/, retrieved: July 2019

[726] National Institute of Standards and Technology (NIST): Engineering Metrology Toolbox. http://emtoolbox.nist.gov/

[727] Neeser, W.: *Test und Inbetriebnahme von DEPJFET-Detektoren*, Universität Bonn, diploma thesis, 1996. BONN-IR-96-31.

[728] Nefkens, B.M.K.; Starostin, A.B. (Crystal Ball Collaboration): New results on meson physics with the Crystal Ball detector. Proc.: *6th International Workshop Meson'2000, Production, properties and interaction of mesons,* Cracow, Poland. In: *Acta Phys. Polon.* B 31 (2000), p. 2669. https://www.actaphys.uj.edu.pl/R/31/10/2669/pdf

[729] Neiser, A. (A2 Collaboration): Current status and performance of the Crystal Ball and TAPS calorimeter. Proc.: *Proceedings of 16th Int. Conference on Calorimetry in High Energy Physics (CALOR 2014),* Giessen, Germany. In: *J. Phys. Conf. Ser.* 587 (2015), p. 012041. doi: 10.1088/1742-6596/587/1/012041

[730] Nemethy, P.; Oddone, P.J.; Toge, N.; Ishibashi, A.: Gated time projection chamber. In: *Nucl. Inst. Meth.* 212 (1983), p. 273. doi: 10.1016/0167-5087(83)90702-0

[731] Nesladek, M. et al.: Charge transport in high mobility single crystal diamond. Proc.: *18th European Conference on Diamond, Diamond-Like Materials, Carbon Nanotubes, Nitrides and Silicon Carbide,* Berlin, Germany. In: *Diam. Relat. Mater.* 17 (2008), p. 1235. doi: 10.1016/j.diamond.2008.03.015

[732] Neuert, H.: *Kernphysikalische Messverfahren.* Karlsruhe: Verlag G. Braun, 1966.

[733] Newman-Holmes, C.; Schmidt, Eugene E.; Yamada, R.: Measurement of the magnetic field of the CDF magnet. In: *Nucl. Inst. Meth. A* 274 (1989), p. 443. doi: 10.1016/0168-9002(89)90174-5

[734] Neyret, D. et al.: New pixelized Micromegas detector for the COMPASS experiment. In: *JINST* 4 (2009), p. P12004. doi: 10.1088/1748-0221/7/03/C03006

[735] Ni, K.: *Development of a Liquid Xenon Time Projection Chamber for the XENON Dark Matter Search*, Columbia University, Diss., 2006. http://xenon.astro.columbia.edu/thesis/Kaixuan.Ni_Thesis.pdf

[736] Niebuhr, C.: Aging in the central jet chamber of the H1 experiment. In: *Nucl. Inst. Meth. A* 515 (2003), p. 43. doi: 10.1016/j.nima.2003.08.128

[737] NIM Committee: Standard NIM instrumentation system / DOE/ER. DOE/ER-0457T. 1990. doi: 10.2172/7120327

[738] Nishida, S. et al.: Studies of a proximity focusing aerogel RICH for the Belle upgrade. Proc.: *IEEE Symposium Conference Record Nuclear Science*, vol. 3 (2004), p. 1951. doi: 10.1109/NSSMIC.2004.1462628

[739] NMDB: Neutron Monitor Database. http://www.nmdb.eu

[740] Novotny, R.: Performance of the BaF-2 calorimeter TAPS. In: *Nucl. Phys. Proc. Suppl.* 61B (1998), p. 137. doi: 10.1016/S0920-5632(97)00552-5

[741] Nygren, D.R.; Marx, J.N.: The time projection chamber. In: *Phys. Today* 31N10 (1978), p. 46. doi: 10.1063/1.2994775

[742] Nyquist, H.: Thermal agitation of electricity in conductors. In: *Phys. Rev.* 32 (1928), p. 110. doi: 10.1103/PhysRev.32.110

[743] Oed, A.: Position sensitive detector with microstrip anode for electron multiplication with gases. In: *Nucl. Inst. Meth. A* 263 (1988), p. 351.

[744] Ohm, S.: H.E.S.S. performance parameters. 2018. Private communication.

[745] Okubo, S.; Tanaka, H.K.M.: Imaging the density profile of a volcano interior with cosmic-ray muon radiography combined with classical gravimetry. In: *Meas. Sci. Technol.* 23 (2012), p. 042001. doi: 10.1088/0957-0233/23/4/042001

[746] Olive, K.A. et al. (Particle Data Group): Review of particle physics (RPP). In: *Chin. Phys.* C38 (2014), p. 090001. doi: 10.1088/1674-1137/38/9/090001. – An updated RPP is published every 2 years. The online access to the current and previous issues is available from http://pdg.lbl.gov/

[747] Omar, M. A.: *Elementary Solid State Physics: Principles and Applications*. Addison-Wesley, 1975.

[748] Oppenheim, A.V.; Willsky, A.S.: *Signals and Systems*. 2nd ed. Pearson, 2014.

[749] Oreglia, M. et al.: A study of the reaction: $\psi' \to \gamma\gamma J/\psi$. In: *Phys. Rev. D* 25 (1982), p. 2259. doi: 10.1103/PhysRevD.25.2259

[750] Oxford Physics: BEBC. Version: 2014. http://www.physics.ox.ac.uk/dwb/BEBC.pdf

[751] Oyama, K. et al.: The transition radiation detector for ALICE at the LHC. In: *Nucl. Inst. Meth. A* 623 (2010), p. 362. doi: 10.1016/j.nima.2010.02.249

[752] Palladino, V.; Sadoulet, B.: Application of the classical theory of electrons in gases to multiwire proportional and drift chambers. LBL-3013. 1974. http://www.osti.gov/bridge/servlets/purl/4270437-MpNuaH/4270437.pdf

[753] Palladino, V.; Sadoulet, B.: Application of classical theory of electrons in gases to drift proportional chambers. In: *Nucl. Inst. Meth.* 128 (1975), p. 323. doi: 10.1016/0029-554X(75)90682-5

[754] Palmonari, F.M. (CMS Collaboration): CMS tracker performance. In: *Nucl. Inst. Meth. A* 699 (2013), p. 144. doi: 10.1016/j.nima.2012.06.010

[755] Pan, L.S.; Kania, D.R.: *Diamond: Electronic Properties and Applications*. Kluwer Academic Publ., 1995. doi: 10.1007/978-1-4615-2257-7

[756] Pancheshnyi, S. et al. (The LXCat team): LXcat Database. http://www.lx

cat.net, retrieved: 2014
[757] Panta, R.K. et al.: Energy calibration of the pixels of spectral X-ray detectors. In: *IEEE Trans. Med. Imag.* 34 (2015), p. 697. doi: 10.1109/TMI.2014.2337881
[758] Papini, P. et al.: In-flight performances of the PAMELA satellite experiment. In: *Nucl. Inst. Meth. A* 588 (2008), p. 259. doi: 10.1016/j.nima.2008.01.052
[759] Parker, S.I.: A proposed VLSI pixel device for particle detection. In: *Nucl. Inst. Meth. A* 275 (1989), p. 494. doi: 10.1016/0168-9002(89)90736-5
[760] Parkes, C. et al. (LHCb Collaboration): First LHC beam induced tracks reconstructed in the LHCb VELO. In: *Nucl. Inst. Meth. A* 604 (2009), p. 1. doi: 10.1016/j.nima.2009.01.215
[761] Parkhomchuck, V.V.; Pestov, Y.N.; Petrovykh, N.V.: A spark counter with large area. In: *Nucl. Inst. Meth.* 93 (1971), p. 269. doi: 10.1016/0029-554X(71)90475-7
[762] Particle Data Group: Atomic and Nuclear Properties of Materials. http://pdg.lbl.gov/2018/AtomicNuclearProperties. The current version is linked under http://pdg.lbl.gov/
[763] Particle Data Group: Review of particle physics (RPP). http://pdg.lbl.gov
[764] Paschke, J.; Teichert, A.M. (QuarkNet): How to Build a Cosmic-Ray Cloud Chamber. Version: 2008. http://quarknet.fnal.gov/resources/QN_CloudChamberV1_4.pdf
[765] Patrignani, C. et al. (Particle Data Group): Review of particle physics. In: *Chin. Phys.* C40 (2016), p. 100001. doi: 10.1088/1674-1137/40/10/100001
[766] Pawloski, G. (MINOS): The MINOS experiment. Proc.: *13th International Conference on Heavy Quarks and Leptons (HQL 2016),* Blacksburg, VA. In: *PoS* HQL2016 (2017), p. 004. doi: 10.22323/1.274.0004
[767] PCIMG: Open Modular Computing Specifications. Version: 2014. http://www.picmg.org/
[768] Pelgrom, Marcel J.: *Analog-to-Digital Conversion.* Springer, 2013.
[769] Pellegrini, G. et al.: Recent Technological Developments on LGAD and iLGAD Detectors for Tracking and Timing Applications. In: *Nucl. Inst. Meth. A* 831 (2016), p. 24. doi: 10.1016/j.nima.2016.05.066
[770] Peric, I.: A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology. In: *Nucl. Inst. Meth. A* 582 (2007), p. 876. doi: 10.1016/j.nima.2007.07.115
[771] Peric, I. et al.: The FEI3 readout chip for the ATLAS pixel detector. In: *Nucl. Inst. Meth. A* 565 (2006), p. 178. doi: 10.1016/j.nima.2006.05.032
[772] Perkins, D.H.: *Introduction to High Energy Physics.* Cambridge University Press, 2000.
[773] Perl, M.: The discovery of the tau-lepton. In: *[529]*, p. 79. doi: 10.1017/CBO9780511471094.007
[774] Perl, M.L. et al.: Evidence for anomalous lepton production in e^+e^- annihilation. In: *Phys. Rev. Lett.* 35 (1975), p. 1489. doi: 10.1103/PhysRevLett.35.1489
[775] Pernegger, H.: The pixel detector of the ATLAS experiment for LHC Run-2. Proc.: *7th International Workshop on Semiconductor Pixel Detectors for Particles and Imaging (PIXEL 2014),* Niagara Falls, Ontario. In: *JINST* 10 (2015), p. C06012. doi: 10.1088/1748-0221/10/06/C06012
[776] Pernegger, H. et al.: Charge-carrier properties in synthetic single-crystal diamond measured with the transient-current technique. In: *J. Appl. Phys.* 97

(2005), p. 073704. doi: 10.1063/1.1863417
[777] Peso, J. del; Ros, E.: On the energy resolution of electromagnetic sampling calorimeters. In: *Nucl. Inst. Meth. A* 276 (1989), p. 456. doi: 10.1016/0168-9002(89)90571-8
[778] Pestotnik, R. et al.: Aerogel RICH for forward PID at Belle II. In: *Nucl. Inst. Meth. A* 732 (2013), p. 371. doi: 10.1016/j.nima.2013.06.080
[779] Pestov, Y.N.: Status and future developments of spark counters with a localized discharge. In: *Nucl. Inst. Meth.* 196 (1982), p. 45. doi: 10.1016/0029-554X(82)90614-0
[780] Pestov, Y.N.: The Status of Spark Counters with a Localized Discharge. In: *Nucl. Inst. Meth. A* 265 (1988), p. 150. doi: 10.1016/0168-9002(88)91066-2
[781] Philipp, K.: Zur Existenz der weitreichenden α-Strahlen des Radium C. In: *Naturwissenschaften* 14 (1926), p. 1203. doi: 10.1007/BF01451770
[782] Philippot, J.C.: Automatic Processing of Diode Spectrometry Results. In: *IEEE Trans. Nucl. Sci.* NS-17/3 (1970), p. 446. doi: 10.1109/TNS.1970.4325723
[783] Pierre Auger Collaboration : Pierre Auger Observatory. http://auger.org
[784] Pintilie, I.; Buda, M.; Fretwurst, E.; Lindstrom, G.; Stahl, J.: Stable radiation-induced donor generation and its influence on the radiation tolerance of silicon diodes. In: *Nucl. Inst. Meth. A* 556 (2006), p. 197. doi: 10.1016/j.nima.2005.10.013
[785] Pintilie, I.; Fretwurst, E.; Lindstrom, G.: Cluster related hole traps with enhanced-field-emission: the source for long term annealing in hadron irradiated Si diodes. In: *Appl. Phys. Lett.* 92 (2008), p. 024101. doi: 10.1063/1.2832646
[786] Pintilie, I.; Lindstroem, G.; Junkes, A.; Fretwurst, E.: Radiation-induced point- and cluster-related defects with strong impact on damage properties of silicon detectors. In: *Nucl. Inst. Meth. A* 611 (2009), p. 52. doi: 10.1016/j.nima.2009.09.065
[787] Pinto, S.D.: Micropattern gas detector technologies and applications the work of the RD51 collaboration. Proc.: *IEEE Nuclear Science Symposium (NSS/MIC)*, (2010), p. 802. doi: 10.1109/NSSMIC.2010.5873870
[788] Pinto, S.D. (for the RD51 collaboration): Micropattern gas detector technologies and applications, the work of the RD51 collaboration. In: *arXiv:1011.5529* (2010).
[789] Pitzl, D. et al.: The H1 silicon vertex detector. In: *Nucl. Inst. Meth. A* 454 (2000), p. 334. doi: 10.1016/S0168-9002(00)00488-5
[790] Plewnia, S. et al.: A sampling calorimeter with warm-liquid ionization chambers. In: *Nucl. Inst. Meth. A* 566 (2006), p. 422. doi: 10.1016/j.nima.2006.07.051
[791] Polyanskiy, M.: Refractive Index Info. http://refractiveindex.info
[792] Porro, M. et al.: Spectroscopic performance of the DePMOS detector/amplifier device with respect to different filtering techniques and operating conditions. In: *IEEE Trans. Nucl. Sci.* 53 (2006), p. 401. doi: 10.1109/TNS.2006.869850
[793] Powell, C.F.: Nobel Lecture: The cosmic radiation. http://www.nobelprize.org/nobel_prizes/physics/laureates/1950/powell-lecture.html
[794] Powell, C.F.: Mesons. In: *Rep. Prog. Phys.* 13 (1950), p. 350. doi: 10.1088/0034-4885/13/1/309
[795] Press, W.H.: Flicker Noises in Astronomy and Elsewhere. In: *Comments*

Astrophys. 7 (1978), p. 103.
[796] Prosenjit, R.-C.: *Handbook of Microlithography, Micromachining, and Microfabrication.* Institution of Engineering and Technology, 1997.
[797] Qiu, Xi-Yu et al.: Position reconstruction in fission fragment detection using the low pressure MWPC technique for the JLab experiment E02-017. In: *Chinese Physics C* 38 (2014), p. 074003. doi: 10.1088/1674-1137/38/7/074003
[798] Quadt, A.: Representation of stereo wire layers in a cylindrical drift chamber. 2003. Private communication.
[799] Radeka, V.: Low-Noise Techniques in Detectors. In: *Annu. Rev. Nucl. Part. Sci.* 38 (1988), p. 217. doi: 10.1146/annurev.ns.38.120188.001245
[800] Radeka, V.: The sign in Ramo's equation. 2011. Private communication.
[801] Radeka, V.; Rehak, P.: Second Coordinate Readout in Drift Chambers by Charge Division. In: *IEEE Trans. Nucl. Sci.* 25 (1978), p. 46.
[802] Radu, R. et al.: Investigation of point and extended defects in electron irradiated silicon: dependence on the particle energy. In: *J. Appl. Phys.* 117 (2015), p. 164503. doi: 10.1063/1.4918924
[803] Raether, H.: Electron Avalanches and Breakdown in Gases. In: *Butterworth Advanced Physics Series, London* (1964).
[804] Raizer, Y.P.: *Gas Discharge Physics.* Springer-Verlag, 1991. doi: 10.1007/978-3-642-61247-3
[805] Ramanantsizehena, P.; Gresser, J.; Schultz, G.: Computations of Drift Velocities for Chambers Working in Magnetic Fields. In: *Nucl. Inst. Meth.* 178 (1980), p. 253. doi: 10.1016/0029-554X(80)90886-1
[806] Ramo, S.: Currents Induced by Electron Motion. In: *Proceedings of the I.R.E* 27 (1939), p. 584. doi: 10.1109/JRPROC.1939.228757
[807] Ramsauer, C.: Über den Wirkungsquerschnitt der Gasmoleküle gegenüber langsamen Elektronen. In: *Ann. Phys.* 369 (1921), p. 513. doi: 10.1002/andp.19213690603
[808] Ratcliff, B.: The DIRC counter: A New type of particle identification device for B factories. In: *Conf. Proc.* C921117 (1992), p. 331.
[809] Ratti, L. et al.: CMOS MAPS with Fully Integrated, Hybrid-pixel-like Analog Front-end Electronics. In: *eConf* C0604032 (2006), p. 0008. http://www.slac.stanford.edu/econf/C0604032/papers/0008.PDF
[810] Raymond, M. et al.: The CMS tracker APV25 0.25-mu-m CMOS readout chip. Proc.: *6th Workshop on Electronics for LHC Experiments,* Krakow, Poland, (2000), p. 130. doi: 10.5170/CERN-2000-010.130
[811] Rehak, P. et al.: Progress in Semiconductor Drift Detectors. In: *Nucl. Inst. Meth. A* 248 (1986), p. 367. doi: 10.1016/0168-9002(86)91021-1
[812] Reilly, D.; Ensslin, N.; Smith, H.: *Passive Non-Destructive Assay of Nuclear Materials.* The Commission, 1991.
[813] Reines, F.; Cowan, C.: The Reines-Cowan experiments: Detecting the Poltergeist. In: *Los Alamos Sci.* 25 (1997), p. 4. http://library.lanl.gov/cgi-bin/getfile?00326606.pdf
[814] Reines, F.; Cowan, C.L.: Detection of the free neutrino. In: *Phys.Rev.* 92 (1953), p. 830. doi: 10.1103/PhysRev.92.830
[815] RENO Collaboration: Reactor Experiment for Neutrino Oscillation. http://hcpl.knu.ac.kr/neutrino/neutrino.html
[816] Rice-Evans, P.: Spark and streamer chambers. In: *J. Phys. E: Sci. Instrum.*

2 (1969), p. 221. doi: 10.1088/0022-3735/2/3/201
[817] Rice-Evans, P.: *Spark, Streamer, Proportional and Drift Chambers*. Richelieu Press, 1974.
[818] Rideout, V.L.: A review of the theory and technology for ohmic contacts to group III–V compound semiconductors. In: *Solid State Electron.* 18 (1975), p. 541. doi: 10.1016/0038-1101(75)90031-3
[819] Riege, H.: High-frequency and pulse response of coaxial transmission cables with conductor, dielectric and semiconductor losses. CERN-70-04. 1970.
[820] Riegler, W.: Extended theorems for signal induction in particle detectors. Proc.: *10th Vienna Conference on Instrumentation*. In: *Nucl. Inst. Meth. A* 535 (2004), p. 287. doi: 10.1016/j.nima.2004.07.129
[821] Riegler, W.: Electric fields, weighting fields, signals and charge diffusion in detectors including resistive materials. In: *JINST* 11 (2016), p. P11002. doi: 10.1088/1748-0221/11/11/P11002
[822] Riegler, W.: An application of extensions of the Ramo–Shockley theorem to signals in silicon sensors. In: *Nucl. Inst. Meth. A* 940 (2019), p. 453. doi: 10.1016/j.nima.2019.06.056
[823] Riegler, W.; Aglieri Rinella, G.: Point charge potential and weighting field of a pixel or pad in a plane condenser. In: *Nucl. Inst. Meth. A* 767 (2014), p. 267. doi: 10.1016/j.nima.2014.08.044
[824] Riegler, W.; Lippmann, C.: The physics of resistive plate chambers. In: *Nucl. Inst. Meth. A* 518 (2004), p. 86. doi: 10.1016/j.nima.2003.10.031
[825] Riegler, W.; Lippmann, C.; Veenhof, R.: Detector physics and simulation of resistive plate chambers. In: *Nucl. Inst. Meth. A* 500 (2003), p. 144. doi: 10.1016/S0168-9002(03)00337-1
[826] Robinson, F.N.H.: *Noise in Electrical Circuits*. Oxford University Press, 1962.
[827] Rochester, G.D.; Wilson, J.G.: *Cloud chamber photographs of the cosmic radiation*. Academic Press/Pergamon Press, 1952.
[828] Rodnyi, P.A.: Core-valance transitions in wide-gap ionic crystals. In: *Sov. Phys. Solid State* 34 (1992), p. 1053.
[829] Rodnyi, P.A.: Progress in fast scintillators. In: *Radiat. Meas.* 33 (2001), p. 605. doi: 10.1016/S1350-4487(01)00068-3
[830] Rogalla, M.: *Systematic Investigation of Gallium Arsenide Radiation Detectors for High Energy Physics Experiments*. Shaker, 1997.
[831] ROOT Development Team: class ROOT::Math::Vavilov. http://root.cern.ch/root/html/ROOT__Math__Vavilov.html
[832] ROOT Development Team: ROOT web page. http://root.cern.ch/drupal
[833] Rose, M.E.; Korff, S.A.: An investigation of the properties of proportional counters I. In: *Phys. Rev.* 59 (1941), p. 850. doi: 10.1103/PhysRev.59.850
[834] Rossi, B.; Greisen, K.: Cosmic-ray theory. In: *Rev. Mod. Phys.* 13 (1941), p. 240. doi: 10.1103/RevModPhys.13.240
[835] Rossi, B.B.: *High Energy Particles*. 1st ed. Prentice-Hall, 1952.
[836] Rossi, B.B.: *Cosmic Rays*. McGraw-Hill, 1964. https://archive.org/details/cosmicrays0000ross
[837] Rossi, L.; Fischer, P.; Rohe, T.; Wermes, N.: *Pixel Detectors: From Fundamentals to Applications*. Springer, 2006.
[838] Ruch, J.G.; Kino, G.S.: Measurement of the velocity-field characteristics of gallium arsenide. In: *Appl. Phys. Lett.* 10 (1967), p. 40.

[839] Ruchti, R.C.: The use of scintillating fibers for charged-particle tracking. In: *Annu. Rev. Nucl. Part. Sci.* 46 (1996), p. 281. doi: 10.1146/annurev.nucl.46.1.281

[840] Ruland, A.M.: Performance and operation of the BaBar calorimeter. In: *J. Phys. Conf. Ser.* 160 (2009), p. 012004. doi: 10.1088/1742-6596/160/1/012004

[841] Rutherford, E.: The scattering of α and β particles by matter and the structure of the atom. In: *Phil. Mag.* 21 (1911), p. 669. doi: 10.1080/14786440508637080

[842] Rutherford, E.; Geiger, H.; Harling, J.: An electrical method of counting the number of α-particles from radio-active substances. In: *Proc. Roy. Soc. A* 81 (1908), p. 141. doi: 10.1098/rspa.1908.0065

[843] Sadrozinski, H. F. W. et al.: Ultra-fast silicon detectors. Proc.: *Proceedings, 9th International Conference on Radiation Effects on Semiconductor Materials Detectors and Devices.* In: *Nucl. Inst. Meth. A* 730 (2013), p. 226. doi: 10.1016/j.nima.2013.06.033

[844] Sadrozinski, H.F.-W. et al.: Ultra-fast silicon detectors (UFSD). Proc.: *10th International 'Hiroshima' Symposium on the Development and Application of Semiconductor Tracking Detectors (HSTD-10),* Xian, China. In: *Nucl. Inst. Meth. A* 831 (2016), p. 18. doi: 10.1016/j.nima.2016.03.093

[845] Sadrozinski, H.F.W.; Seiden, A.; Cartiglia, N.: 4D tracking with ultra-fast silicon detectors. In: *Rep. Prog. Phys.* 81 (2018), p. 026101. doi: 10.1088/1361-6633/aa94d3

[846] Sanfilippo, S.: Hall probes: physics and application to magnetometry. In: *arXiv:1103.1271* (2011).

[847] Santonico, R.; Cardarelli, R.: Development of resistive plate counters. In: *Nucl. Inst. Meth.* 187 (1981), p. 377. doi: 10.1016/0029-554X(81)90363-3

[848] Sauli, F.: GDD: Gaseous Detector Development, CERN web page. http://gdd.web.cern.ch/GDD/

[849] Sauli, F.: Principles of operation of multiwire proportional and drift chambers. In: *[397]*, p. 79. doi: 10.1142/9789814355988_0002. Reprint of CERN-Report 77-09

[850] Sauli, F. (ed.): *Instrumentation in High Energy Physics.* World Scientific, 1992 (Advanced series on directions in high energy physics). doi: 10.1142/1356

[851] Sauli, F.: GEM: A new concept for electron amplification in gas detectors. In: *Nucl. Inst. Meth. A* 386 (1997), p. 531. doi: 10.1016/S0168-9002(96)01172-2

[852] Sauli, F.: Fundamental understanding of aging processes: review of the workshop results. In: *Nucl. Inst. Meth. A* 515 (2003), p. 358. doi: 10.1016/j.nima.2003.09.024

[853] Sauli, F.: *Gaseous Radiation Detectors – Fundamentals and Applications.* 1st ed. Cambridge University Press, 2014. doi: 10.1017/CBO9781107337701

[854] Sauli, F.; Sharma, A.: Micropattern gaseous detectors. In: *Annu. Rev. Nucl. Part. Sci.* 49 (1999), p. 341. doi: 10.1146/annurev.nucl.49.1.341

[855] Sauter, E.: *Grundlagen des Strahlenschutzes.* Leipzig: Fachbuchverlag, 1983.

[856] Saviano, G. et al.: Properties of potential eco-friendly gas replacements for particle detectors in high-energy physics. In: *JINST* 13 (2018), p. P03012. doi: 10.1088/1748-0221/13/03/P03012

[857] Schade, P.; Kaminski, J. (LCTPC Collaboration): A large TPC prototype for a linear collider detector. In: *Nucl. Inst. Meth. A* 628 (2011), p. 128. doi:

10.1016/j.nima.2010.06.300

[858] Schardt, D.; Elsässer, T.; Schulz-Ertner, D.: Heavy-ion tumor therapy: physical and radiobiological benefits. In: *Rev. Mod. Phys.* 82 (2010), p. 383. doi: 10.1103/RevModPhys.82.383

[859] Scheer, J.J.; Laar, J. van: GaAs-Cs: a new type of photoemitter. In: *Solid State Commun.* 3 (1965), p. 189. doi: 10.1016/0038-1098(65)90289-9

[860] Schieck, J. (for the DEPFET Collaboration): DEPFET pixels as a vertex detector for the Belle II experiment. In: *Nucl. Inst. Meth. A* 732 (2013), p. 160. doi: 10.1016/j.nima.2013.05.054

[861] Schiff, L.I.: Energy-angle distribution of thin target bremsstrahlung. In: *Phys. Rev.* 83 (1951), p. 252. doi: 10.1103/PhysRev.83.252

[862] Schiff, L.I.: *Quantum mechanics*. 3rd ed. McGraw-Hill, 1987. https://archive.org/details/quantummechanics0000schi

[863] Schmidt, A. (CMS Collaboration): Performance of track and vertex reconstruction and b-tagging studies with CMS in p p collisions at $\sqrt{s} = 7$ TeV. In: *PoS* KRUGER2010 (2011), p. 032. doi: 10.22323/1.130.0032

[864] Schmidt, B.: *Drift und Diffusion von Elektronen in Methan und Methan-Edelgas-Mischungen*, Universität Heidelberg, Dissertation, 1986. http://d-nb.info/910174938

[865] Schmidt, T.: *Aufbau und Funktionsnachweis eines Optischen Moduls mit optisch-analoger Pulsübertragung für den AMANDA-II- und IceCube-Detektor*, Humboldt-Universität Berlin, Diss., 2002. doi: 10.18452/14910

[866] Schneider, B.: *Computation of the Space Drift Time Relation in Arbitrary Magnetic Fields (in German)*, Universität Bonn, diploma thesis, 1987. https://lib-extopc.kek.jp/preprints/PDF/2000/0030/0030077.pdf. BONN-IR-87-19

[867] Schottky, W.: Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. In: *Ann. Phys.* 362 (1918), p. 541. doi: 10.1002/andp.19183622304

[868] Schottky, W.: Zur Berechnung und Beurteilung des Schroteffektes. In: *Ann. Phys.* 373 (1922), p. 157. doi: 10.1002/andp.19223731007

[869] Schottky, W.: Small-shot effect and flicker effect. In: *Phys. Rev.* 28 (1926), p. 74. doi: 10.1103/PhysRev.28.1331

[870] Schreiner, A.: *Aging Studies of Drift Chambers of the HERA-B Outer Tracker Using CF_4-based Gases*, Humboldt-Universität zu Berlin, Diss., 2001. doi: 10.18452/14703

[871] Schultz, G.; Gresser, J.: A study of transport coefficients of electrons in some gases used in proportional and drift chambers. In: *Nucl. Inst. Meth.* 151 (1978), p. 413. doi: 10.1016/0029-554X(78)90151-9

[872] Schumacher, J.O.; Wettling, W.: *Device physics of silicon solar cells*. London: Imperial College Press, 2001.

[873] Schuster, P.M.: The scientific life of Victor Franz (Francis) Hess (June 24, 1883 – December 17, 1964). In: *Astropart. Phys.* 53 (2014), p. 33. doi: 10.1016/j.astropartphys.2013.05.005

[874] Schwabl, F.: *Statistical Mechanics*. Springer, 2006. doi: 10.1007/3-540-36217-7

[875] Schwandt, J.; Klanner, R.: On the weighting field of irradiated silicon detectors. In: *Nucl. Inst. Meth. A* 942 (2019), p. 162418. doi: 10.1016/j.nima.2019.162418

[876] Schwank, J.R. et al.: Radiation effects in MOS oxides. In: *IEEE Trans. Nucl.*

Sci. 55 (2008), p. 1833. doi: 10.1109/TNS.2008.2001040

[877] Schwanke, U.: *Aufbau und Durchführung von Testexperimenten mit Waben-driftkammern für das HERA-B Experiment*, Humboldt-Universität zu Berlin, diploma thesis, 1996. doi: 10.18452/14005. DOI 10.18452/14005

[878] Schwiening, J.: BaBar DIRC. 2013. Private communication.

[879] Schyns, E.: Status of large area CsI photocathode developments. In: *Nucl. Inst. Meth. A* 494 (2002), p. 441. doi: 10.1016/S0168-9002(02)01520-6

[880] Sclar, N.: Neutral impurity scattering in semiconductors. In: *Phys. Rev.* 104 (1956), p. 1559. doi: 10.1103/PhysRev.104.1559

[881] Seeger, K.: *Semiconductor Physics*. Springer, 2004. doi: 10.1007/978-3-662-09855-4

[882] Segal, J.D. et al.: Second generation monolithic full-depletion radiation sensor with integrated CMOS circuitry. In: *IEEE Nucl. Sci. Symp. Conf. Rec.* 2010 (2010), p. 1896. doi: 10.1109/NSSMIC.2010.5874104

[883] Seguinot, J.; Ypsilantis, T.: Evolution of the RICH Technique. In: *Nucl. Inst. Meth. A* 433 (1999), p. 1. doi: 10.1016/S0168-9002(99)00543-4

[884] Seitz, F.: On the theory of the bubble chamber. In: *Physics of Fluids* 1 (1958), p. 2. doi: 10.1063/1.1724333

[885] Seltzer, S.M.; Berger, M.J.: Evaluation of the collision stopping power of elements and compounds for electrons and positrons. In: *Int. J. Appl. Radiat. Isotopes* 33 (1982), p. 1189. doi: 10.1016/0020-708X(82)90244-7

[886] Seltzer, S.M.; Berger, M.J.: Bremsstrahlung spectra from electron interactions. In: *Nucl. Inst. Meth.* B 12 (1985), p. 95. doi: 10.1016/0168-583X(85)90707-4

[887] Seo, E.S.: Direct measurements of cosmic rays using balloon borne experiments. In: *Astropart. Phys.* 39–40 (2012), p. 76. doi: 10.1016/j.astropartphys.2012.04.002

[888] Serrano, J. et al.: The White Rabbit Project. Proc.: *12th International Conference on Accelerator & Large Experimental Physics Control Systems (ICALEPCS2009),* Kobe, Japan, (2009), p. 93. http://accelconf.web.cern.ch/AccelConf/ICALEPCS2009/papers/tuc004.pdf

[889] Shapiro, M.M.: Nuclear emulsions. In: *[330]*, p. 342. doi: 10.1007/978-3-642-45903-0_8

[890] Sharma, A.: Properties of some gas mixtures used in tracking detectors. SLAC-J-ICFA-16-3. 1998. http://www.slac.stanford.edu/pubs/icfa/summer98/paper3/paper3.pdf

[891] Shockley, A.: Currents to conductors induced by a moving point charge. In: *J. Appl. Phys.* 9 (1938), p. 635. doi: 10.1063/1.1710367

[892] Shockley, W.: The theory of p-n junctions and p-n junction transistors. In: *Bell Syst. Tech. J.* 28 (1949), p. 435.

[893] Shockley, W.: *Electrons and Holes in Semiconductors: With Applications to Transistor Electronics*. Van Nostrand Reinhold, 1950 (The Bell Telephone Laboratories series). https://archive.org/details/ElectronsAndHolesInSemiconductors

[894] Shockley, W.; Read, W. T.: Statistics of the recombinations of holes and electrons. In: *Phys. Rev.* 87 (1952), p. 835. doi: 10.1103/PhysRev.87.835

[895] Singh, B. et al. (PANDA): Technical design report for the PANDA barrel DIRC detector. In: *arXiv:1710.00684* (2017).

[896] Singh, J.: *Physics of Semiconductors and their Heterostructures*. McGraw-Hill, New York, 1993.

[897] Sipos, R.: The DAQ for the single-phase DUNE prototype at CERN. In: *IEEE Trans. Nucl. Sci.* 66 (2019), p. 1210. doi: 10.1109/TNS.2019.2906411

[898] Sirri, G.: Fast automated scanning of OPERA emulsion films. In: *Nucl. Phys. Proc. Suppl.* 172 (2007), p. 324. doi: 10.1016/j.nuclphysbps.2007.08.144

[899] Smirnov, D. (D0 Collaboration): Status of the D0 fiber tracker and preshower detectors. In: *Nucl. Inst. Meth. A* 598 (2009), p. 94. doi: 10.1016/j.nima.2008.08.085

[900] Smith, P.; Inoue, M.; Frey, J.: Electron velocity in Si and GaAs at very high electric fields. In: *Appl. Phys. Lett.* 37 (1980), p. 797.

[901] Snoeys, W. et al.: A process modification for CMOS monolithic active pixel sensors for enhanced depletion, timing performance and radiation tolerance. In: *Nucl. Inst. Meth. A* 871 (2017), p. 90. doi: 10.1016/j.nima.2017.07.046

[902] Sonnenschein, L. (CMS Collaboration): Drift velocity monitoring of the CMS muon drift chambers. In: *PoS* HCP2009 (2009), p. 101. http://pos.sissa.it/archive/conferences/102/101/HCP2009_101.pdf

[903] Spicer, W. E.; Herrera-Gomez, A.: Modern theory and applications of photocathodes. Proc.: *SPIE's 1993 International Symposium on Optics, Imaging and Instrumentation,* San Diego. SPIE, vol. 2022 (1993), p. 18. doi: 10.1117/12.158575

[904] Spicer, W.E.: Photoemissive, photoconductive, and optical absorption studies of alkali-antimony compounds. In: *Phys. Rev.* 112 (1958), p. 114. doi: 10.1103/PhysRev.112.114

[905] Spieler, H.: *Semiconductor Detector Systems*. Oxford University Press, 2005. doi: 10.1093/acprof:oso/9780198527848.001.0001

[906] Sproul, A.B.; Green, M.A.: Intrinsic carrier concentration and minority-carrier mobility of silicon from 77 to 300 K. In: *J. Appl. Phys.* 73 (1993), p. 1214. doi: 10.1063/1.353288

[907] Sproul, A.B.; Green, M.A.; Zhao, J.: Improved value for the silicon intrinsic carrier concentration at 300 K. In: *Appl. Phys. Lett.* 57 (1990), p. 255. doi: 10.1063/1.103707

[908] Srour, J.R.; Marshall, C.J.; Marshall, P.W.: Review of displacement damage effects in silicon devices. In: *IEEE Trans. Nucl. Sci.* 50 (2003), p. 653. doi: 10.1109/TNS.2003.813197

[909] Srour, J.R.; McGarrity, J.M.: Radiation effects on microelectronics in space. In: *Proc. IEEE* 76 (1988), p. 1443. doi: 10.1109/5.90114

[910] Srour, J.R.; Palko, J.W.: Displacement damage effects in irradiated semiconductor devices. In: *IEEE Trans. Nucl. Sci.* 60 (2013), p. 1740. doi: 10.1109/TNS.2013.2261316

[911] Stabin, M.G.: *Radiation Protection and Dosimetry: An Introduction to Health Physics*. Springer, 2007. doi: 10.1007/978-0-387-49983-3

[912] Stanev, T.; Vankov, C.; Streitmatter, R.E.; Ellsworth, R.W.; Bowen, T.: Development of ultrahigh-energy electromagnetic cascades in water and lead including the Landau–Pomeranchuk–Migdal effect. In: *Phys. Rev. D* 25 (1982), p. 1291. doi: 10.1103/PhysRevD.25.1291

[913] Sternheimer, R. M.: The density effect for ionization loss in materials. In: *Phys. Rev.* 88 (1952), p. 851. doi: 10.1103/PhysRev.88.851

[914] Sternheimer, R.M.; Berger, M.J.; Seltzer, S.M.: Density effect for the ionization loss of charged particles in various substances. In: *At. Data and Nucl. Data Tables* 30 (1984), p. 261. doi: 10.1016/0092-640X(84)90002-0

[915] Stokes, T. (Hamamatsu Inc.): Avalanche photodiodes theory and applications. Version: 2005. http://www.photonicsonline.com/doc/avalanche-photodiodes-theory-and-applications-0001

[916] Stone, S.L. et al.: Characteristics of electromagnetic shower sampling counters. In: *Nucl. Inst. Meth.* 151 (1978), p. 387. doi: 10.1016/0029-554X(78)90148-9

[917] Struder, L.: Recent developments in semiconductor detectors and on chip electronics. In: *Nucl. Inst. Meth. A* 283 (1989), p. 387. doi: 10.1016/0168-9002(89)91390-9

[918] Strüder, L. et al.: Fully depleted, backside illuminated, spectroscopic active pixel sensors from the infrared to X-rays. Proc.: *X-Ray Optics, Instruments, and Missions III*. SPIE, vol. 4012 (2000), p. 200. doi: 10.1117/12.391556

[919] Struder, L. et al.: The European Photon Imaging Camera on XMM-Newton: The pn-CCD camera. In: *Astron. Astrophys.* 365 (2001), p. L18. doi: 10.1051/0004-6361:20000066

[920] Strüder, L.; Soltau, H.: High resolution silicon detectors for photons and particles. In: *Radiat. Prot. Dosimetry* 61 (1995), p. 39. http://rpd.oxfordjournals.org/content/61/1-3/39.abstract

[921] Super-Kamiokande: official web site. http://www-sk.icrr.u-tokyo.ac.jp/sk

[922] Suyama, M. et al.: A hybrid photodetector (HPD) with a III–V photocathode. Proc.: *1997 IEEE Nuclear Science Symposium Conference Record*. In: *IEEE Trans. Nucl. Sci.* 45 (1998), p. 572. doi: 10.1109/NSSMIC.1997.672701

[923] Swiss Aviation Resources: ICAO Standard Atmosphere. https://www.aviation.ch/tools-atmosphere.asp, retrieved: Jan. 2020

[924] Swiss Wafers AG: Mono-Silizium Wafers. http://www.swisswafers.ch/d/products/monosiliconwa.html

[925] SYNOPSYS: TCAD: Technology Computer Aided Design. https://www.synopsys.com/silicon/tcad.html, retrieved: 2020

[926] Sze, S.M.: *Semiconductor Devices, Physics and Technology*. 2nd ed. Wiley, 1985.

[927] Sze, S.M.: *Semiconductor Sensors*. Wiley, 1994.

[928] Sze, S.M.; Lee, M.K.: *Semiconductor Devices, Physics and Technology*. 3rd ed. Wiley, 2012.

[929] Sze, S.M.; Ng, K.K.: *Physics of Semiconductor Devices*. 3rd ed. Wiley, 2007. doi: 10.1002/0470068329

[930] Tabata, M. et al.: Development of silica aerogel with any density. Proc.: *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2005*, (2005), p. 816. doi: 10.1109/NSSMIC.2005.1596380

[931] Taft, B.A.; Hickey, B.M.; Wunsch, C.; Baker, D.J., Jr.: Equatorial undercurrent and deeper flows in the central Pacific. In: *Deep Sea Res.* 21 (1974), p. 403.

[932] Tanabashi, M. et al. (Particle Data Group): Review of particle physics. In: *Phys. Rev. D* 98 (2018), p. 030001. doi: 10.1103/PhysRevD.98.030001

[933] Tang, S. et al.: Development and data path integration test of the global Feature EXtractor for the ATLAS Level-1 calorimeter trigger system Phase-I upgrade. In: *Nucl. Inst. Meth. A* 938 (2019), p. 5. doi: 10.1016/j.nima.2019.05.068

[934] Tani, T.: Characterization of nuclear emulsions in overview of photographic emulsions. Proc.: *24th International Conference on Nuclear Tracks in Solids*. In: *Radiat. Meas.* 44 (2009), p. 733. doi: 10.1016/j.radmeas.2009.10.051

[935] Taylor, R.E.: Nobel Lecture: Deep Inelastic Scattering: The Early Years. http://www.nobelprize.org/nobel_prizes/physics/laureates/1990/taylor-lecture.html

[936] Ter-Mikaelian, M.L.: Emission of fast particles in a heterogeneous medium. In: *Nucl. Phys.* 24 (1961), p. 43.

[937] Ter-Mikaelian, M.L.: *High Energy Electromagnetic Processes in Condensed Media*. Wiley-Interscience, 1972.

[938] The Harshaw Chemical Company: Harshaw Scintillation Phosphors. 1978.

[939] Thomson, E.J. et al.: Online track processor for the CDF upgrade. In: *IEEE Trans. Nucl. Sci.* 49 (2002), p. 1063. doi: 10.1109/TNS.2002.1039615

[940] Thomson, M.A.: Particle flow calorimetry and the PandoraPFA algorithm. In: *Nucl. Inst. Meth. A* 611 (2009), p. 25. doi: 10.1016/j.nima.2009.09.009

[941] Thornber, K.K.: Relation of drift velocity to lowfield mobility and highfield saturation velocity. In: *J. Appl. Phys.* 51 (1980), p. 2127.

[942] Tosi, D.: *Measurement of acoustic attenuation in South Pole ice with a retrievable transmitter*, Humboldt Universität zu Berlin, Diss., 2010. doi: 10.18452/16186

[943] Townsend, J.S.: *Electrons in Gases*. Hutchinson, 1948.

[944] Townsend, J.S.; Tizard, H.: The motion of electrons in gases. In: *Proc. Roy. Soc. A* 88 (1913), p. 336. doi: 10.1098/rspa.1913.0034

[945] Toyama, T. et al. (CTA Consortium): Novel photo multiplier tubes for the Cherenkov Telescope Array project. In: *arXiv:1307.5463* (2013).

[946] Treis, J. et al.: DEPMOSFET active pixel sensor prototypes for the XEUS wide field imager. In: *IEEE Trans. Nucl. Sci.* 52 (2005), p. 1083. doi: 10.1109/TNS.2005.852673

[947] Tsagli, S. et al. (NESTOR Collaboration): Recent measurements on the Hamamatsu 13-in., R8055, photomultiplier tubes. In: *Nucl. Inst. Meth. A* 567 (2006), p. 511. doi: 10.1016/j.nima.2006.05.176

[948] Tsai, Y.-S.: Pair production and bremsstrahlung of charged leptons. In: *Rev. Mod. Phys.* 46 (1974), p. 815. doi: 10.1103/RevModPhys.46.815. – (Erratum in [949])

[949] Tsai, Y.-S.: Erratum: Pair production and bremsstrahlung of charged leptons. In: *Rev. Mod. Phys.* 49 (1977), p. 421. doi: 10.1103/RevModPhys.49.421

[950] Tsung, J.-W. et al.: Signal and noise of diamond pixel detectors at high radiation fluences. In: *JINST* 7 (2012), p. P09009. doi: 10.1088/1748-0221/7/09/P09009

[951] Turchetta, R.: Spatial resolution of silicon microstrip detectors. In: *Nucl. Inst. Meth. A* 335 (1993), p. 44. doi: 10.1016/0168-9002(93)90255-G

[952] Turchetta, R. et al.: A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology. In: *Nucl. Inst. Meth. A* 458 (2001), p. 677. doi: 10.1016/S0168-9002(00)00893-7

[953] Ulrici, J. et al.: Imaging performance of a DEPFET pixel bioscope system in tritium autoradiography. In: *Nucl. Inst. Meth. A* 547 (2005), p. 424. doi: 10.1016/j.nima.2005.03.133

[954] Unal, G. (NA48 Collaboration): Performances of the NA48 liquid krypton

calorimeter. Proc.: *9th Conference on Calorimetry in High Energy Physics (CALOR 2000)*, Annecy, France. In: *Frascati Phys. Ser.* 21 (2001), p. 361. arXiv:hep-ex/0012011.

[955] Van Overstraeten, R.; De Man, H.: Measurement of the ionization rates in diffused silicon p-n junctions. In: *Solid State Electron.* 13 (1970), p. 583. doi: 10.1016/0038-1101(70)90139-5

[956] Vavilov, P.V.: Ionization losses of high-energy heavy particles. In: *Sov. Phys. JETP* 5 (1957), p. 749.

[957] Vavra, J: Physics and chemistry of aging: Early developments. In: *Nucl. Inst. Meth. A* 515 (2003), p. 1. doi: 10.1016/j.nima.2003.08.124

[958] Veenhof, R.: Garfield – simulation of gaseous detectors. http://garfield.web.cern.ch/garfield/

[959] Veenhof, R.: GARFIELD, recent developments. In: *Nucl. Inst. Meth. A* 419 (1998), p. 726. doi: 10.1016/S0168-9002(98)00851-1

[960] VERITAS Collaboration: VERITAS. http://veritas.sao.arizona.edu

[961] Voigt, B.: *Sensitivity of the IceCube detector for ultra-high energy electron-neutrino events*, Humboldt Universität zu Berlin, Diss., 2008. doi: 10.18452/15850

[962] Voss, R.F.; Clarke, J.: '1/f noise' in music and speech. In: *Nature* 258 (1975), p. 317. doi: 10.1038/258317a0

[963] Wagner, A.: Distributions of measurements with the test drift chamber FSP of the OPAL experiment. 1991. Private communication.

[964] Wakely, S. P.; Muller, D.; Gahbauer, F.; Plewnia, S.; Horandel, J. R.: Transition radiation detectors for energy measurements at high Lorentz factors. In: *Nucl. Inst. Meth.* A531 (2004), p. 435. doi: 10.1016/j.nima.2004.04.245

[965] Walck, C.: Handbook on statistical distributions for experimentalists. Internal Report SUF-PFY/96-01, University of Stockholm. 2007. http://inspirehep.net/record/1389910/files/suf9601.pdf

[966] Wald, F.V.; Bell, R.O.: Halogen-doped cadmium telluride for detection of gamma-rays. In: *Nat. Phys. Sci.* 237 (1972), p. 13. doi: 10.1038/physci237013a0

[967] Walenta, A.H.; Fischer, J.; Okuno, H.; Wang, C.L.: Measurement of the ionization loss in the region of relativistic rise for noble and molecular gases. In: *Nucl. Inst. Meth.* 161 (1979), p. 45. doi: 10.1016/0029-554X(79)90360-4

[968] Walenta, A.H.; Heintze, J.; Schuerlein, B.: The multiwire drift chamber, a new type of proportional wire chamber. In: *Nucl. Inst. Meth.* 92 (1971), p. 373. doi: 10.1016/0029-554X(71)90413-7

[969] Ward, M.L.; Greenwood, P.E.: 1/f noise. In: *Scholarpedia* 2 (2007), p. 1537. doi: 10.4249/scholarpedia.1537. – revision #137265

[970] Watanabe, S. et al.: High energy resolution hard X-ray and gamma-ray imagers using CdTe diode devices. In: *IEEE Trans. Nucl. Sci.* 56 (2009), p. 777. doi: 10.1109/TNS.2008.2008806

[971] Wayne, M.R.: Visible light photon counters and the D0 scintillating fiber tracker. In: *Nucl. Inst. Meth.* 387 (1997), p. 278. doi: 10.1016/S0168-9002(96)01007-8

[972] Weber, S. G.; Andronic, A.: ALICE event display of a Pb–Pb collision at 2.76 A TeV. ALICE-PHO-GEN-2015-004. 2015. https://cds.cern.ch/record/2032743

[973] Weekes, T.C. et al.: Observation of TeV gamma rays from the Crab nebula

using the atmospheric Cerenkov imaging technique. In: *Astropphys. J.* 342 (1989), p. 379. doi: 10.1086/167599

[974] WEIGEL, U.: *Cold Intense Electron Beams from Gallium Arsenide Photocathodes*, Universität Heidelberg, Diss., 2003. doi: 10.11588/heidok.00004513

[975] WEILAND, T. et al.: The Finite Integration Technique and the MAFIA™ software package. http://www.temf.tu-darmstadt.de/forschung_5/fitmafia/fit.de.jsp

[976] WEISSTEIN, E.: Wolfram MathWorld. http://mathworld.wolfram.com/

[977] WERMES, N.: Trends in pixel detectors: Tracking and imaging. In: *IEEE Trans. Nucl. Sci.* 51 (2004), p. 1006. doi: 10.1109/TNS.2004.829438

[978] WESTCOTT, C.H.: A study of expected loss rates in the counting of particles from pulsed sources. In: *Proc. Roy. Soc. A* 194 (1948), p. 508. doi: 10.1098/rspa.1948.0094

[979] WIGMANS, R.: On the energy resolution of uranium and other hadron calorimeters. In: *Nucl. Inst. Meth. A* 259 (1987), p. 389. doi: 10.1016/0168-9002(87)90823-0

[980] WIGMANS, R.: High resolution hadronic calorimetry. In: *Nucl. Inst. Meth. A* 265 (1988), p. 273. doi: 10.1016/0168-9002(88)91081-9

[981] WIGMANS, R.: Advances in hadron calorimetry. In: *Annu. Rev. Nucl. Part. Sci.* 41 (1991), p. 133. doi: 10.1146/annurev.ns.41.120191.001025

[982] WIGMANS, R.: *Calorimetry*. Oxford Science Publications, 2000.

[983] WIGMANS, R.: The DREAM project – towards the ultimate in calorimetry. Proc.: *11th Pisa Meeting on Advanced Detectors*. In: *Nucl. Inst. Meth. A* 617 (2010), p. 129. doi: 10.1016/j.nima.2009.09.118

[984] WIKIPEDIA: 65-nanometer process. https://en.wikipedia.org/wiki/65-nanometer_process, retrieved: 2018

[985] WIKIPEDIA: CMOS. https://en.wikipedia.org/wiki/CMOS, retrieved: 2018

[986] WIKIPEDIA: Diamond structure. https://en.wikipedia.org/wiki/Diamond_cubic, retrieved: 2014

[987] WIKIPEDIA: Ethernet. https://en.wikipedia.org/wiki/Ethernet, retrieved: 1.8.2014

[988] WIKIPEDIA: Field-effect transistor. https://en.wikipedia.org/wiki/Field-effect_transistor, retrieved: 2014

[989] WIKIPEDIA: International Standard Atmosphere. https://en.wikipedia.org/wiki/International_Standard_Atmosphere, retrieved: Jan. 2020

[990] WIKIPEDIA: Law of mass action. https://en.wikipedia.org/wiki/Law_of_mass_action, retrieved: 2014

[991] WIKIPEDIA: MOSFET. https://en.wikipedia.org/wiki/MOSFET, retrieved: 2014

[992] WIKIPEDIA: Zincblende structure. https://en.wikipedia.org/wiki/Cubic_crystal_system#Zincblende_structure, retrieved: 2014

[993] WILKINSON, D.H.: A stable ninety-nine channel pulse amplitude analyser for slow counting. In: *Proc. Cambridge Phil. Soc.* 46(3) (1950), p. 508.

[994] WILLE, K.: *The Physics of Particle Accelerators: An Introduction*. Oxford University Press, 2000.

[995] WILLEKE, F.J.: Experiences with the HERA Lepton-Proton Collider. Proc.: *4th Particle Accelerator Conference (APAC 2007)*. RRCAT, Indore, India, (2007), p. 842. http://accelconf.web.cern.ch/AccelConf/a07/PAPERS/FRYMA01.PD

F
[996] Williams, E.J.; Terroux, F.R.: Investigation of the passage of "fast" β-particles through gases. In: *Proc. Roy. Soc. A* 126 (1930), p. 289. doi: 10.1098/rspa.1930.0008
[997] Williams, M.C.S.: Particle identification using time of flight. In: *J. Phys. G* 39 (2012), p. 123001. doi: 10.1088/0954-3899/39/12/123001
[998] Wilson, C.T.R.: Nobel Lecture: On the Cloud Method of Making Visible Ions and the Tracks of Ionizing Particles. http://www.nobelprize.org/nobel_prizes/physics/laureates/1927/wilson-lecture.html
[999] Wilson, E.J.N.: *An Introduction to Particle Accelerators*. Oxford University Press, 2001. doi: 10.1093/acprof:oso/9780198508298.001.0001
[1000] Wink, R. et al.: The miniaturized proportional counter HD-2 (Fe)/(Si) for the GALLEX solar neutrino experiment. In: *Nucl. Inst. Meth. A* 329 (1993), p. 541. doi: 10.1016/0168-9002(93)91289-Y
[1001] Winn, D.R.; Worstell, W.A.: Compensating hadron calorimeters with Cherenkov light. In: *IEEE Trans. Nucl. Sci.* 36 (1989), p. 334. doi: 10.1109/23.34459
[1002] Winokur, P. S.: Radiation-induced interface traps. In: Ma, T.P. (ed.); Dressendorfer, P.V. (ed.): *Ionizing Radiation Effects in MOS Devices and Circuits*. Wiley, 1989, p. 193.
[1003] Witherell, M.S.: The PEP-II B factory and the BABAR detector. In: *Nuovo Cim.* A109 (1996), p. 1061. doi: 10.1007/BF02823647
[1004] Wiza, J.L.: Microchannel plate detectors. In: *Nucl. Inst. Meth.* 162 (1979), p. 587. doi: 10.1016/0029-554X(79)90734-1
[1005] Wizemann, F.; Gisen, A.; Kröninger, K.; Weingarten, J.: Temperature scaling of reverse current generated in proton irradiated silicon bulk. In: *JINST* 14 (2019), p. P07008. doi: 10.1088/1748-0221/14/07/p07008
[1006] Wright, A. G.: *The Photomultiplier Pandbook*. Oxford University Press, 2017. doi: 10.1093/oso/9780199565092.001.0001
[1007] Wu, J. et al.: The performance of the TOFr tray in STAR. In: *Nucl. Inst. Meth. A* 538 (2005), p. 243. doi: 10.1016/j.nima.2004.08.105
[1008] Wunstorf, R.: *Systematische Untersuchungen zur Strahlenresistenz von Silizium-Detektoren für die Verwendung in Hohenergiephysikexperimenten*, Universität Hamburg, Diss., 1992. http://www-library.desy.de/preparch/desy/int_rep/fh1k-92-01.pdf
[1009] Xu, Y.C. et al.: A new beam profile diagnostic system based on the industrial Ethernet. Proc.: *1st International Particle Accelerator Conference (IPAC10)*, Kyoto, Japan, (2010), p. 1044. http://accelconf.web.cern.ch/AccelConf/IPAC10/papers/mope033.pdf
[1010] Yajima, K. et al.: Measurements of cosmic-ray neutron energy spectra from thermal to 15 MeV with Bonner Ball neutron detector in aircraft. In: *J. Nucl. Sci. Technol.* 47 (2010), p. 31. doi: 10.1080/18811248.2010.9711934
[1011] Yamamoto, A.; Makida, Y.: Advances in superconducting magnets for high energy and astroparticle physics. In: *Nucl. Inst. Meth. A* 494 (2002), p. 255. doi: 10.1016/S0168-9002(02)01477-8
[1012] Yamamoto, H.: dE/dx particle identification for collider detectors. Proc.: *4th Workshop physics and experiments with future linear e^+e^- colliders (LCWS'99)*, Sitges, Spain, vol. 2 (1999), p. 1022. *arXiv:hep-ex/9912024*.
[1013] Yasuda, H.: New insights into aging phenomena from plasma chemistry. In:

Nucl. Inst. Meth. A 515 (2003), p. 15. doi: 10.1016/j.nima.2003.08.125

[1014] Yates, E.C.; Crandall, D.G.: Decay times of commercial organic scintillators. In: *IEEE Trans. Nucl. Sci.* 13/3 (1966), p. 153. doi: 10.1109/TNS.1966.4324093

[1015] Ypsilantis, T.; Seguinot, J.: Theory of ring imaging Cherenkov counters. In: *Nucl. Inst. Meth. A* 343 (1994), p. 30. doi: 10.1016/0168-9002(94)90532-0

[1016] Ypsilantis, T.; Seguinot, J.: Development of ring imaging Cherenkov counters for particle identification. In: *NATO Sci. Ser. B* 352 (1996), p. 551. doi: 10.1007/978-1-4613-1147-8_29

[1017] Yuan, L. C. L.; Wang, C. L.; Uto, H.; Pruenster, S.: Formation-zone effect in transition radiation due to ultrarelativistic particles. In: *Phys. Rev. Lett.* 25 (1970), p. 1513. doi: 10.1103/PhysRevLett.25.1513

[1018] Zatsepin, G.T.; Kuzmin, V.A.: Upper limit of the spectrum of cosmic rays. In: *JETP Lett.* 4 (1966), p. 78.

[1019] Zeuner, T. (HERA-B Collaboration): The MSGC-GEM inner tracker for HERA-B. In: *Nucl. Inst. Meth. A* 446 (2000), p. 324. doi: 10.1016/S0168-9002(00)00042-5

[1020] ZEUS Collaboration; Holm, U. (ed.): The ZEUS Detector. Version: 1993, status report (unpublished). http://www-zeus.desy.de/bluebook/bluebook.html

[1021] Zhao, S.: *Characterization of the Electrical Properties of Polycrystalline Diamond Films*, Ohio State University, Columbus, OH, Diss., 1994. https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1394810346

[1022] Zhu, R.-Y.: Radiation Damage in Scintillating Crystals. In: *Nucl. Inst. Meth. A* 413 (1998), p. 297. doi: 10.1016/S0168-9002(98)00498-7

[1023] Zhu, R.-Y.: The next generation of crystal detectors. Proc.: *Hard X-Ray, Gamma-Ray, and Neutron Detector Physics XVII*. SPIE. In: *Proc. SPIE Int. Soc. Opt. Eng.* 9593 (2015), p. 84. doi: 10.1117/12.2190459

[1024] Zlotnick, F.: A Comparison of LVDS, CMOS, and ECL. ON Semiconductor Application Note AND8059/D. 2001. https://www.onsemi.com/pub/Collateral/AND8059-D.PDF

[1025] Zorn, C.: Fast scintillators for high radiation levels: II Plastic and liquid organic scintillators. In: *[850]*, p. 218. doi: 10.1142/9789814360333_0004

[1026] Zuber, K.: *Neutrino Physics*. 2nd ed. CRC Press, 2011. doi: 10.1201/b11065

Abbreviations

In this list only multiply occurring abbreviations are included. The page number refers to where the respective abbreviation is explained for the first time. Names of experiments or facilities are not included in the abbreviation list, but appear in the index.

AC	alternating current 142
ADC	analog-to-digital converter 749
ANSI	American National Standards Institute 803
APD	avalanche photo diode; photodiode with avalanche amplification 424
APS	acoustic phonon scattering 120
ASIC	application specific integrated circuit; electronic chip for custom applications 738
ATCA	advanced telecommunications computing architecture; readout system 804
bcc	body-centred cubic 262
c.o.g.	centre of gravity 833
CAM	content addressable memory; associative memory 813
CAMAC	computer automated measurement and control; electronic readout system and standard 803
CB	conduction band; energetically highest band in semiconductors 263
CC	charged current 691
CCD	charge collection distance 354
CCD	charge-coupled device 323
CCE	charge collection efficiency 354
CMB	cosmic microwave background 699
CMOS	complementary metal–oxide semiconductor; an integrated circuit technology 759
CNGS	CERN Neutrinos to Gran Sasso 166
CSA	charge-sensitive amplifier; amplifier converting charge into voltage 722
CSDA	continuous slowing down approximation; approximation used in range calculation programs like ESTAR 52
CT	computed tomography (with X-rays) 19
Cz	Czochralski; a method of crystal growth 261
DAC	digital-to-analog converter 753
DAQ	data acquisition 799
DC	direct-current (e. g. DC voltage) 10
DEPFET	depleted field-effect transistor; a special pixel detector 328
DIRC	detection of internally reflected Cherenkov light; Cherenkov detector of the BaBar experiment 472
DKFZ	Deutsches Krebsforschungszentrum; cancer research centre in Germany 52
DLC	diamond-like carbon; technique used e.g. for coating 217
DLTS	deep level transient spectroscopy; technique to identify energy levels of semiconductor defects. 360
DM	dark matter 704
DMAPS	depleted monolithic active pixel sensor; depleted MAPS 333
DME	dimethyl ether 215
DOM	digital optical module 703
dSiPM	digital SiPM; a variant of the SiPM 434
DSP	digital signal processor 813
DTL	diode–transistor logic; logic standard used in early integrated circuits 758
e/h	electron–hole pair 148
EAS	extensive air showers 678
EC	electron capture 820
ECC	emulsion cloud chamber 165

ECL	emitter coupled logic; logic standard used in integrated circuits 759
ENC	equivalent noise charge; noise charge equivalent to a signal of 1 e 794
ENOB	effective number of bits; a measure of the dynamic range of an ADC 750
ESONE	European Standards on Nuclear Electronics 803
FADC	flash analog-to-digital converter 232
fcc	face-centred cubic 262
FIFO	first-in-first-out; shift register 813
FOV	field of view, e.g. of telescopes 667
FPGA	field programmable gate array; freely programmable logic circuitry 813
FWHM	full width at half maximum 42
FZ	float zone process process; a method of crystal growth 260
GCD	gate controlled diode; MOS controlled diode structure to measure oxide charges 296
GEM	gas electron multiplier 217
GSI	Gesellschaft für Schwerionenforschung; research centre, Germany 51
HAPD	hybrid avalanche photodiode 427
HPD	hybrid photodiode 427
HV	high voltage 142
IACT	imaging atmospheric Cherenkov telescopes 686
IC	integrated circuit 297
IC	inverse Compton effect 81, 758
IEEE	Institute of Electrical and Electronics Engineers 803
IIS	ionised impurity scattering 120
InGRID	integrated grid; micro-grid fabricated on top of a chip 220
IP	interaction point 377
ISA	international standard atmosphere 668
JES	jet energy scale 648
LAN	local area network 805
LAr	liquid argon 610
LDF	lateral distribution function 677
LKr	liquid krypton 610
LPM	Landau–Pomeranchuk–Migdal effect 590
LS	least squares method 841
LSB	least significant bit; lowest value bit 749
LUT	look-up table 813
LVCMOS	low voltage CMOS; CMOS technology operating with low voltage levels 759
LVDS	low voltage differential signalling; electronic standard for signal transmission 760
MA	main amplifier 213
MAPS	monolithic active pixel sensor; a special pixel detector type 333
micromegas	MICRO-MEsh GAseous Structure 219
mip	minimum-ionising particle 31
MOS	metal–oxide semiconductor; semiconductor interface 274
MOSFET	MOS field-effect transistor 297
MOST	MOS transistor 297
MPGDs	micro pattern gas detector 214
mpv	most probable value of a distribution 42
MRPC	multigap resistive plate chamber; wire chamber with resistive anodes and multiple planes 551
MSGC	microstrip gas chamber 215
MVA	multivariate analysis; a statistical analysis technique 581
mwe	meter water equivalent; thickness of an equivalent water layer in metres 65
MWPC	multiwire proportional chamber 204
NC	neutral current 691
NIEL	non-ionising energy loss; energy lost to lattice displacement damage 344
NIM	nuclear instrumentation module; standard for modular electronics for signal and trigger processing 803
NIRS	National Institute of Radiological Sciences, Japan 52

NIS	neutral impurity scattering 120
NKG	Nishimura–Kamata–Greisen; NKG function 676
NMOS	MOS transistor with electrons as channel current carriers 297
NMR	nuclear magnetic resonance 380
NTP	normal temperature and pressure 21
OpAmp	operational amplifier 719
PA	preamplifier 213
PCI	peripheral component interconnect; bus standard 805
PDG	Particle Data Group; issuer of the 'Review of Particle Properties' 13
PET	positron emission tomography 19
PFA	particle flow analysis 646
PLD	programmable logic device 802
PMOS	MOS transistor with holes as channel current carriers 297
PMT	photo multiplier tube; photomultiplier 413
POT	protons on target 698
PWC	proportional wire-chamber 610
RAM	random addressable memory 813
RBE	relative biological effectiveness 819
RICH	ring imaging Cherenkov detector; Cherenkov ring detector 448
rms	root-mean-square; square root of the mean squares of a set of numbers 68
RoI	region-of-interest 801
RPC	resistive plate chambers 197
RTL	resistor–transistor logic; logic standard used in early integrated circuits 758
SDR	space–drift-time relation 233
SEM	scanning electron microscope 165
SI	Système International d'Unités, English: International System of Units 20
SiDC	silicon drift chamber 320
SiPM	silicon photomultiplier; a pixel photon detector 429
SNR	signal-to-noise ratio 301, 720
SNU	solar neutrino unit 692
SOI	silicon-on-insulator; sensor fabrication technology 337
SPAD	single photon avalanche diode; APD sensitive to single photons 426
SPECT	single-photon-emission computed tomography 19
SPT	superconducting phase-transition thermometer 708
SRH	Shockley–Read–Hall; model framework for radiation damage description 346
SSM	standard solar model 696
STP	standard temperature and pressure 20
TDAQ	trigger and data acquisition 816
TDC	time-to-digital converter 232
TEA	triethylamine; photon absorbing molecule vapour 460
TEM	transmission electron microscopy; method to resolve the chemical composition of semiconductor defects. 360
TES	transition edge sensor 708
TMAE	tetrakis-(dimethylamino)-ethylene; photon absorbing molecule vapour 460
TMS	tetramethylsilane 617
TOF	time-of-flight; measurement of the flight time 548
ToT	time-over-threshold; method to measure the amplitude of a pulse 727
TPC	time projection chamber 241
TR	transition radiation 480
TRD	transition radiation detector 490
TSC	thermally stimulated current; technique to identify energy levels of semiconductor defects. 360
TTL	transistor–transistor logic; logic standard used in integrated circuits 758
TWP	twisted-pair cable; twisted conductor leads as twin signal conductor 761
VB	valence band; next to highest energy band in semiconductors, usually filled with electrons 263
VEM	vertical equivalent muon 680
VHE	very high energy 686

VLSI	very large scale integration; very densely integrated electronic circuit 738
VME	Versa Module Eurocard; readout system 803
VPT	vacuum phototriodes 614
w.e.	water equivalent 679
WIMP	weakly interacting massive particle; candidate for dark matter 704
WLS	wavelength shifter; scintillating dye emitting at longer wavelength than absorbing 507

Index