Springer Tracts in Modern Physics 275

Frank Hartmann

Evolution of Silicon Sensor Technology in Particle Physics

Second Edition

Springer

Springer Tracts in Modern Physics

Volume 275

Series editors

Yan Chen, Department of Physics, Fudan University, Shanghai, China
Atsushi Fujimori, Department of Physics, University of Tokyo, Tokyo, Japan
Thomas Müller, Institut für Experimentelle Kernphysik, Universität Karlsruhe, Karlsruhe, Germany
William C. Stwalley, Storrs, CT, USA

Springer Tracts in Modern Physics provides comprehensive and critical reviews of topics of current interest in physics. The following fields are emphasized:

- Elementary Particle Physics
- Condensed Matter Physics
- Light Matter Interaction
- Atomic and Molecular Physics
- Complex Systems
- Fundamental Astrophysics

Suitable reviews of other fields can also be accepted. The Editors encourage prospective authors to correspond with them in advance of submitting a manuscript. For reviews of topics belonging to the above mentioned fields, they should address the responsible Editor as listed in "Contact the Editors".

More information about this series at http://www.springer.com/series/426

Frank Hartmann

Evolution of Silicon Sensor Technology in Particle Physics

Second Edition

Springer

Frank Hartmann
Karlsruhe Institute of Technology (KIT)
Eggenstein-Leopoldshafen
Germany

ISSN 0081-3869 ISSN 1615-0430 (electronic)
Springer Tracts in Modern Physics
ISBN 978-3-319-64434-9 ISBN 978-3-319-64436-3 (eBook)
DOI 10.1007/978-3-319-64436-3

Library of Congress Control Number: 2017949127

1st edition: © Springer-Verlag Berlin Heidelberg 2009
2nd edition: © Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

In the post era of the Z and W discovery, after the observation of jets at UA1 and UA2 at CERN, John Ellis visioned at a HEP conference at Lake Tahoe, California, in 1983 "To proceed with high energy particle physics, one has to tag the flavour of the quarks!"

This statement reflects the need for a highly precise tracking device, being able to resolve secondary and tertiary vertices within high-particle densities. Since the distance between the primary interaction point and the secondary vertex is proportional to the lifetime of the participating particle, it is an excellent quantity to identify particle flavor in a very fast and precise way. In colliding beam experiments, this method was applied especially to tag the presence of b quarks within particle jets. It was first introduced in the DELPHI experiment at LEP, but soon followed by all collider experiments to date. The long expected t quark discovery was possible mainly with the help of the CDF silicon vertex tracker, providing the b quark information. In the beginning of the twenty-first century, the new LHC experiments were taking shape. CMS with its $206\,m^2$ of silicon area is perfectly suited to cope with the high-luminosity environment. Even larger detectors are envisioned for the far future, like the SiLC project for the International Linear Collider. Silicon sensors matured from small 1 inch single-sided devices to large 6 in double-sided, double metal detectors and to 6 (8) inch single-sided radiation hard sensors. A large group of researchers inside the high energy physics community is steadily developing and investigating new devices and is pushing the technology to new limits. These larger and larger devices are the driving force to improve industrial processing and quality control. Also the readout electronics evolved from microsecond to nanosecond integration times and to radiation-tolerant sub-micron technology devices.

This review describes the successes in high energy physics as well as the developments in technology from the early days of NA11 to the current detector CMS at the LHC and the design of the future one at High-Luminosity Large Hadron Collider HL-LHC, ending with an outlook for detectors in an early design phase for the future linear collider and the Future Circular Collider. The day-to-day life in a

silicon laboratory and the practical handling and testing strategies and R&D tools are also described. Particle Physics examples will be given to underline the importance of silicon tracking devices for high energy physics.

In the second edition, the aspects about radiation are discussed in greater detail of radiation studies, defect engineering, and concepts for very high radiation levels. This edition also introduces measurement tools for sensor R&D. More details about tracking systems are given, how they are operated and how their data is being treated. Some sections demonstrate how dedicated silicon sensors are finding their way into large high granular calorimeters and ultra-fast timing detectors.

Karlsruhe Frank Hartmann
June 2017

Acknowledgements

I truly thank all my colleagues from DELPHI, CDF, CMS, and RD50 of the last two decades and more for either supporting me or making my life interesting! I am well aware that unfortunately I am not able to credit every person individually for all their invaluable impact on all the projects. The author lists of DELPHI, CDF, CMS, and RD50 are simply too long. I apologize if your name does not explicitly appear where it should.

I am especially grateful to Wim de Boer and Thomas Müller, Joe Incandela, and Luigi Rolandi who early on supported me so much and who always had such confidence in me.

I like to thank RD50 and especially Michael Moll for the great work they are doing towards the understanding of radiation effects in silicon, for sharing it with everybody and for giving a lively platform for all experts to work together. And for allowing me to present their work here.

With great pleasure I also acknowledge

Hans Dijkstra, Paula Collins, and Gunnar Maehlum for insights during my undergraduate thesis and for the DELPHI structures which I was able to borrow – the source of many photos here;

Pia Steck, Tobias Barvich, Thomas Eichhorn and Christian Barth from Karlsruhe who relentlessly made many photos and drawings, now published here;

Duccio Abbaneo, Nicola Bacchetta, Marta Baselga, Thomas Bergauer, Gaelle Boudoul, Gino Bolla†, Peter Blüm, Erik Butz, Gianluigi Casse, Alexander Dierlamm, Guido Dirkes, Diana Fellner, Jean-Charles Fontaine, Karl Gill, Oliver Holme, Joe Incandela, Will Johns, Gregor Kramberger, Manfred Krammer, Jacek Marczewski, Marcello Mannelli, Stefano Mersi, Michael Moll, Thomas Müller, Daniel Münstermann Steve Nahn, Anca Ioana Patru, Marcel Stanitzki, Werner Riegler, Robert Stringer, Marco Rovere, Matthias Schröder, Jorgen Christiansen, Francois Vasez, Nicolo Cartiglia, Ivan Vila, Katja Klein and Marcos Fernandez Garcia; without their meticulous proof reading and invaluable comments, publishing this monograph would not have been impossible;

The German Bundesministerium für Bildung und Forschung BMBF, the Deutsche Forschungsgemeinschaft and the Land of Baden-Württemberg making the described research possible by financially supporting it;

The people dearest to me: Dana and Gwen, my daughter Sara and my mother Marianne who endured many hours of absence either due to my "detector obligations" or writing this monograph.

Several schematics are inspired by Bergmann and Schäfer, Sze, from Hyper-script of H. Föll and lectures of Carl Haber, Alan Honma, Paula Collins, Manfred Krammer, Gregor Kramberger, and Michael Moll.

Missing even a single one of the dear people mentioned above, I never would have been able to finish this monograph.

Frank Hartmann

Contents

List of Figures

Chapter 1
Basic Principles of a Silicon Detector

This chapter introduces the basic silicon properties and their technical application to set the scene and provide understanding of the silicon sensors functionality. The writing concentrates on examples of detectors used in particle physics experiments – in the **H**igh **E**nergy **P**hysics HEP. It also describes the working principle of silicon sensors as particle detectors, together with an explanation of their production processes and design parameter considerations. A lot of teasers and links to information about detectors of the last and future 40 years are scattered in this chapter to underline the introduced parameters with real life examples.

The examples are taken from **DE**tector with **L**epton, **P**hoton and **H**adron **I**dentification DELPHI (Chap. 4) at the former **L**arge **E**lectron **P**ositron collider LEP at CERN in Geneva, the **C**ollider **D**etector at **F**ermilab CDF residing at the TEVATRON collider at Fermilab near Chicago (Chap. 5), the **C**ompact **M**uon **S**olenoid CMS experiment (Chap. 6) at the **L**arge **H**adron **C**ollider LHC at CERN and it HL-LHC upgrade (Chap. 7) including the future high granularity silicon-based calorimeter (Sect. 7.2) and finally the conceptual design considerations for the detectors at the future **I**nternational **L**inear **C**ollider ILC and the **F**uture **C**ircular **C**ollider FCC (Chap. 8). The chapter also briefly introduces silicon non-strip sensors. Also the R&D methods DLTS, TSC and TCT are being introduced. It concludes with a quality assurance description and a walk-through prominent historic sensor failures.

1.1 Fundamental Silicon Properties

The Bible teaches us “In the beginning God created heaven and earth”. Today, we still do not know for sure what heaven is made of, but earth consists of, at least the upper crust, silicon and oxygen with some dirt (in the form of the other 90 elements) thrown in for added value.

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_1

Silicon, the element which revolutionized the development of electronics, is known as an important and multi-useable material, dominating today's technology. Its properties have been thoroughly investigated and are well known. Silicon is used in solar cells, computers and telecommunication systems. Initially, silicon comes from the Latin word *silex; silicis*, meaning "flint". The physical properties of silicon can be described as a hard, dark-grey solid with a bluish tinge. At ordinary temperatures, silicon is impervious to air. At high temperatures, silicon reacts with oxygen, forming an inert layer of silica. Silicon is important in plant and animal life. Diatoms in both fresh and salt water extract silica from the water to use as a component of their cell walls. Silicon does not occur in the free, elemental state, but is found in the form of silicon dioxide and complex silicates. Jöns Jacob Berzelius is generally credited with the discovery of silicon in 1824. Henri Deville prepared crystalline silicon in 1854, a second allotropic form of the element. Silicon is a semiconductor and a solid, which isolates at low temperatures and shows a measurable conductivity at higher temperatures. The specific conductivity of 10^{3}–$10^{-4}\Omega^{-1}\mathrm{cm}^{-1}$ lies somewhere between metals and insulators. Since the development of quantum mechanics the electric conductivity can be explained with the covalent bond in the crystal lattice. Bound electrons can be excited by inducing energy above threshold energy (gap energy), e.g. energy from temperature, light, x-ray, β-particles. Since the 1960s semiconductors have been used as particle detectors. Initially, they were operated in fixed-target experiments as calorimeters and as detectors with a high-precision track reconstruction. Since the late 1980s they have been widely used in collider experiments as silicon microstrip or silicon pixel detectors near the primary interaction point. Silicon sensors have a very good intrinsic energy resolution: for every 3.6 eV released by a particle crossing the medium, one electron–hole pair is produced. Compared to about 30 eV required to ionise a gas molecule in a gaseous detector, one gets 10 times the number of particles in silicon. The average energy loss of a minimum ionising particle in silicon is 390 eV/μm, creating 108 (electron–hole pairs)/μm; these values being high due to the high density of silicon.

The usefulness and success of silicon can be explained in a handful of keywords:

- existence in abundance
- favourable energy band gap
- possibility to engineer the gap properties by deliberate addition of certain impurity atoms (dopants)
- the existence of a natural oxide

Since these topics are exhaustively discussed in standard literature like [169, 265, 283], this book keeps the discussion on basic silicon properties, such as band gap, doping and *pn*-junctions brief. On the other hand, concepts and formulas necessary for the understanding of the sensor functionality are presented as well as an overview of "standard" sensor designs and the reasons for certain ranges of parameters.

Table 1.1 Silicon properties

Parameter	Symbol	Unit	Value
Atomic number			14
Relative atomic weight			28.0855
Structure			Diamond
Lattice constant	a_0	Å	5.4307
Lattice orientation			⟨111⟩
Electron configuration:			$1s^2 2s^2 2p^6 3s^2 3p^2$
Density	ρ_m	gcm^{-3}	2.328
Melting point	T_m	°C	1414
Boiling point	T_b	°C	2355
Gap energy (300 K)/(0 K)	E_g	eV	(1.124)/(1.170)
Dielectric constant	ε_r		11.7
Intrinsic carrier density	n_i	cm^{-3}	$1.45 \cdot 10^{-10}$
Mobility			
– of the electrons	μ_e	$cm^2\ [Vs]^{-1}$	1350
– of the holes	μ_h	$cm^2\ [Vs]^{-1}$	450
Effective density of states			
– of the conductance band	N_c	cm^{-3}	$3.22 \cdot 10^{19}$
– of the valence band	N_v	cm^{-3}	$1.83 \cdot 10^{19}$
Max. electric field	E_{max}	$V\mu m^{-1}$	30
Thermal expansion coefficient		1/°C	$2.5 \cdot 10^{-6}$
Intrinsic resistivity	ρ	kΩcm	235

1.1.1 Just Silicon and Some Impurities

Quantitative properties of silicon are shown in Table 1.1. Silicon, as every semiconductor, has a forbidden region in the energy band structure, the band gap. At low temperature and in the absence of impurities ("intrinsic silicon") the valence band is full and the conduction band is empty. Without any impurities, the concentration of electrons n in the conduction band and holes[1] p in the valence band are equal to the intrinsic concentration $n_i = n = p$. This is also reflected in the global charge neutrality at equilibrium. The mechanism to alter conductivity behaviour is to insert additional states in the forbidden region to increase the probability to excite electrons or holes in the Fermi–Dirac sense (see Fig. 1.1). The technical expression is called "doping".

Silicon is a type IV material (four valence electrons), it can be doped with impurities to alter its free charge carrier concentration. One can produce "n-type" silicon

[1]Concept of "holes": if an electron absorbs $E \geq \Delta E_g$ (E_{Gap}) it can enter the conduction band leaving a vacancy called "hole" representing a positive charge in the valence band which can move in an applied electric field. These holes are treated like particles and follow the Fermi–Dirac statistic.

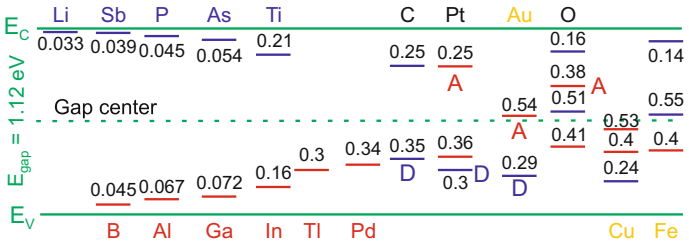


Fig. 1.1 The numerous impurities establish additional levels in the forbidden zone, e.g. boron near the valence band decreases the energy needed for hole excitation from 1.12 to 0.045 eV. Useful dopants add levels near the conduction or valence band. The A stands for acceptor, the D for donor. The levels introduced in the mid-gap region such as those from gold or copper are undesired and only increase leakage current but are not useful as attributed dopants

by adding type V material, e.g. phosphorus (donor impurity; excess of electrons as majority charge carriers); "p-type" material can be realized by adding type III material, e.g. boron (acceptor impurities; excess of holes as majority charge carriers).

Typical concentration levels are

- Si atoms $5 \cdot 10^{22}\,\mathrm{cm}^{-3}$
- $n_i = 1.45 \cdot 10^{10}\,\mathrm{cm}^{-3}$ at 27°C
- HEP silicon sensor bulk $10^{12}\,\mathrm{cm}^{-3}$
- HEP strip/pixel/p^+-stop implant doping $10^{15} - 10^{19}\,\mathrm{cm}^{-3}$
  - on the higher end for latest generations
- HEP dedicated contacts $10^{20}\,\mathrm{cm}^{-3}$
- light doping (IC industry) $10^{16}\,\mathrm{cm}^{-3}$
- heavy doping (IC industry) $10^{19}\,\mathrm{cm}^{-3}$

In thermal equilibrium the probability for occupancy of one state at the energy E for the temperature T is given – taking into account the Pauli principle – by the Fermi–Dirac statistics

$$f(E,T) = \frac{1}{e^{(E-E_F)/k_B T} + 1} \tag{1.1}$$

where k_B is the Boltzmann constant and E_F the Fermi energy[2] or Fermi level.

$$f(E = E_F) = \frac{1}{2} \tag{1.2}$$

E_F can be regarded as the energy where exactly half of the available levels are occupied. With this knowledge, the location in the middle of the band gap for intrinsic silicon and in-between the additional levels and the nearest band becomes obvious (see Fig. 1.1). The important point to remember is that the Fermi level is shifted

[2]Fermi energy: "the highest possible energy of a fermion at $T = 0$ K".

depending on the impurities. The actual value of E_F level is essential, whenever a contact of two materials exists, because it defines naturally the relative levels of all the other bands (see Fig. 1.4).

In thermal equilibrium, the Fermi energy is constant everywhere. The final number of free charge carriers is calculated as

$$n = \int_0^\infty D(E) \cdot f(E,T) \cdot dE \tag{1.3}$$

with $D(E)$ as the density of states, a pure quantum mechanical distribution counting all possible energy levels up to the energy E.

The technologically important point is that for all practical temperatures, the majority charge carrier concentration is identical to the dopant concentration, visible in the location of the corresponding E_F (see Fig. 1.1). The densities of electrons n and holes p then derive to

$$n = N_C e^{-\frac{E_C - E_F}{k_B T}} \quad \text{with} \quad N_C = 2\left(\frac{2\pi m_e^* k_B T}{h^2}\right)^{\frac{3}{2}} \tag{1.4}$$

and

$$p = N_V e^{-\frac{E_F - E_V}{k_B T}} \quad \text{with} \quad N_V = 2\left(\frac{2\pi m_h^* k_B T}{h^2}\right)^{\frac{3}{2}} \tag{1.5}$$

with N_C for conduction band and N_V as effective state density in the valence band. E_C, E_F and E_V are the energies of the condition band, Fermi level and valence band. The m_e^* and m_h^* stand for the effective masses of electrons and holes, respectively, h is the Planck constant and the factor 2 is derived from the two possible spin states of the electrons. Obviously the charge neutrality still holds, but free electrons find their hole counterpart in the additional doping levels and no longer in the lower valence band. The famous mass action law holds for intrinsic as well as doped material:

$$n \cdot p = n_i^2 = N_C \cdot N_V e^{-\frac{E_g}{k_B T}} \tag{1.6}$$

where $E_g = E_C - E_V$.

In ultrapure silicon at room temperature the intrinsic carrier concentration is $n_i = 1.45 \cdot 10^{10}\,\mathrm{cm}^{-3}$ and with approximately $10^{22}\,\mathrm{cm}^{-3}$ only about 1 in 10^{12} silicon atoms is ionised.

To finally understand the electric conduction mechanism, one has to know about mobility μ and drift velocity $v_D = \mu E$. It is not difficult to convince oneself that conduction depends on the number of free charges, their ability to move and their "motivation" to move. Translating this into a mathematical formula, the conductivity σ and resistivity ρ of doped silicon material comes to

$$\sigma = e(\mu_e n + \mu_h p) \text{ and therefore } \rho = \frac{1}{e(\mu_e n + \mu_h p)} \tag{1.7}$$

where e is the electron charge, μ_e and μ_h are the mobilities of electrons and holes, respectively, with mobility $\mu = \frac{v_D}{E}$ (E the electric field, v_D the drift velocity), n and p are the densities of electrons and holes, respectively. For the pure intrinsic silicon case at T = 300 k we get

$$\rho = \frac{1}{1.6 \cdot 10^{-19}\text{C}\left(1350\frac{\text{cm}^2}{Vs} \cdot 1.45 \cdot 10^{10}\text{cm}^{-3} + 450\frac{\text{cm}^2}{Vs}1.45 \cdot 10^{10}\text{cm}^{-3}\right)} \approx 235\ \text{k}\Omega\text{cm} \tag{1.8}$$

Microscopically, mobility μ is given by

$$\mu_{e,h} = \frac{e\tau_s}{m^*_{e,h}} \tag{1.9}$$

with τ_s the time between two scattering processes

1. at crystal defects, like dislocations or undesired impurities – this effect is not dominant before irradiation, but is the dominant part after
2. at intentionally introduced impurities, namely the doping atoms
3. at phonons, the thermally stimulated lattice vibrations

Since the mobility μ drops with increasing dopant concentration, the conductivity σ does not increase linearly with doping concentration. The mobilities in silicon are quite high and therefore suited for the use as HEP detectors. The mobility for electrons and holes is $\mu_e = 1350\,\text{cm}^2/\text{Vs}$ and $\mu_h = 450\,\text{cm}^2/\text{Vs}$, resulting in about 10 ps/μm with E > $5 \cdot 10^4$ V/cm or below 10 ns collection/readout time in silicon sensors of several 100 μm thickness.

It is worthwhile to note that at high fields the velocity is not any more proportional to the electric field; or said differently the mobility is not constant any more [283]. At very high fields the velocity even saturates, at around $E = 10^7$ V/cm for Silicon at T=300 k, which is an interesting property of the base material. Linearity is given for holes up to $E \approx 2 \cdot 10^4$ V/cm but already ends around $E \approx 7 \cdot 10^3$ V/cm for electrons in high-purity silicon materials. The effect is therefore close to negligible for sensors used before the era of the LHC and even then only lightly relevant for the pixel sensors collecting electrons. An example of hole and electron mobilities and their dependence on field and doping concentrations in a strip sensor is presented later in Fig. 1.18.

A very important factor for silicon sensors is the base resistivity ρ. In material dominated by one type of impurity, e.g. the donor dopant density N_d is much larger than the intrinsic carrier concentration, the following expression for the resistivity ρ is valid:

$$\rho = \frac{1}{e(\mu N_d)} \tag{1.10}$$

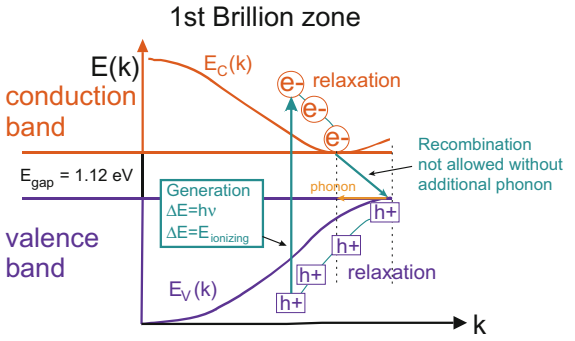


Fig. 1.2 Electron–hole pair generation by a photon in silicon. To obey energy and momentum conservation, recombination is only possible with phonon exchange. Recombination is basically impossible and thus lifetimes τ_L in ideal semiconductors without impurities can range up to milliseconds

This is a very important parameter, which has to be carefully chosen for sensors. It will be shown later that it is inversely proportional to the minimal detector operation voltage. Generally, high resistivity (5 – 10 kΩcm) material is preferred, but low[3] resistivity (1 – 3 kΩcm) material was chosen for the CMS experiment at the LHC – details and reason, see Sect. 6.4.1.

So far, the number of free charge carriers and the neutrality of the solid object are considered only in a static way. In the real device, we have a dynamic equilibrium. Electron–hole pairs are generated and recombine all the time, only the average concentration remains constant.

Figure 1.2 is more educational than representative for a real device. In an ideal indirect[4] semiconductor (like silicon) an exited electron (hole) cannot recombine on its own. A phonon is needed to simultaneously conserve energy and momentum. In a real device the recombination rate is completely dominated by second order effects – additional levels in the forbidden gap (Fig. 1.3). The lifetime τ_L is completely defined by the impurities, like interstitial atoms, such as Fe, Ni, Cu, Au, the dopant atoms, crystal dislocation or grain boundaries:

$$\tau_L = \frac{1}{\sigma v_{th} N_t} \tag{1.11}$$

with v_{th} as the thermal velocity ($\approx 10^7$ cm/s), the charge carrier cross-section $\sigma(\approx 10^{-15}\,\text{cm}^2)$ and impurity/trap concentration N_t.

The important message from the Shockley–Read–Hall theory (schematically displayed in Fig. 1.3) is the role played by the impurities, acting not only as traps but also as generation centres for electron-hole pairs. The direct band to band ($E_C \leftrightarrow E_V$) generation or recombination is technically negligible. Dislocations or precipitates in the active device regions or point defects in the silicon crystal and especially metal impurities like Cu, Au, Fe, Cr, etc. introduce new 'deep' levels inside the band gap.

[3] Not to compare with low resistivity IC silicon material with 10 – 100 Ωcm.

[4] The lowest conduction band energy does not lie at the same position in k-space as the highest valence band energy.

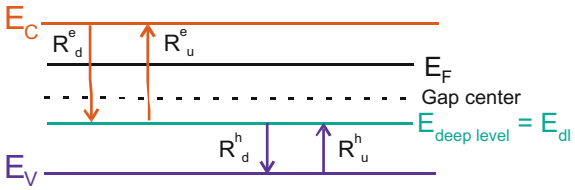


Fig. 1.3 Shockley–Read–Hall in a nutshell: The Shockley–Read–Hall theory (Shockley and Read [270] and Hall [126]) describes in detail the second order effects, which in reality are the most relevant mechanisms. Defects take the function of step stones to generate or recombine free charge carriers. Depending on the placement of the levels, reflecting different species of atoms, lifetimes τ_L range from ns to ms. R^e_d: transit rate of electrons from conduction band to deep level; electron capture
R^e_u: transit rate of electrons from deep level to conduction band; electron emission
R^h_u: transit rate of holes from valence band to deep level; hole capture
R^h_d: transit rate of holes from deep level to valence band; hole emission

The deep levels are also often called generation-recombination centres or briefer recombination centres or also intermediate-level states. The mechanism leads to a reduction of minority carrier lifetime and thus influences device characteristics.

The emission/capture[5] rates are primarily defined by the deep level cross-section $\sigma_{e,h}$, the energy difference to the valence/conduction band and the fraction of deep level states occupied by electrons n_{dl} (holes p_{dl}) in the first place, naturally dependent on temperature. In steady state/thermal equilibrium[6] we have $R_u = R_d$ (rf. Fig. 1.3). Therefore, similar to formulas 1.4 and 1.5, the emission rates R^e_u and R^h_d can be quantified as

$$R^{e,h} = (n,\, p)_{dl} \cdot c_{e,h} \cdot N_{C,V} \cdot e^{\pm \frac{E_{dl}-E_{C,V}}{k_B T}} = (n,\, p)_{dl} \cdot \sigma_{e,h} \cdot v_{th;e,h} \cdot N_{C,V} \cdot e^{-\frac{E_{a;e,h}}{k_B T}} \tag{1.12}$$

with the activation energy $E_{a;e,h} \propto E_{dl} - E_{C,V}$, the capture coefficient $c_{e,h}$, the deep level cross-section $\sigma_{e,h}$ and the thermal velocity $v_{th;e,h}$.

Shortly, the influence of τ_L on the intrinsic leakage current of silicon sensors will be introduced.

1.1.2 The pn*-Junction*

In an intrinsic silicon substrate – the size of a standard silicon sensor used in high-energy physics – there are $\sim 10^9$ free charge carriers but *only* $\sim 2 \cdot 10^4$ generated electrons induced by an ionising particle. The resulting signal would be lost in the

[5] The capture rate is similar, but here also the concentration of free electrons (holes) in the conduction (valence) band to "feed" them are relevant.

[6] According to the principle of *detailed balance* the concentrations of free electrons in the conduction band and free holes in the valence band have to be constant.

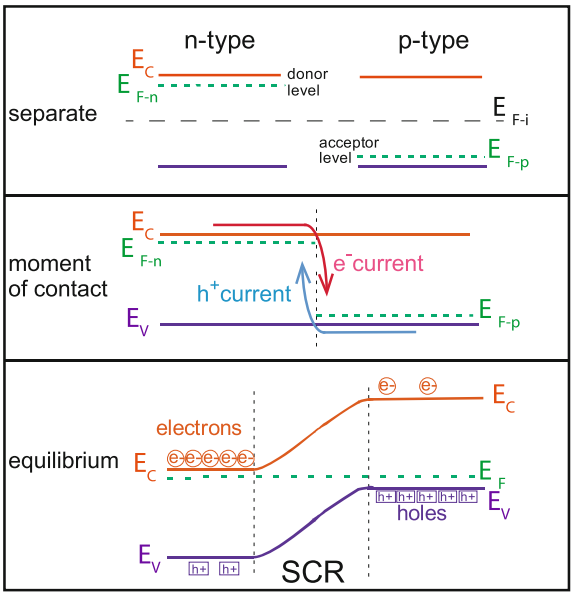


Fig. 1.4 How does p-type silicon behave when brought into contact with its neighbour n-type. By joining p- and n-type together electrons move to the lower Fermi levels and holes to the higher, building up a space charge region SCR, where in equilibrium, the Fermi energy E_F is constant everywhere. To visualize it, first draw band diagrams for the p and n region with their defined Fermi levels E_{F-p}, E_{F-n} separately. The dashed line depicts where the Fermi level E_{F-i} would be in an intrinsic sensor. Second, join the two parts, the electrons move to the material with the lower Fermi energy, while the opposite is true for the holes. Last, in the state of equilibrium, a space charge is built up and the potentials are shifted accordingly, with the Fermi energy constant everywhere

number of free charge carriers. Therefore, the free charge carriers have to be reduced by several orders of magnitude. This could be achieved by cooling to very low temperatures or by depleting the silicon volume of free charge carriers, using p- and n-type silicon in a *reverse-biased pn-junction* configuration.

Cryogenic cooling would be highly unpractical for large detector applications, therefore the basic diode properties of a pn-junction are exploited. The scope here only describes a pn-junction in equilibrium and the actual use of high-doped material in reverse bias at intermediate temperatures.

The principal concept of depleting a semiconductor volume, thus creating a **s**pace **c**harge **r**egion SCR, is visualized in Fig. 1.4. As discussed earlier the Fermi energy in thermal equilibrium must be constant. Nature establishes the equilibrium by the movement of the free majority charge carriers into the opposite regions (electrons move to the p and holes vice versa to the n-region) leaving behind ionised charged dopant atoms. The diffusion current $j_{\text{diffusion}}$ is described by Fick's first law $j_{\text{diffusion}} = -e\text{D}\nabla n$, equalizing the carrier density according to its gradient ∇n with the diffusion constant D.

Some basic math is needed to achieve a "feeling" for the device. At the pn-junction diffusion and recombination produces a space charge layer, creating an electric field $\boldsymbol{E}$ and preventing further diffusion. The mathematical expressions are visualized in Fig. 1.5. A dynamic equilibrium is created: diffusion flow and field current of both charge carriers are compensating each other at the pn-junction. The Poisson equation describes the electrostatic potential $\phi(x)$:

$$\frac{\partial^2 \phi}{\partial x^2} = -\frac{1}{\varepsilon_0 \varepsilon_{\mathrm{Si}}} \varrho(x) \tag{1.13}$$

with $\varepsilon_{\mathrm{Si}}$ as relative dielectric constant os silicon in the SCR region. Assuming complete ionisation the **charge density** $\varrho(x)$ (Fig. 1.5d), with the impurity densities N_A and N_D (Fig. 1.5b) (acceptor and donor, respectively) plus the mobile charge densities $n(x)$ and $p(x)$ (Fig. 1.5c), is described by

$$\varrho(x) = -q\big[n(x) - p(x) + N_A - N_D\big] \tag{1.14}$$

In the very localized contact region of n- and p-type silicon, the free charges compensate for the charges of ionised and uncompensated impurities.

In Fig. 1.5d the depleted boundary layer is drawn. This leads to the negatively (positively) charged x_p (x_n) region in the p (n) doped volume summing up to the space charge region[7] w:

$$w = x_p - x_n \tag{1.15}$$

Integration of the Poisson equation including the depletion approximation[8] leads to a linear behaviour of the **electric field strength** $|E|$ (Fig. 1.5e) n-type region $-x_n \leq x \leq 0$ and p-type region $0 \leq x \leq x_p$, respectively, are

$$|E_n(x)| = +\frac{qN_D}{\varepsilon_0 \varepsilon_{\mathrm{Si}}}(x + x_n); \quad |E_p(x)| = +\frac{qN_A}{\varepsilon_0 \varepsilon_{\mathrm{Si}}}(x - x_p) \tag{1.16}$$

Integrating twice leads to the parabolic behaviour of the **potential** $\phi(x)$ (Fig. 1.5f) with the boundary condition $\phi(x = 0) = 0$

The potentials for the n-type region for $-x_n \leq x \leq 0$ and p-type region $0 \leq x \leq x_p$, respectively, are then described by

$$\phi_n(x) = -\frac{1}{2}|E_{max}| \cdot x_n \cdot \left[\left(\frac{x}{x_n}\right)^2 + 2\frac{x}{x_n}\right]; \quad \phi_p(x) = +\frac{1}{2}|E_{max}| \cdot x_p \cdot \left[\left(\frac{x}{x_p}\right)^2 - 2\frac{x}{x_p}\right] \tag{1.17}$$

The total difference of potential in the space charge region gives the **diffusion or built-in voltage** $V_{\mathrm{diffusion}}$

[7] Depleted of free charge carriers.

[8] Assuming an "abrupt" change of $\varrho(x)$ (see Fig. 1.5d – full line).

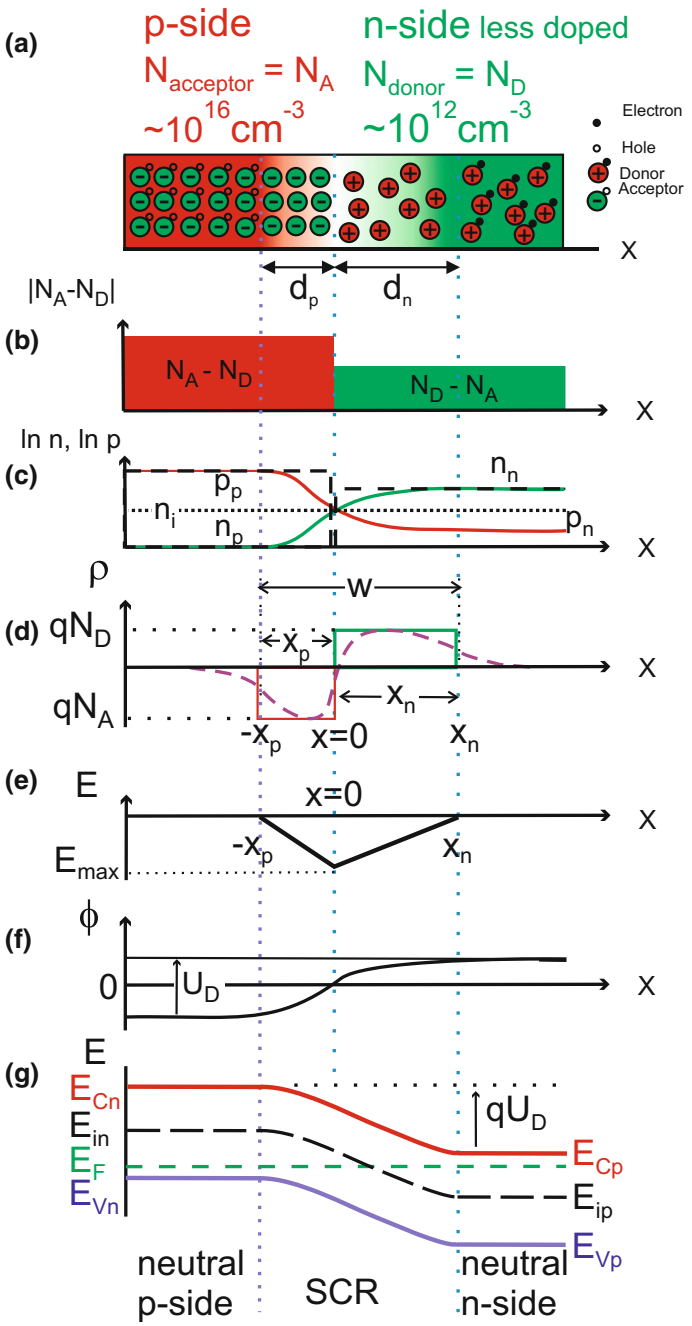


Fig. 1.5 These diagrams display (**a**) A simple visualization of the atomic and charge configuration. (**b**) The doping profile. (**c**) The mobile charge density. (**d**) The space charge density. (**e**) The electric field configuration. (**f**) The electric potential. (**g**) Electron energy across the *pn*-junction. All states are depicting the equilibrium state, without any external voltage

$$V_{\text{diffusion}} = \phi_p(+x_p) - \phi_n(-x_n) = \frac{1}{2}|E_{max}|w \overset{eq.\ 1.10,\ 1.16}{=} \frac{1}{2\mu\rho\varepsilon}w^2 \quad (1.18)$$

with mobility μ, resistivity ρ and $\varepsilon = \varepsilon_0\varepsilon_{\text{Si}}$. From another viewpoint the potentials ϕ_p and ϕ_n are defined by the differences of intrinsic and extrinsic Fermi energy levels and their carrier concentrations (see also Figs. 1.4 and 1.5):

$$E_{F-n} - E_{F-i} = e\phi_n = k_B T \ln \frac{N_D}{n_i}; \quad E_{F-i} - E_{F-p} = e\phi_p = -k_B T \ln \frac{N_A}{n_i} \quad (1.19)$$

resulting in

$$V_{\text{diffusion}} = \phi_n - \phi_p = \frac{k_B T}{e}\left(\ln \frac{N_D}{n_i} + \ln \frac{N_A}{n_i}\right) = \frac{k_B T}{e} \ln \frac{N_A \cdot N_D}{n_i^2} \quad (1.20)$$

As an example with a p doping concentration of $N_A = 10^{15}\text{cm}^{-3}$ and $N_D = 10^{12}\text{cm}^{-3}$ for the n region the widths would be $x_p = 0.02\,\mu\text{m}$ and $x_n = 23\,\mu\text{m}$ respectively and with

$$V_{\text{diffusion}}(T = 300\text{ K}) = \phi_n - \phi_p = \frac{k_B T}{e} \ln \frac{10^{15} \cdot 10^{12}}{(1.45 \cdot 10^{10})^2} \approx 0.4\text{ V} \quad (1.21)$$

The whole system is completely defined by the energy barriers, which are fully defined by the doping concentrations. Intrinsically the generation and recombination of charges are in equilibrium and the system is stable. Creation of large volumes with pure doping concentration difference is technically impossible; $V_{\text{diffusion}}$ is of the order of a few to some hundreds of millivolts with space charge regions of some tens of micrometers. An additional technique is needed to increase the depleted volume.

An **external voltage** $\pm V$ will disturb this equilibrium of spontaneous, generation and recombination of electrons/holes. The external voltage increases or decreases – depending on the polarity – the intrinsic potential barrier of the pn-junction. As a result of the external voltage the depletion width decreases or increases (see Figs. 1.6 and 1.7).

Silicon sensors are operated in reverse bias mode and the forward case will not be considered any further. For the detector case, charge carriers created in the SCR can be collected at the junction, while charge created in the non-depleted zone recombines with free majority carriers or with the generation partner, and is lost. Operation conditions, namely voltage V_{external}, is therefore such that the full volume is depleted. With $V_{\text{external}} = V_{\text{bias}} \gg V_{\text{diffusion}}$ and Eq. (1.18) w results to

$$w = \sqrt{2\varepsilon\mu\rho V_{\text{bias}}} \quad (1.22)$$

and vice versa

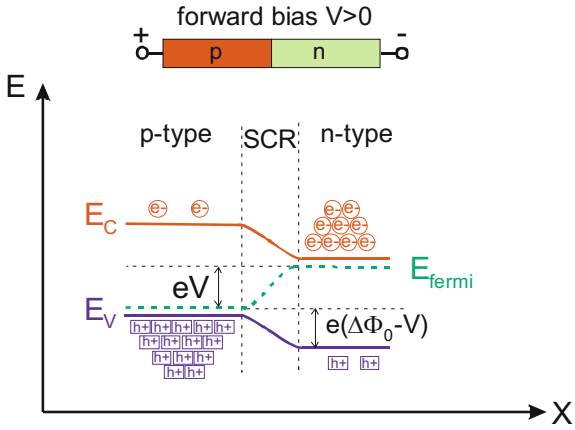


Fig. 1.6 Forward (bias) voltage: In the forward case, the barrier decreases significantly, the majority carriers flow freely through the diode

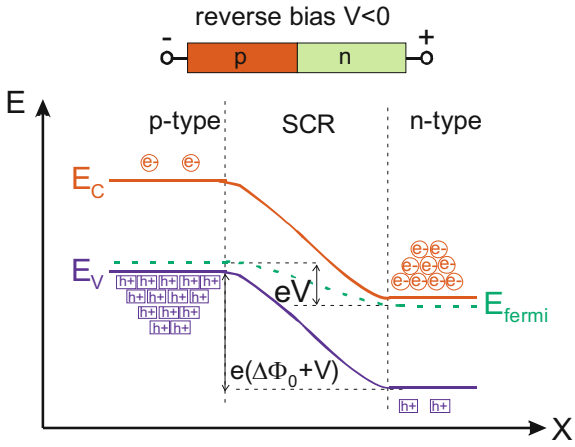


Fig. 1.7 Reverse (bias) voltage In the reverse bias case, the potential barrier as well as the depletion width increases

$$V_{\text{full depletion}} = V_{FD} = \frac{D^2}{2\varepsilon\mu\rho} \tag{1.23}$$

with $w = D$ as the full sensor thickness and resistivity ρ. V_{FD} is one of the most important design parameters, describing the minimal operation value – the voltage the sensor has to sustain without going into current breakdown. As an example, sensors in the inner layers of the CMS tracker are 320 µm thick with a resistivity ρ range of 1.5 – 3 kΩ, the depletion voltage V_{FD} is therefore within

$$V_{FD} = \frac{(0.032)^2\,\text{cm}^2}{2 \cdot (8.85 \cdot 10^{-14}\text{F cm}^{-1}) \cdot 11.7 \cdot (1350\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}) \cdot (1.5-3)\text{k}\Omega} \approx 122 - 244\,\text{V}$$

To run over-depleted the operation voltage V_{bias} is therefore set to 400 V.

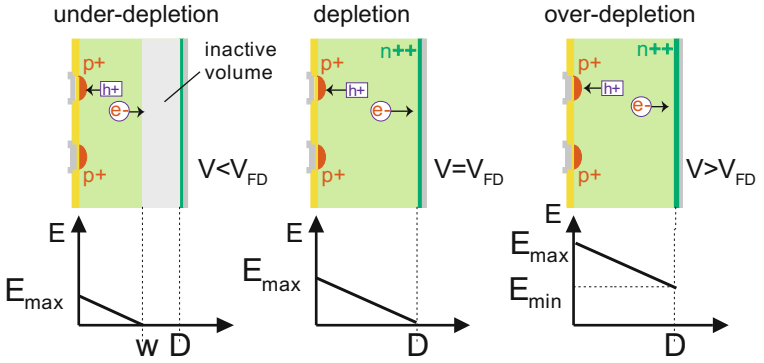


Fig. 1.8 The field in the bulk silicon is linear, depleting a certain volume. The *left* part shows an under-depleted $V_{bias} < V_{FD}$ sensor, while the *right* scheme shows the standard over-depletion $V_{bias} > V_{FD}$ case, the picture in the *middle* depicts full depletion $V_{bias} = V_{FD}$. The max field E_{max} is on the segmented side for *p*-in-*n*, where the strips are on GND potential and the backplane is on high voltage potential

Electric field configurations for under-, full- and over-depletion conditions are shown in Fig. 1.8. The absolute electric field values $|E|$ for the different external voltage configurations are given for different depths x by:

Under-depletion: $V_{\text{bias}} < V_{FD}$

$$E(x) = \frac{2V}{D}\left(1 - \frac{x}{w}\right) \text{ with } E_{max} = \frac{2V_{\text{bias}}}{w} \tag{1.24}$$

Depletion: $V_{\text{bias}} = V_{FD} = \frac{D^2 q_e N_{eff}}{2e}$

$$E(x) = \frac{q_e N}{\varepsilon}(D - x) \text{ with } E_{max} = \frac{2V_{\text{bias}}}{D} \tag{1.25}$$

Over-depletion: $V_{\text{bias}} > V_{FD}$

$$E(x) = \frac{2V_{FD}}{D}\left(1 - \frac{x}{D}\right) + \frac{V_{\text{bias}} - V_{FD}}{D} \text{ with } E_{max/min} = \frac{V_{\text{bias}} \pm V_{FD}}{D} \tag{1.26}$$

With formula (1.26) the maximum electric fields can be, for example, calculated for a sensor used during the LEP era and one from the CMS era. The bias voltage at DELPHI was set to 60 V with an effective full depletion voltage of 40 V on a 300 µm thick sensor resulting in a minimum electric field of 666 V/cm and maximum one of 3.3 kV/cm. With $V_{\text{bias}} = 400$ V and $V_{FD} = 200$ V applied to a 300 µm thick sensor minimum and maximum electric field strengths climb to $E_{min} = 6.6$ kV/cm and $E_{max} = 20$ kV/cm.

With $V_{\text{bias}} > V_{FD}$, the equilibrium is disturbed and an electric field is now established, sweeping the thermally generated electron–hole pairs in the SCR (generation rate inside the space charge region G_{SCR}) out of the depletion region. As mentioned earlier, the emission process is dominated by the Shockley–Read–Hall transitions/emissions. Of course this is now an undesired effect, resulting in a reverse current also called "leakage current" j_R (I_L), described by

$$j_R = e{\cdot}G_{SCR}{\cdot}w = \frac{1}{2}e\frac{n_i}{\tau_L}w = \frac{1}{2}e{\cdot}n_i{\cdot}\sigma{\cdot}v_{th} \cdot N_t{\cdot}w \tag{1.27}$$

and with the surface A of the junction

$$I_L = \frac{1}{2}e\frac{n_i}{\tau_L}w \cdot A = \frac{1}{2}e \cdot n_i \cdot \sigma \cdot v_{th} \cdot N_t \cdot w \cdot A \tag{1.28}$$

In summary, the leakage current is completely dominated by the effective lifetime τ_L (refer to formula (1.11), the generation lifetime of minority carriers), namely the impurity states N_t near mid-gap, e.g. Au and all noble metals are "lifetime killers", see Fig. 1.1. The temperature dependence enters indirectly via $n_i \propto T^2 \cdot e^{-\frac{E_g}{2k_B T}}$.

For example, a CMS sensor has a leakage current of ~0.5 nA/strip at room temperature, where a strip is $A = 100 \times 0.05\,\text{mm}^2$ with a depletion width of 300 μm, the sensor thickness.

$$\tau_L = \frac{1}{2}e\frac{n_i}{I_L}w{\cdot}A = \frac{0.5(1.6 \cdot 10^{-19}\text{C})(1.45 \cdot 10^{10}\,\text{cm}^{-3})(0.03 \cdot 10 \cdot 0.006\,\text{cm}^3)}{0.5 \cdot 10^{-9}\,\text{A}} \sim 4\,\text{ms}$$

The impurity concentration N_t then amounts to

$$N_t = \frac{1}{\sigma v_{th}\tau_L} = \frac{1}{10^7 \cdot 10^{-15} \cdot 4 \cdot 10^{-3}\,\text{cm}^3} = 2.5 \cdot 10^{10}/\text{cm}^3$$

The current increases linearly with $w \propto \sqrt{V}$ until the detector is fully depleted. At higher bias voltages an electric breakdown is observed, where the current starts to increase dramatically. The breakdown can either be explained by "avalanche breakdown", due to charge multiplication in charge collisions with the lattice or by "Zener breakdown", based on the quantum mechanical "tunnel effect". Figure 1.9 shows $I \propto \sqrt{V}$ behaviour, as well as a breakdown. In Chap. 6 a design is introduced, which allows bias voltages above even the maximum electric field of silicon.

The full capacitance of a sensor can be calculated by regarding the two planes of the SCR as plates capacitor with silicon as dielectric inside. The capacitance C decreases linearly with w until the depletion zone reaches through the full sensor thickness $w = D$ and therefore $\sim \sqrt{V}$:

$$C_{\text{bulk}} = \begin{cases} A\sqrt{\frac{\varepsilon_0\varepsilon_{\text{Si}}}{2\rho\mu V_{\text{bias}}}} & V_{\text{bias}} \leq V_{FD} \\ A\frac{\varepsilon_0\varepsilon_{\text{Si}}}{D} = const. & V_{\text{bias}} > V_{FD} \end{cases} \tag{1.29}$$

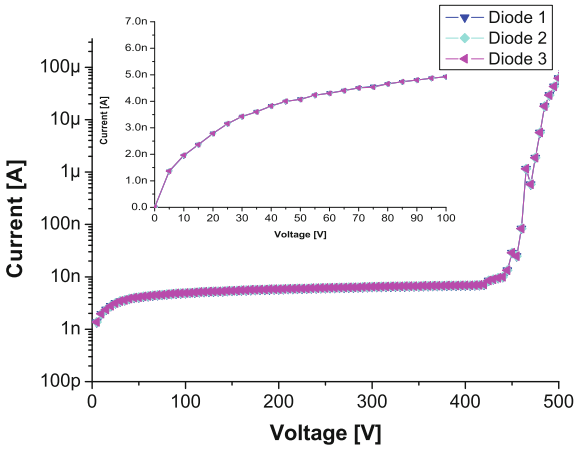


Fig. 1.9 The current–voltage characteristic for a silicon diode in the reverse bias direction is depicted. The expanded view shows the $I \propto \sqrt{V}$ dependence (for $V_{\text{bias}} < V_{FD}$), while the global view shows the full scan including breakdown at higher voltages

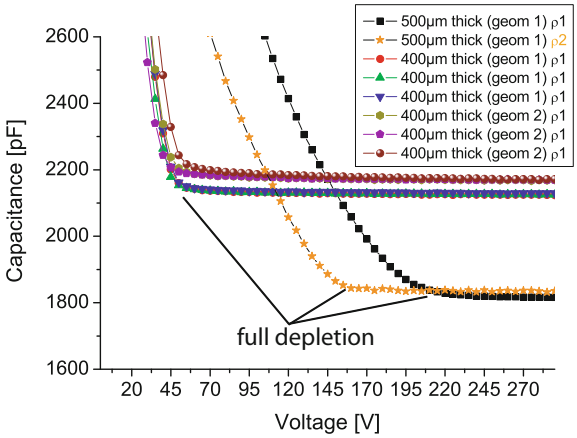


Fig. 1.10 The measurement plots describe the capacitance dependence on area and thickness quite clearly. The x-axis coordinate of the kink shows the depletion voltage, defined by material resistance and thickness. The y-coordinate of the plateau shows the minimal capacitance, defined by area and thickness. The two upper bands depict sensors of two different geometries with slightly different areas and same high resistivity material, both $D = 400$ µm thick. The lower CV curves describe $D = 500$ µm thick sensors. With increased thickness, C becomes smaller and $V_{\text{depletion}} \propto D^2$ becomes larger. The different depletion voltages of the lower curves derive from two different resistivities $\rho 2 > \rho 1$

Figure 1.10 expresses the capacitance C and V_{FD} dependency on area, thickness and resistivity ρ. The **c**apacity–**v**oltage characteristic CV or $1/C^2$ versus voltages behaviour is used as a standard method to determine V_{FD}. The kink determines V_{FD}.

1.1.3 SiO_2

Silicon dioxide SiO_2 is very stable and chemically inert. It is a near-perfect dielectric and one of the best insulators, with an extremely high breakdown field strength. Only the co-existence of Si and SiO_2 defines the real technological value. Pure silicon is simply too chemically reactive, without a thin layer of SiO_2 it will react with anything. The excellent electrical properties of the Si – SiO_2 interface are also very important. This interface has a very low density of energy states (especially for ⟨100⟩ crystal orientation) in the band gap and therefore neither provides recombination centres nor introduces fixed charges. SiO_2 is used as

- gate oxide for transistors
- dielectric in capacitors
- passivation and protection
- structure masking during etching, diffusion and ion implantation
- insulation

For tracking sensors it serves mainly as coupling capacitor oxide and as final passivation. Figure 1.11 shows the passivation of all Si areas leaving metal pads free for connectivity.

The growth, application and structuring of SiO_2 are relatively easy and will be described in Sect. 1.9.2, especially how the selective etching possibilities of Si or SiO_2 allow dedicated processing.

Si – SiO_2 Interface

In modern silicon microstrip sensors, the surface consists of a sandwich of high-doped silicon strips, a SiO_2 insulation layer with metal strips on top to be connected to the readout electronics, called AC coupling. This **m**etal-**o**xide **s**emiconductor forms the classical MOS structure. The subject is widely discussed for example in [194] or in any electronic or solid state physics literature. The basics are discussed with

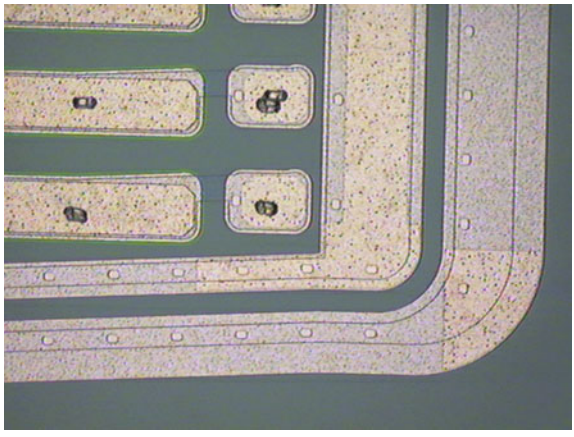


Fig. 1.11 The picture shows a small section of a CMS strip sensor, where the metal pads are "open" for connectivity, while the rest is all covered with SiO_2 for protection

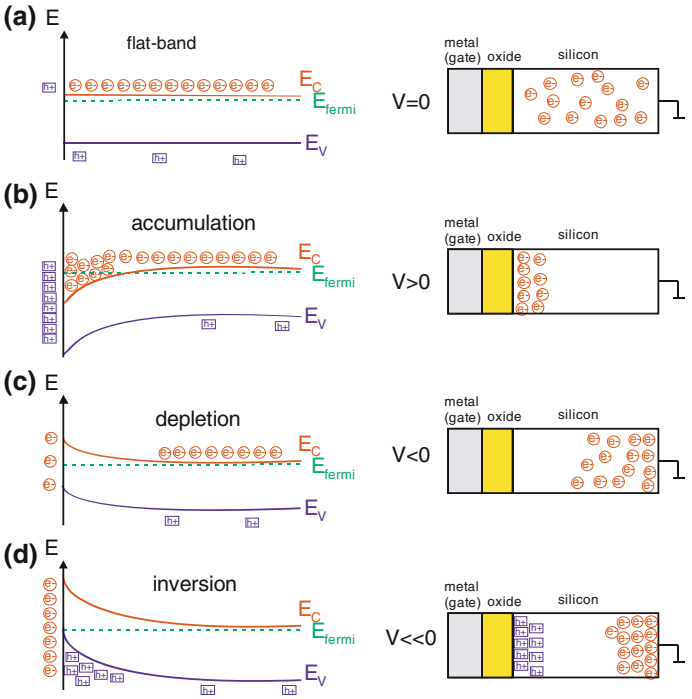


Fig. 1.12 The diagrams show the different optimal cases of a MOS structure: **(a)** Flat-band, **(b)** Accumulation, **(c)** surface depletion and **(d)** inversion. The *left* part of the diagram illustrates schematically the energy bands of the $Si - SiO_2$ interface (not including the metal), taking into account the charge or potential present on the oxide side, independent of their origin. The *right* part displays a more volume-based picture, with a defined metal gate, where the metal potential defines the interface case. In summary the band structure is defined by energy band shifts originating from voltage or fixed charges on the SiO_2 side

an emphasis on the importance of a very pure oxide to have a well-defined deep depleted surface for a good strip isolation. The $Si - SiO_2$ interface condition is defined by the oxide charge or voltage potential applied on the metal part, simply two aspects of the same physical electron/hole attraction or repelling mechanism. The mechanism acts on the majority carriers of the bulk silicon. The basic cases and their potential structures are shown in Fig. 1.12: flat-band condition, accumulation, surface depletion and surface inversion. In Sect. 2.3 the surface property change caused by radiation is discussed.

The SiO_2 layer is considered as a thin surface with a constant homogeneous charge density ϱ, the integrated sum of electrons, holes, ionised doping atoms or undesired contaminations.

As for the *pn*-junction, the Poisson equation almost describes the field and band interface situation, only without any net current in the insulator.

The Fermi level E_F is constant, no charge carriers from the semiconductor can neutralize the charges. The surface charges cause an electric field penetrating into the *Si*-bulk bending the energy bands.

Two majorly different configurations have to be discussed: the surface charge has the same polarity as the bulk majority carriers, the majority charge carriers are driven into the bulk. When a space charge layer builds up, this case is called **surface depletion**, the mandatory case for a HEP sensor. It guarantees the isolation between adjacent strips. Without a possible compensating current, but at a constant E_F the bands are bent. $E_C - E_F$ increases towards the surface. With a very low majority carrier concentration, scaling as $e^{-(E_C - E_F)}$ at the interface, the minority carriers concentration increases due to the mass action law, finally becoming the majority carriers. This configuration is called **surface inversion** and can ruin the strip-to-strip isolation.

In the second configuration, charges from the bulk are attracted to and accumulate at the interface, due to the different polarity in the surface layer compared to the majority carrier in the bulk. The bands are then bent inwards. This case is called **accumulation** and of course also short-circuits neighbouring strips. One prominent example, where an electron accumulation layer forms, is the n^+ face of a double-sided sensor or in an n-in-p sensor (more later). There the n^+ layer attracts electrons at the n^+–n-interface. The technical solution for the strip–strip isolation in this case is presented in Sect. 1.6.3. The **flat-band** configuration is the configuration just in-between depletion (outward bending) and accumulation (inward bending). In this case, the external charge is just compensating the charge due to intrinsic surface states. The potentials/bands in the silicon bulk are flat.

With an additional metal on top of the SiO_2 layer, the situation changes. The work functions of metal and semiconductor are not the same and the flat-band case is now intrinsically impossible, it can only be achieved via an external voltage, the flat-band voltage $V_{\text{flat-band}}$ (often simply V_{FB}). In reality, a low $V_{\text{flat-band}}$ is a quality factor of the oxide and refers to a low contamination level. More globally, an external voltage can always shift between the four configurations:

- **accumulation** ($V > V_{\text{flat-band}}$)
- **flat-band condition** ($V = V_{\text{flat-band}}$)
- **surface depletion** ($V < V_{\text{flat-band}}$)
- **inversion** ($V \ll V_{\text{flat-band}}$)

With all the above, it has to be taken into account that in standard operation of strip sensors, no voltage is applied to the metal strips. Implant and metal are at almost the same potential and only oxide contamination, charge-up or contamination in the overlaying passivation can change the needed surface depletion case. In short, in a strip sensor undesired charges in the oxide can easily decrease the inter-strip resistance by several orders of magnitude. CCDs on the other hand make use of exactly the possibility of the band shifts, attracting and repelling charges to shift from CCD cell to cell. In the IC industry, the MOS structure is the most important means to open or close a conductive channel, e.g. for a **F**ield **E**ffect **T**ransistor FET.

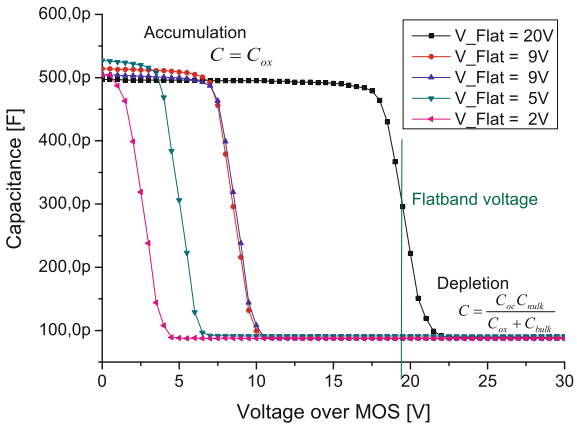


Fig. 1.13 With a capacitance over voltage scan on a MOS structure the point of inflection reveals its flat-band voltage. $V_{\text{flat-band}}$ measures oxide thickness and purity. A low value guarantees a low contamination level. High values arise often from significant amounts of oxide traps and/or interface states in the oxide. Levels for HEP sensors range between $V_{\text{flat-band}} = 1$ and 10 V, where the example of 20 V flat-band voltage belongs to a rejected sensor

Dedicated MOS structures on the sensor wafer allow the measurement of the flat-band voltage and therefore provide a means to qualify oxide quality and have a hint of the oxide thickness. By varying the voltage over a MOS structure, the accumulation configuration changes through flat-band case to depletion. In the accumulation configuration the majority carriers are attracted to the surface and the measured capacitance is $C = C_{ox}$, while in inversion case the serial capacitance of oxide plus bulk silicon $C = C_{ox} C_{\text{bulk}} / (C_{ox} + C_{\text{bulk}})$ is measured. The flat-band voltage marks the inflection of the CV characteristic.

Figure 1.13 displays a CV curve in a MOS measurement with a $V_{\text{flat-band}}$ evaluation of several structures of different quality during the quality assurance campaign of the CMS experiment.

It has to be mentioned that the measurement frequency of the LCR[9] device plays a role. A high measurement frequency is used to avoid any majority or minority charge carrier movement during the individual C measurement.

1.1.4 Summary of Silicon Properties

To summarize, the properties of intrinsic silicon, and the concept of changing electrical properties by adding donor or acceptor atoms were introduced. The connection of p- and n-doped silicon was described, where at equilibrium the Fermi level E_F is constantly forming a potential barrier and a space charge region. The situation so far

[9]Device measuring L inductance, C capacitance and R resistance.

describes a single diode, which in reverse bias mode is devoid of free charge carriers and any created charge from ionisation in the SCR can be collected at the electrodes.

The Si to SiO_2 affinity allows an easy integration of a capacitive coupling of diode to metal contact, thereby allowing the use of a charge amplifying chip.

For a tracking detector further segmentation of these "diodes" is needed. The production of raw silicon wafers is described in Sect. 1.9.1 while the processing steps from wafer to silicon sensor are introduced in Sect. 1.9.2.

1.2 Ingredients to Use Silicon as Detector Basis

All tracking detectors make use of the free charges resulting from the ionisation of a passing charged particle in a medium, e.g. gas or a semiconductor. The average energy loss of a charged particle in a medium is described by the Bethe formula.

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\left(\frac{2m_e c^2\beta^2\gamma^2 T_{max}}{I^2}\right) - \beta^2 - \frac{\delta(\gamma)}{2}\right] \quad (1.30)$$

In this formula z is the charge of the incident particle, T_{max} the maximum kinetic energy which can be imparted to a free electron in a single collision, I the mean excitation energy, Z the atomic number, A the atomic mass, N_A the Avogadro's number, m_e the electron mass, c the speed of light, r_e the classical electron radius, $\beta = v/c$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ and δ density effect correction. The full function of average energy loss of a traversing charged particle is given in Fig. 1.14. The most prominent part is the minimum at approximately $\beta\gamma = 3$ – the minimum deposited energy in the medium. Every detector must be able to keep its noise well below this energy to be able to detect these **M**inimum **I**onizing **P**articles MIPs.

In addition, there are statistical fluctuations, a subject investigated in depth by Landau. The number of collisions in a finite medium as well as the energy transfer per scattering varies. The first effect can be described by a Poisson distribution, while the latter is described by a "straggling function" first deduced by Landau. In rarer cases, called δ-rays or δ-electrons, the transferred energy is large; these δ-electrons are responsible for the asymmetric long tail towards high charge deposits. All in all the **M**ost **P**robable **V**alue MPV of energy transfer is about 30% lower than the average value. For silicon, the average energy used for the creation of one electron–hole pair in the indirect semiconductor is 3.6 eV, about three times larger than the band gap of 1.12 eV, deriving from the fact that part of the deposited energy is used for phonon creation. For a MIP, the most probable number of electron–hole pairs generated in 1 μm of silicon is 76, while the average is 108. A resulting Landau distribution is shown in Fig. 1.15.

Whereas gas detectors utilize charge amplification in electron avalanche clouds by applying high voltages, solid state devices have to cope with the raw signal of

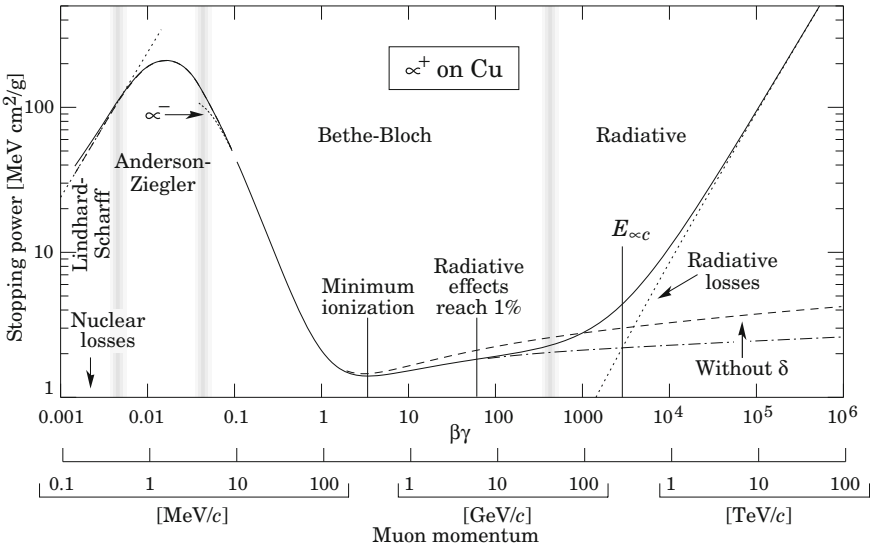


Fig. 1.14 Charged particles lose energy, when traversing material. The figure shows the stopping power $\langle dE/dX \rangle$ of copper for traversing muons [326]. Exactly this effect is the fundamental principle of all ionising detectors. The sensor design needs to make sure to detect the **M**inimum **I**onizing **P**article MIP with a momentum ∼300 MeV/c with a significant signal/noise (S/N) ratio. The plot includes the corrections to the Bethe formula at low and high energies, whose explanations are beyond the scope of this book

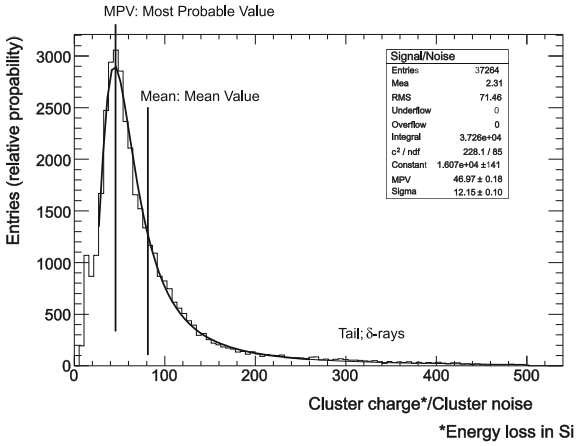


Fig. 1.15 A Landau distribution. The distribution displays the ratio of ionisation signal charge to noise in 500 μm silicon from cosmic particles (MIP) arriving at normal incidence in a 3.8 T field. The distribution shows well the difference between Mean Value and **M**ost **P**robable **V**alue MPV, and the long tail of the distribution from delta-rays is clearly visible [340]

the primary ionisation[10]. The high density ($\rho_m = 2.33\text{g/cm}^3$) of silicon causes an average energy loss of about 390 $\frac{\text{eV}}{\mu\text{m}}$ for a MIP, resulting in the above mentioned 108 $\frac{e-h}{\mu\text{m}}$. In a standard volume of $1 \times 1 \times 0.3\,\text{cm}^3$ at room temperature, there are about $4.5 \cdot 10^8$ free charge carriers versus $3.2 \cdot 10^4$ electron–hole pairs created by a MIP, this explains the need to work with fully depleted sensors. Due to the Landau fluctuations and the most probable signal of 22400 electron–hole pairs $= 0.7 \cdot 32000$, the most probable collected charge is therefore 22400 holes or[11] electrons collected at either side of the detector, equal to 3.6 fC.

1.3 Working Principle of a Silicon Tracking Device

An early basic overview of silicon sensors can be found in [231]. The working principle of a silicon microstrip detector is illustrated in Figs. 1.16 and 1.17. An ionising particle penetrates through a fully depleted silicon n-doped slice. The generated holes drift along the electric field, created by the bias voltage, to the p^+ doped strips[12] while the electrons drift to the n^{++} backplane. The backplane has a higher doping concentration thus lower resistivity to achieve a good ohmic connection 'avoiding' a Schottky contact[13]. Figure 1.8 on p. 15 expresses the need to over-deplete to collect charges, in the p-in-n example holes, from the whole volume and to enforce a fast drift to the electrodes.

The charge induced on the doped strips are then capacitively coupled (AC) to the aluminium readout strips, which are directly (DC) connected[14] to the charge preamplifier of the readout chip. The custom ASIC (application specific integrated circuit) then amplifies and shapes the signal. Also, analogue-to-digital conversion, zero suppression, or baseline correction may be realized here. In principle, the capacitor does not need to be implemented on the silicon sensor, it can also be implemented inside the readout chip or in-between, this was for example the case for the NA11 experiment, see Sect. 3.2 or any pixel sensor deployed in high energy physics. Sensors with

[10](1) DEPEFTs are silicon sensors with intrinsic amplification, they are introduced in Sect. 1.12.6. (2) There are also photo-avalanche diodes and **si**licon **p**hoto-**m**ultiplier SiPM with intrinsic charge amplification, but these are not used as segmented sensors for high energy tracking sensors. (3) LGADs, a recent development, are similar to SiPM and introduced in Sect. 1.12.8.

[11]In the most common sensor p-in-n, holes are collected at the segmented side, but for n-in-p, n-in-n or double-sided sensors electrons are also collected.

[12]In an n-in-n, n-in-p or a double-sided detector, electrons drift to the n^+ doped strips.

[13]Often, highest doping concentration in the backplane, here n^{++}, lower doping for the strips/pixel, here p^+ and lowest doping for the bulk, here n to achieve high resistivity in the bulk and low resistivity elsewhere. The extra high doping for the backplane is not done at every vendor.

[14]Most often by ultrasonic wire-bonding.

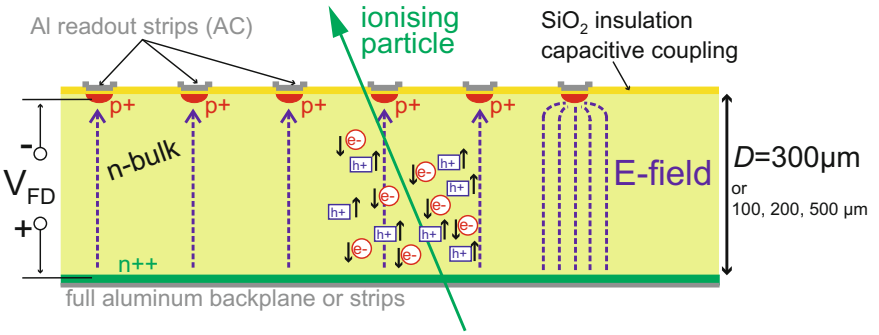


Fig. 1.16 Working principle of a p-in-n AC-coupled silicon microstrip detector. Electron–hole pairs, resulting from the ionisation of the crossing charged particle, generated according to the Bethe formula, travel to the electrodes on the sensor planes guided by the electric field. The segmentation in the pn-junctions allows to collect the charges on a small individual strips, where they capacitively couple to the Al readout strips. These are then connected to the readout electronics, where the intrinsic signal is shaped and amplified. In the case of segmented p^+-strip implants in an n-bulk silicon material, holes are collected at the p^+ strips. The field concentrates on the strips as illustrated on the right end of the figure. Electrons are moving towards the n^{++}-backplane

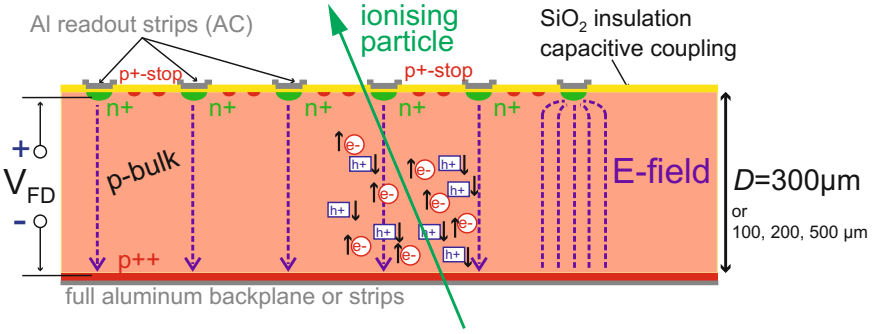


Fig. 1.17 Working principle of an n-in-p AC-coupled silicon microstrip detector. An n-in-p sensor is basically the inversion of a p-in-n sensor plus p^+-stop[15] implants necessary to isolate the individual strips – see later explanation (Sect. 1.3.4). Logically also the bias voltage is reversed. It registers electrons instead of holes on the readout electrode

integrated capacitors are called AC-coupled and otherwise DC-coupled. Because the capacitor value needs to be large, the full strip length consists of a "p^+–oxide–metal sandwich". Examples are given in the DELPHI (Chap. 4), CDF (Chap. 5) and CMS (Chap. 6) experiments.

[15]For a simpler figure we avoid the use of $^{+++}$ and write p^+ for p^+-stop although its doping concentration is lower then for the n^+ strip.

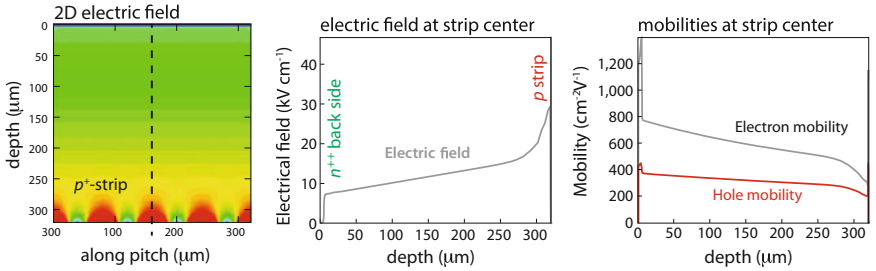


Fig. 1.18 A two-dimensional electric field configuration, emphasizing the discrete strips (*left*). The field linearity across the bulk volume and sharp rise at the strip (*middle*). Mobilities of holes and electrons, dependent on the electric field and impurity/doping concentrations (*right*). Sensor parameters and voltage settings taken from CMS (rf. Chap. 6). Plots adapted from [134]

1.3.1 Charge Collection – An Illustration

For better illustration, we simulated a CMS-like sensor: a $D = 300\ \mu\text{m}$ thick, with strip-to-strip distance (pitch) of 80 µm, a bias voltage of 400 V is used with respect to a depletion voltage of 160 V. The dynamic drift of electrons and holes, after the crossing of a MIP particle at a 45° angle, are then simulated and presented. Figure 1.18 shows the electric field configuration of the sensor and the corresponding electron and hole mobilities. The left plot shows a two-dimensional representation of the electric field with a linear behaviour throughout the bulk of the sensor but highly peaking at the p^+-strips (electrodes), where by design we want to collect the charge. The electric field at the centre of one strip is plotted in the middle figure, again the strong rise at the strip is seen, a linear behaviour for the main part and a drop to zero on the backside where the drop close to the surface reflects the significant thickness of the conductive back-n^{++} layer. The electric field strengths are at values where the mobility is not constant any more, as depicted in the right plot of Fig. 1.18 – see also Sect. 1.1.1. The slope in the middle reflects the change due to electric field strength, the sharp drop/rise at the edges is due to the change of doping concentration and change in field. The electric field guides the drifting holes towards the strip-like electrodes while the electrons drift to the common backplane (electrode).

Figure 1.19 shows the absolute sum of current density, meaning the colour code does not distinguish between electrons and holes. In the simulation a MIP has traversed the sensor at an angle of 45° at time 0°. Already after around 200 ps the electrons and holes are largely separated and the strip near the particle entrance collects holes. The holes collection is distributed to different strips a bit later depending on the drift length. After 4 – 5 ns the electrons with a higher mobility are gone while the hole collection continues. Figure B.1 in the Appendix shows more time slots of the same simulation. For strip and pixel configuration the drift and collection mechanism is the main factor defining the integrated signal; the signal in a p-in-n strip sensor consists mainly of holes (see signal creation in the next section).

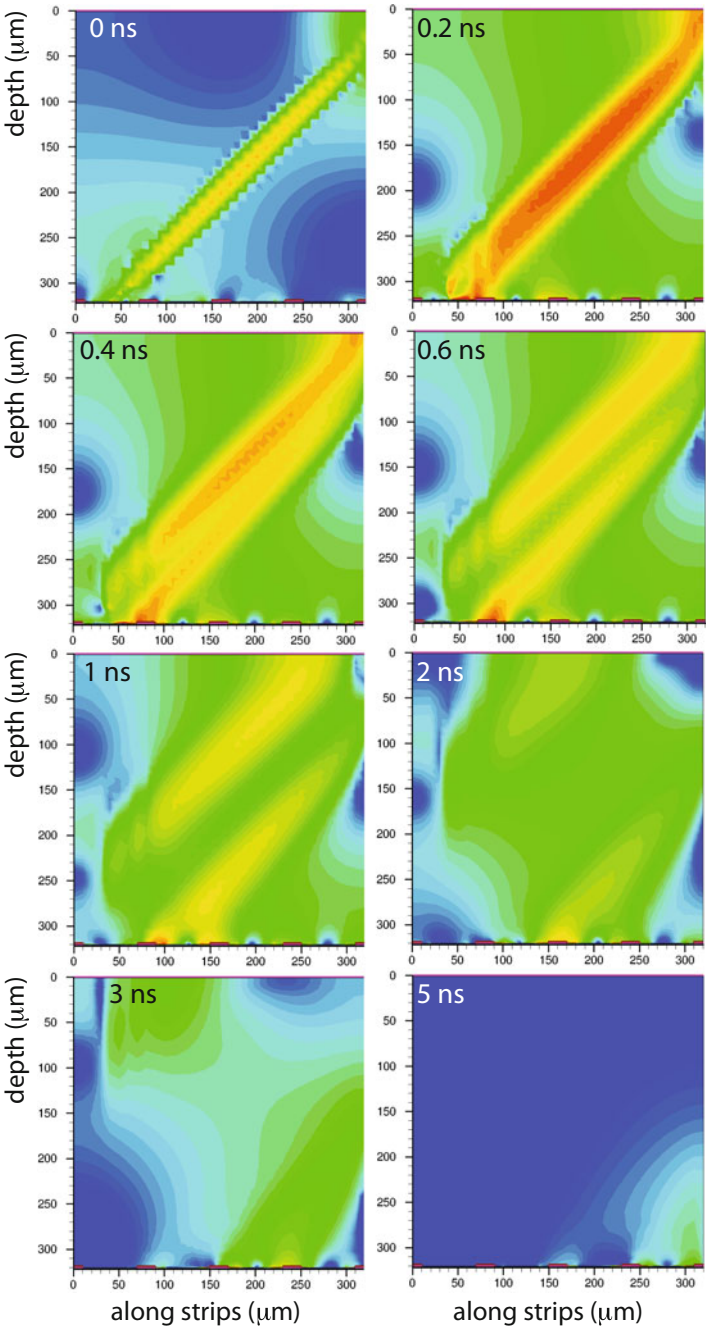


Fig. 1.19 Simulation: An ionising particle traverses the *p*-in-*n* sensor at a 45° angle, disturbing the static situation – same configuration as in Fig. 1.18. The eight plots show the absolute sum of the current densities (electrons and holes) for different times. The strips, collecting holes, are at the *bottom* and the backplane at the *top*. Plots adapted from [134]

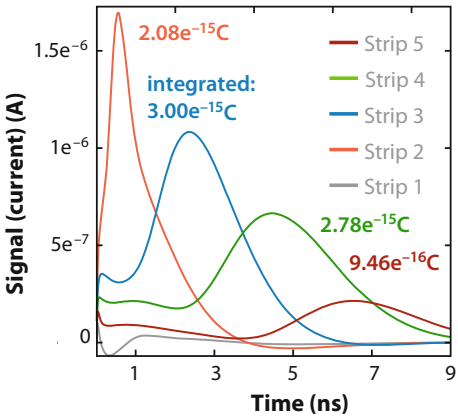


Fig. 1.20 The resulting current of the 45° angle traversing ionising particle, collected at the individual strips in a time-resolved way. The integrated sums are also provided

Figure 1.20 rounds up the example. It displays the resulting charge, collected at the individual strips in a time-resolved way; the integrated sums are also provided.

1.3.2 Signal via Induction – Shockley–Ramo Theorem

In reality, the signal already starts to form before the charges reach the electrodes, namely, due to induction of electric current, when the charges move. As a matter of fact, the signal stops when the charges reach the electrode. With induction, both charge types contribute to the signal at both electrodes (sensor faces) but depending on the geometry with highly different weights – the voltages and charges on all electrodes are related/linked by their electrostatic capacitance matrix. The mechanism can be fully explained by the laws of electrostatics and has initially been discussed by Shockley (1938) and Ramo (Ramo's theorem 1939). An elegant way is to introduce the concept of the weighting field which defines how the charge couples to the respective electrodes. The weighting field is mostly distinctly different to the electric field which defines the charge drift. Mathematically the weighting field is determined by applying unit potential ($U = 1\,\mathrm{V}$) to the measurement electrode and zero to all others and then solving the Poisson equation. It therefore solely depends on the sensor geometry, mainly the electrode geometry and the positioning of the electrodes with respect to each other (readout and backplane); at first order electrode size versus distance. The weighting field does not depend on the movement of the charges. Similar as above, we assume the moving charge is an infinitesimal small electrode with a capacitive coupling to all electrodes, depending on electrode geometries and inversely on distances to electrodes. Said differently, this coupling mechanism is coded into the weighting field/potential and is therefore only dependent on spacial location, for example highest at a strip/pixel (check Figs. 1.21 and 1.23) or weighting field constant for a diode. For more details, especially for the full mathematical

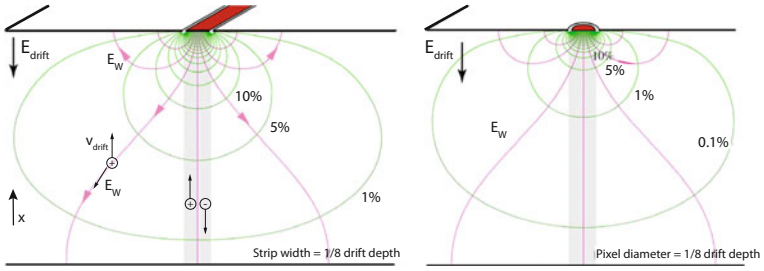


Fig. 1.21 The *lines* indicate the weighting field and weighting potential while the electric field is considered linear for most of the volume (see Fig. 1.18). The *left* figure shows a 2D example – a strip configuration. In the *right* the 3D case of a pixel electrode is shown. As mentioned in the text, to derive the weighting field, the neighbouring electrodes and backplane are set to potential = 0. Figures are taken from [247]

descriptions (reciprocity theorem) and more examples, the reader is referred to [69, 246, 247, 251, 276] and an analytical calculation of some examples (pixel/pads) can be found in [257].

Examples of the weighting field for pixels and strips are shown in Fig. 1.21.

With the knowledge of the weighting field and the charge's velocity as a function of position, driven by the electric (operational) field, the instant induced current can be calculated, while simple integration over the full particle path gives the full induced charge (the signal).

The induced currents/charges can then be described by

$$i = -q\boldsymbol{E}_W \cdot \boldsymbol{v} = q v_s \frac{\partial V_W}{ds} \text{ and } Q_{\text{tot}} = \int_{s_1}^{s_2} i\,dt = q \int_{s_1}^{s_2} \boldsymbol{E}_W d\boldsymbol{s} = q[V_W]_{s_1}^{s_2} \quad (1.31)$$

with $\boldsymbol{v}$ the charge drift velocity driven by the electric field from spacepoint s_1 to s_2, the weighting field $\boldsymbol{E}_W(x, y, z)$, the weighting potential $V_W(x, y, z)$ and the signal (charge) Q_{tot}. The induced charge Q_{tot} is therefore simply given by the difference in the weighting potentials between two positions of the moving charge, independent on path and velocity.

Both electrons and holes induce the same "signed" signal, since with opposite electric charge sign, they also drift in the opposite direction.

First, we discuss the example of a diode to illustrate the dynamic behaviour and then continue with more complex weighting fields/potentials for strips and pixel configurations concentrating on Q_{tot}. In the special case of a two-electrode configuration (with no space charge) the electric field and the weighting field have the same form. For example for a plate capacitor[16] with plate distance D, the electric field is $\boldsymbol{E} = \frac{U}{D}\boldsymbol{e}_d$=constant and the weighting field is $\boldsymbol{E}_W = \frac{1}{D}\boldsymbol{e}_d$=constant (with unit potential U=1 V) and the unity vector $\boldsymbol{e}_d$ the direction pointing from n to p. Subsequently the weighting potential (zero to one) behaves linear (Fig. 1.23(left), diode case). In

[16] Another example would be a cylindrical drift tube.

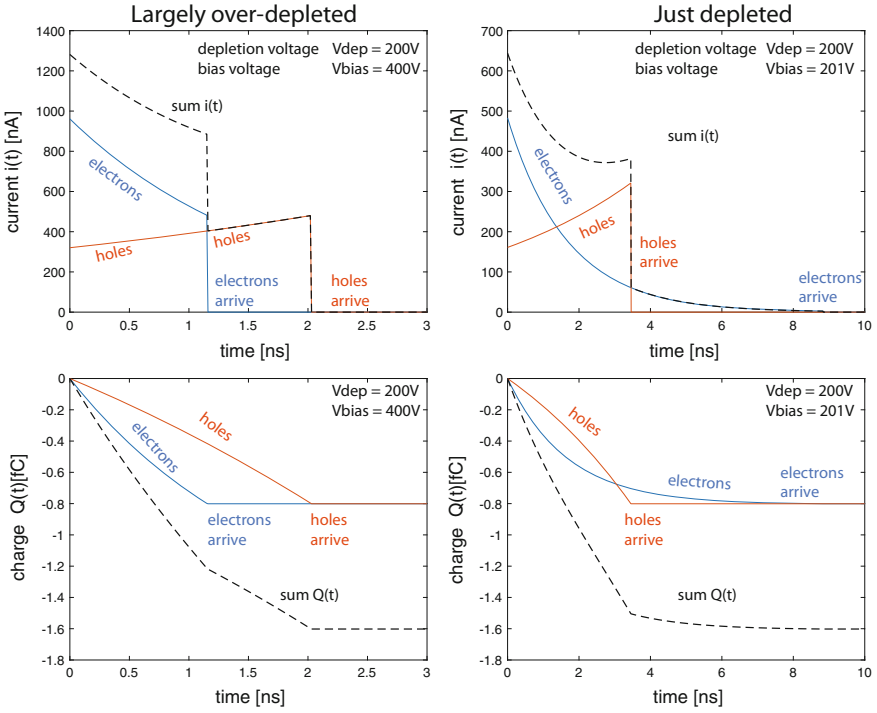


Fig. 1.22 Signal currents and integrated charges of a 300 µm thick diode are plotted in a time resolved manner for different bias voltages. $N = 10000$ electron-hole pairs have been introduced in the middle of the diode bulk. The figure also illustrates the value of over-depletion defining largely signal shape and duration (*left*). With barely the depletion voltage the electric field at the backplane n^{++}-electrode is that low that electrons arrive late (*right*). Mind the different x-axis time-scales. With the linear diode weighting potential and start point in the middle of the sensor, electrons and holes contribute equally to the signal. Figure 1.23 illustrates how this picture changes for a strip sensor and Fig. B.2 in the Appendix shows the case for a charge deposition at a different depth

the case of a silicon *pn*-diode (with space charge) or a large pad configuration (pad size much larger than drift distance) the weighting field $\boldsymbol{E}_W = \frac{1}{D}$ is also constant throughout the whole bulk (both electrodes same geometry). Due to the space charge the electric field is linear and has in this case the same form as the weighting potential. Therefore, in a diode, electrons and holes contribute equally to the induced signal, precisely for charge generation throughout the full volume or deposited in the center.

Figure 1.22 illustrates the dynamic situation in a diode assuming a linear electric field through the entire bulk. See formula 1.31, induced current $i(t)$ is a multiplication of weighting field $\boldsymbol{E}_W = \frac{1}{D}$ and charge velocity $\boldsymbol{v}(t)$ in turn proportional to the electric field $\mu \cdot \boldsymbol{E}(x)$. In a diode, we therefore expect the velocity of holes to increase, moving towards higher electric field thus increasing induced current (vice versa for electrons).

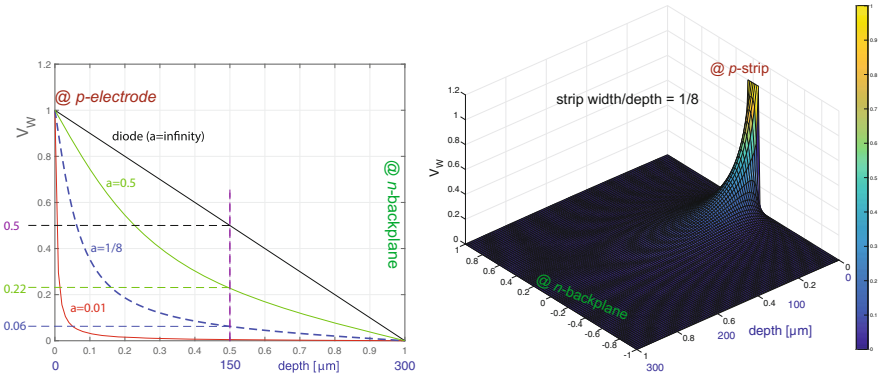


Fig. 1.23 The *left* plot illustrates the weighting potentials of a diode and strip sensors with $a =$ strip widths/ sensor depth. The $a = 1/8$ example corresponds to the geometric configuration as in Fig. 1.21 with $V_W = 1$ at the p electrode. The *right* figure gives the 3D representation of the weighting potential V_W for a strip sensor with $a = 1/8$. With the holes moving to the p-electrode, $Q_h(1 - V_W(x_0))$ and $Q_e(0 - V_W(x_0))$ gives the hole and electron contribution to the induced signal

For pixel or strip sensors, both, the electric field $\boldsymbol{E}$ and the weighting field $\boldsymbol{E}_W$ (and weighting potential V_W) peak at the segmented electrodes. These configurations are also reflected in Figs. 1.18 and 1.21. With the high electric field, also the drift velocity $\boldsymbol{v} = \mu \boldsymbol{E}$ and the charge mobility μ are maximal at (near) the pixel/strip electrode (even more true for a pixel than for a strip). As a result, with high velocity and high weighting field, the main current induction $i = -q\boldsymbol{E}_W \cdot \boldsymbol{v}$ happens in close vicinity of the electrodes; the location of original ionisation becomes less relevant.

As shown earlier $\left(Q = q[V_W]_{s_1}^{s_2}\right)$, with the knowledge of the local weighting potential the integrated induced charge from electrons and holes can simply be calculated by

$$Q_{\text{tot}} = Q_h + Q_e = q_h \cdot [V_W(p) - V_W(x_0)) + q_e \cdot (V_W(n) - V_W(x_0)] \qquad (1.32)$$

assuming the charges are generated at a sensor/diode depths of x_0 with weighting potential $V_W(x_0)$ and boundaries $V_W(p) = 1$ ($V_W(n) = 0$) at the p and n electrode respectively. For example, Fig. 1.23(left) represents the weighting potentials of strip sensors with different strip geometries and a diode versus depths.

With the figure and formula 1.32, one can derive that at mid-depth, electrons and protons contribute equally to the signal at the p electrode for a diode, while for a strip sensor (strip width/depth = 1/8), holes contribute 94% and electrons only 6%. With charge generated at $x_0 = 200$ µm holes would contribute 2/3 (in general $-N \cdot e \cdot x_0/D$) and electrons 1/3 ($N \cdot e \cdot (D - x_0)/D$) in a diode and 3% (electrons) and 97% (holes) in a p-strip. For a MIP generating charge throughout the volume of a diode, electrons and holes contribute 50:50. Since the weighting field $\boldsymbol{E}_W$ only depends on the given geometry it is possible to optimize the electrode design to

maximise signal and to select which charge type mainly contributes to the signal (50:50 for a diode to 90:10 or above for strips and pixels). However, there are more constraints in real sensors, e.g. breakdown voltage, signal-to-noise, position resolution, inter-strip capacitance, inter-strip resistance etc. More discussion about strip parameters are being presented in Sect. 1.6.3. A large scale example of a pad detector will be presented in Sect. 7.2. Calculating the weighting field also becomes more complicated with the introduction of signal routings (additional metal traces), intermediate strips or a metal overhang on the strips, e.g. used in the AC-coupled configuration of the CMS sensors (see also Sect. 6.4.2).

The Shockley–Ramo Theorem description is very powerful, especially when it comes to dynamic situations and timely signal formation. The Transient Current Technique TCT in Sect. 1.8.3 makes use of it to derive the detailed electric field configurations in the sensors from the signal formation. For integration times larger than the drift times, the notion of "charge collection", even if not correct, is most often sufficient to understand the situation, e.g. when evaluating charge distribution to several channels (as illustrated in Fig. 1.20). Very naively, the electric field strength (visually, the density of the field lines) gives a good estimate along which path charges drift (to be collected) and where velocities (current induction process) are high. The plots on the right of Fig. 1.21 show distributions of which percentage of charge is induced in the different electrode configurations. It will be briefly discussed in Sect. 2.2.4 on p. 160 how the situation changes with the onset of trapping after irradiation.

1.3.3 Signal Charge and Particle Position

The final position of the traversing particle is calculated by analysing the pulse height distribution (time integrated signal) on the affected strips (see Fig. 1.24). A top view photo of a sensor with descriptions of the diverse sensor elements is presented in Fig. 1.25. Figure 1.26 displays a 3D view of a standard single-sided p-in-n strip sensor design.

The strip pitch is a very important parameter in the design of the microstrip sensor. In gaseous detectors with a high charge multiplication a signal distribution over several sense wires is welcome to reconstruct the shape of the charge distribution and find the centre. In silicon detectors there is no charge multiplication and small charges would be lost in the noise distribution. Therefore, signal spreading over many strips could result in a loss of resolution. For single-strip events the track position is given by the strip number. The charge cloud on the way to the electrodes also diffuses in the lateral direction. Quantitatively this is mainly affected by bias voltage and temperature. The maximum time available for the lateral drift is the time to travel across the full volume. It can be calculated by $t_{e,h} = d/v_{e,h} = d/\mu_{e,h}E$ with the speed defined by $v_{e,h} = \mu_{e,h}E$. Typical values are thickness $D = 300\,\mu\text{m}$, mobilities $\mu_e = 1350\ \text{cm}^2/\text{Vs}$ and $\mu_h = 450\,\text{cm}^2/\text{Vs}$, mean electric field strengths $|\langle E\rangle| = 2.5$ kV/cm (DELPHI) or $|\langle E\rangle| = 15\,\text{kV/cm}$ (CMS). For these examples,

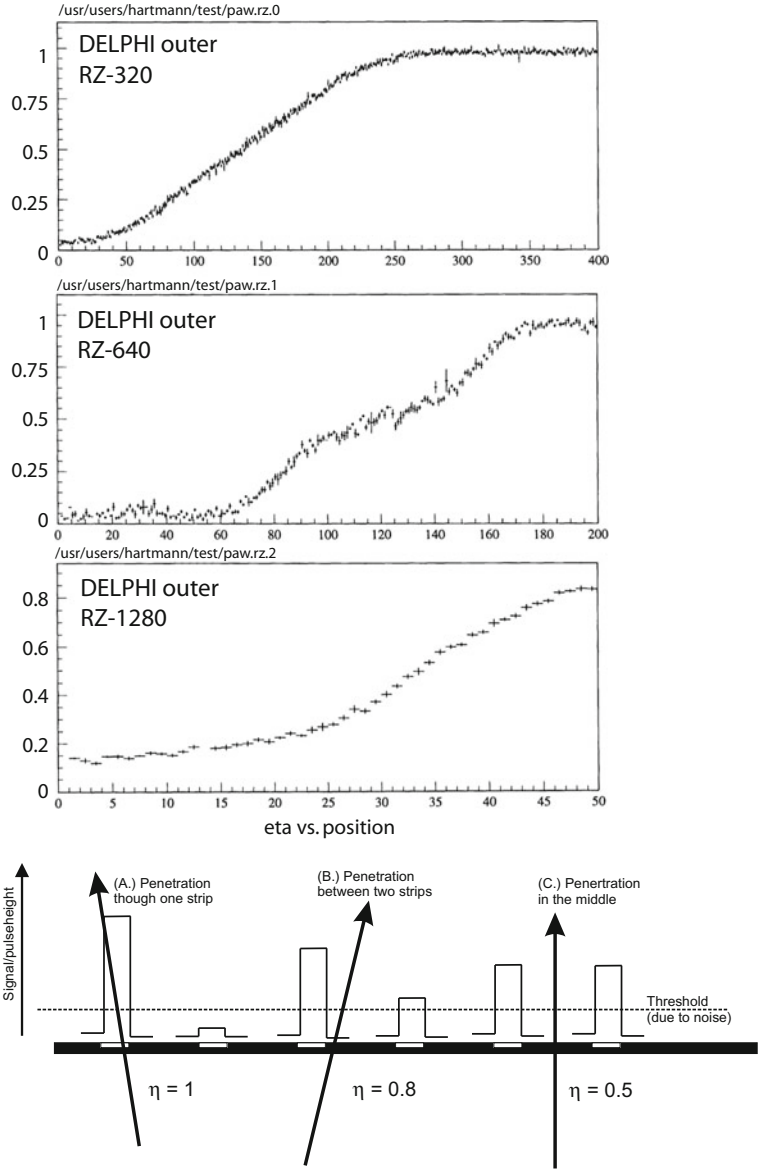


Fig. 1.24 Cluster shape depending on particle location. (**A**) Shows a single-strip cluster, the complete charge is collected on one strip. (**B**) Displays a two-strip cluster the second strip signal lies just above the threshold. (**C**) A distinctive charge distribution provides best localization precision. In the upper part, the η function is plotted against the position of a red laser hit for the three different sensor types of the DELPHI outer detector with 50, 100, 200 μm pitch for sensors RZ-1280, RZ-640 and RZ-320, respectively (1280, 640, 320 is the corresponding number of strips on the sensor). The x-axis counts laser steps of 0.5 μm [137]

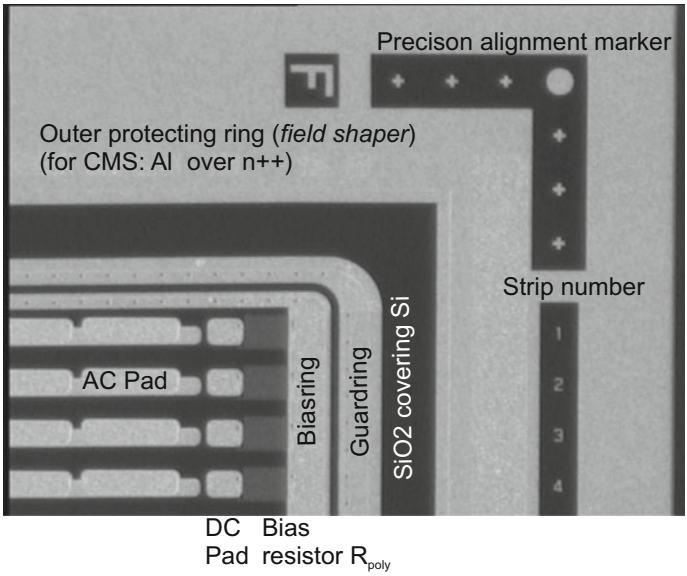


Fig. 1.25 The *top view* of a sensor, the ring structures, like n^{++} active edge protecting ring, guard ring and bias ring are easy to spot. Both guard and bias rings are Al structures located on *top* of the p^{+}-strip implants, they are directly contacted. Precision markers are needed to enable a precision assembly, while the strip numbers facilitate problem reports during quality assurance. The bias resistors connect the p^{+} strip located below the aluminium strips to the bias ring. A number of AC pads are processed at the end of the strips to enable several connections to the readout electronics. The DC pad, a direct contact to the p^{+} strip, enables probing. The backplane features the same high doping concentration n^{++} as the protection ring

the resulting lateral drift times are subsequently $t_e = 9\,\text{ns}$ and $t_h = 27\,\text{ns}$ (DELPHI) or $t_e = 1.5\,\text{ns}$ and $t_h = 4.5\,\text{ns}$ (CMS). With diffusion $\sigma = \sqrt{2\text{D}t}$ and the diffusion constant $\text{D} = \mu k_B T/q$, temperature dependence comes into play and the charge cloud diffusion for electrons and holes is the same, because μ cancels with $t \propto 1/\mu$. This results in a diffusion of approximately 6 μm in the LEP era running detectors at room temperature and below 1 μm in the CMS experiment operating at $T = -10°\text{C}$ (sensor temperature) and lower temperature. For tracks generating enough charge on two strips to exceed the threshold value, the position can be determined more precisely by either calculating the "centre-of-gravity" or better "centre-of-charge"[17] or by using an algorithm that takes into account the actual shape of the charge distribution[18] and the acceptance of the sensor. The best localization is achieved for tracks in the middle of two strips, since the charge is shared equally and the influence of noise is small. The signal for tracks near one strip have poor localization properties, because the remaining small signal on the neighbouring strips is often

[17] Assuming a uniform charge distribution, a track crossing between two strips at $\frac{3}{4} \cdot pitch$ will store $\frac{1}{4} \cdot charge$ on the left strip and $\frac{3}{4} \cdot charge$ on the right strip.

[18] Approximately a Gaussian distribution, due to the diffusion profile.

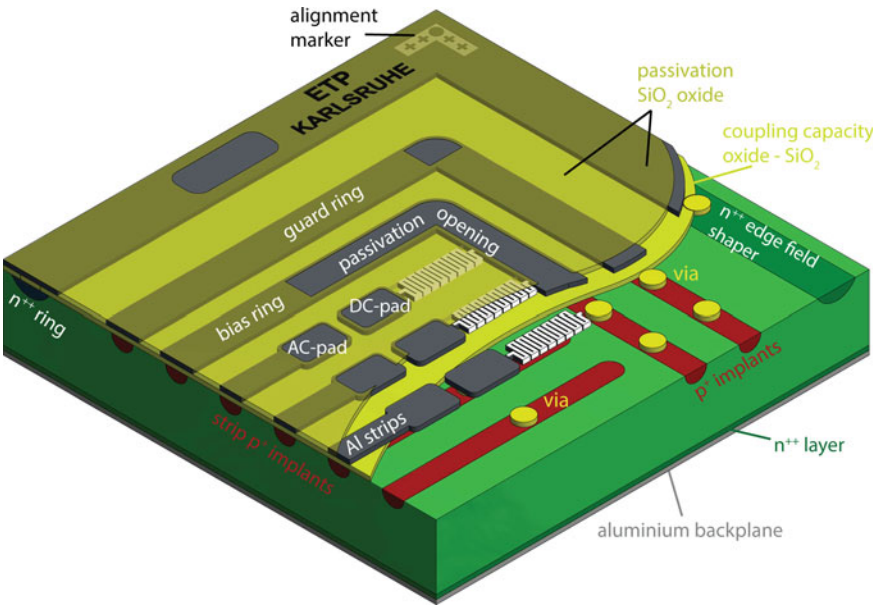


Fig. 1.26 A 3D schematic on a p-in-n sensor is sketched. It shows the baseline of the CMS sensor at the LHC in 2008, but could represent basically any single-sided AC-coupled, $R_{polysilicon}$ biased sensor. In operation, the bias ring is connected to GND potential, which is distributed via the bas resistors to the p^+ strips. The backplane n^{++} is set to positive high voltage depleting the full n-bulk volume by forming a pn-junction p^+-strip to n-bulk. The coupling capacitor is defined between aluminium strip and p^+-implant, the inter-strip capacity between neighbouring strips (both p^+ and Al part). The guard ring shapes the field at the borders. The n^{++} ring defines the volume and prevents high field in the real cut edge regions

in the range of or below the noise level. The "centre-of-gravity" or better "centre-of-charge" method, illustrated for two strips involved, makes use of the parameter η where

$$\eta = \frac{PH_L}{PH_L + PH_R} \tag{1.33}$$

PH_L and PH_R are the pulse heights measured on the left and right strip, respectively. As can be seen in Fig. 1.24; depending on final geometry and e.g. layout of intermediate strips the inter-strip behaviour is not always linear thus a custom function $f(\eta)$ is being used. The position is then calculated as

$$x = x_{\text{strip1}} + \frac{PH_2}{PH_1 + PH_2}(x_{\text{strip2}} - x_{\text{strip1}}) = \frac{PH_1 \cdot x_{\text{strip1}} + PH_2 \cdot x_{\text{strip2}}}{PH_1 + PH_2} \tag{1.34}$$

with a resolution of

$$\sigma_x \propto \frac{\text{pitch}}{\text{signal/noise}} \tag{1.35}$$

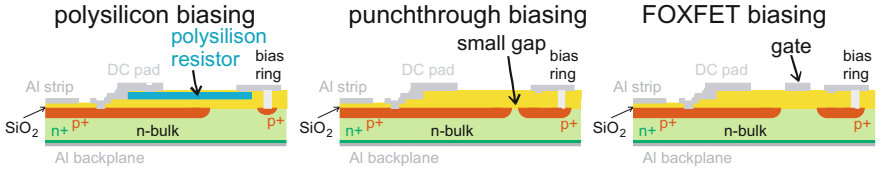


Fig. 1.27 There are three ways to bias an AC-coupled sensor plus a fourth one for a DC-coupled device. The punch-through and FOXFET biasing are the easiest ones but not as radiation hard as the polysilicon resistor, which is more or less the current standard. A photo of a polysilicon resistor can be seen in Fig. 1.43. In addition, a DC-coupled device can be biased via the connections to the electronics, representing the ground potential

As a result, sensors with a pitch of $p = 25$ µm and a signal/noise (S/N) of 50 have a position resolution of 2 – 4 µm. Additional intermediate implant strips in-between readout strips improve the resolution further by capacitively coupling to the neighbour readout strips. The intermediate strips must be kept at the same potential[19] as the readout strips to participate in the charge collection. This technique helps to minimize number of electronic channels while achieving an adequate position resolution. For binary readout, the position resolution is given by $\sigma_x \approx p/\sqrt{12}$.[20]

Design parameter values for the different structures like bias resistors and coupling capacitances for the different needs of different vertex detectors are presented in the following sections. Also structures like guard rings, bias rings, n^{++} rings, etc. will be discussed.

All individual isolated strips need to be at the same potential. This is realized in one of three different ways, illustrated in Fig. 1.27, although nowadays the polysilicon method is most utilized for strip sensors while punch-through is being mainly used for pixelated DC-coupled devices.

Figure 1.28 shows a 3D schematic of an n-in-p single-sided AC-coupled sensor. In an n-in-p sensor additional measures are necessary to isolate the individual n^+-strips from each other. The reason and some technical solutions are illustrated in the next section, while the figure features the p^+-stop "atoll" solution. Section 2.2.4 gives more detail why n-in-p sensors are very interesting despite the need of extra processing steps.

1.3.4 n-Side Isolation of an n-in-n or n-in-p Sensors

Keeping in mind the band structure on the n-side, positive fixed oxide charge on the SiO_2 and Si – SiO_2 interface attracts a layer of mobile electrons (accumulation layer). Without additional precautions, n^+-strips in an n- or p-bulk would be

[19] With the correct potential, electric field lines end here; otherwise they would adjust their potential and not collect any charge.

[20] $\sigma_x \approx p/\sqrt{12}$ arising from geometrical reflections: $< \Delta x^2 >= \frac{1}{p}\int_{-p/2}^{p/2} x^2 dx = \frac{p^2}{12}$.

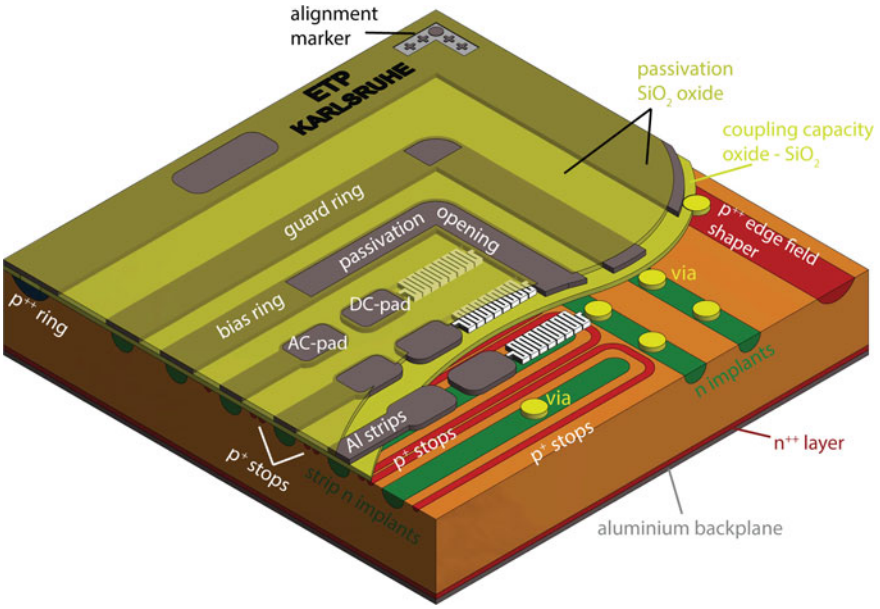


Fig. 1.28 A 3D schematic on a n-in-p sensor is sketched. This represents the current baseline for the CMS HL-LHC Upgrade detector (see also Sect. 7.1) and was initially brought forward by the RD50 collaboration. At first order it is like a negative/inverted version of the p-in-n sensors plus p^+-stops to establish the isolation of the individual strips. The schematic shows the so-called 'atoll' configuration where each individual strip is surrounded by its own p^+-stop; while some designs have a 'common' p^+-stop structure

short-circuited by the electron accumulation layer. Isolation is achieved via implantation of p^+ implants (p^+-stops) surrounding each n^+ strip with an individual or "atoll" configuration. The additional p^+ implants[21] are necessary to dissipate the electron accumulation layer at the Si – SiO_2 boundary thus isolating the readout strips. The aim is to cover as much area as possible of the gap between strips without actually contacting the strips. The production yield should not be lowered by coming too close. Another possibility would be the p-spray technology, where a diffuse layer of p-dopants is introduced. Often a combination of both p^+-stops and p^+-spray techniques are used. These isolation designs are necessary features for n-in-p sensors under investigation for future detectors, like at the HL-LHC. The nomenclature of $^+$ and $^{++}$ here is being used in a lax way, in general the backplane-implant has the highest doping concentration, next comes the strip implants here n^+ then the p^+-stops then the p-spray and ending with the lowest concentration the bulk n or p having a high resistivity. Another way to repel the electrons is to increase the Al strip widths, apply a potential and use them as "field plates" (done for DELPHI). In Fig. 1.29 the

[21] Reminder, the p^+-stops are negatively charged/have a negative space charge.

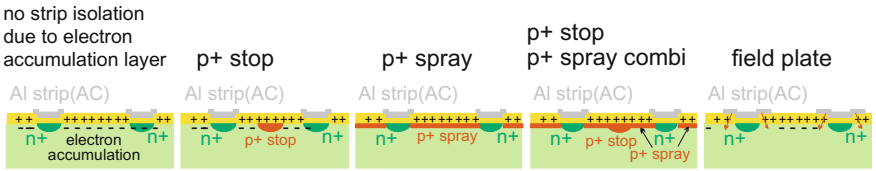


Fig. 1.29 Positive charges in the oxide attract electrons in the n-bulk material to the interface where an electron accumulation layer forms and short-circuits the strips. The basic mechanism to avoid this is to repel the electrons by the introduction of p^+dopant in an isolated region or simply distributed/sprayed in the inter-strip zone. These methods are called p^+-stop, p^+-spray or a combination of both. A fourth method is the use of a metal overhang of the AC pad at slightly negative potential to repel the electrons (A more detailed discussion about the Si – SiO_2 interface can be found in Sect. 1.1.3)

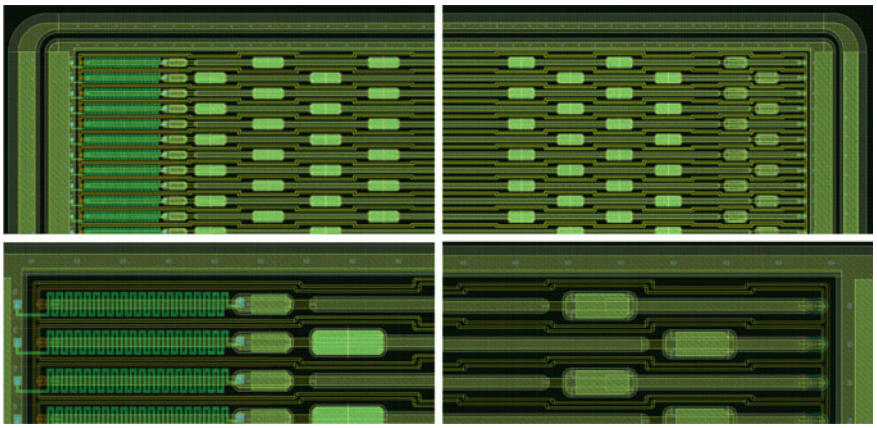


Fig. 1.30 The *upper layouts* show both ends of a sensor; both ends contain two AC-pads per strip and one end features the polysilicon bias resistor supplying the GND potential to the individual strips. The *lower part* depicts a zoom to the individual pads. Every single strip with its pads and bias resistor is surrounded by an individual atoll p^+-stop closing in as much as possible while additionally all strips together are also circumferenced by another p^+-stop ring. Obviously this is more complicated compared to a standard p-in-n sensor. The same layout with p^+-spray isolation would simply miss the additional p^+-stops and look very much like a p-in-n sensor

different isolation methods, namely p^+-stop, p^+-spray and a combination of both are schematically drawn.

Examples of application are to be found in Fig. 4.12 and a photo of p^+-stops can be visited in Figs. 4.11 and 5.19, a schematic can be found in Fig. 5.18. More design trials for the future are shown in Sect. 7.1.1. Figure 7.21 on p. 319 will show examples of p^+-spray, "atoll" and common p^+-stop configurations and more.

Figure 1.30 shows the example of a layout of an "atoll" p^+-stop configuration for a n-in-p strip sensors.

1.4 Single-Sided – Double-Sided, Double Metal

So far, only single-sided sensors have been introduced. Segmentation of the bulk silicon material can be done on both sides with many benefits but also many additional challenges. The obvious benefit is a 2D readout with different strip orientation on each side[22] of a single sensor. Strip implants are then composed of p^+ and n^+ on the two sides, named junction and ohmic side, respectively. As discussed in the previous section the ohmic side, with n^+ strips in an n-bulk, needs special attention to arrange strip isolation due to the presence of an electron accumulation layer with additional p^+ doping in-between n^+ strips or electron-repelling field plates (see Sect. 1.1.3 about accumulation layer at the Si – SiO_2 interface and Sect. 1.6.3 about isolation strategies for n-in-n readout). Another illustrative case is presented in Fig. 4.12 on p. 188, where the DELPHI double-sided sensors are discussed. Double-sided processing is briefly discussed in Sect. 1.9.2. Several additional processing steps are needed and in several cases, the sensor faces are not completely independent, as a critical example on p. 127 in Sect. 1.14 for the CDF experiment proves. Production yield is a key issue with double-sided sensors. In many cases, the second side needs an additional routing layer to lead the signals from the implants to the readout electronics. The routing is normally realized via an additional isolation layer in combination with another layer of aluminium routing strips (called double metal layer) connected individually by "vias" to the implants (see Figs. 4.11 and 4.12). Double-sided sensors are great but their substantially increased complexity makes them unsuitable for very large detector systems.

Routing

Often, the sensor strips do not terminate at the right location to conveniently connect them to the readout electronics (mostly to have a common readout hybrid at a special location e.g. at the very detector end, outside of the tracking volume). In which case a connection to the strips needs to be routed, for example for the orthogonal strips of a double-sided sensor or in special detector geometries, e.g. the **VE**rtex **LO**cator VELO [57, 324, 325] of the LHCb experiment, see Fig. 1.31. Many examples will be given in the chapters describing the DELPHI (Chap. 4) and CDF (Chap. 5) detector. Extra masks and processing steps are necessary to establish this double-metal layer. Figure 1.32 shows the integration of a pitch adapter in the sensor itself, a study for CMS. More complicated and longer routings are evaluated for the 'International Linear Collider' ILC sensors, where the readout chips would be bump bonded to the sensors and their 1024 channels are then routed to the strips. Such a routing is most often achieved by adding an additional insulation layer plus metal routings on top of it with dedicated positions to connect the first and second metal layer via "vias". The schematic can be seen in Fig. 1.33, while Figs. 1.31 and 4.11 show exemplary photos of such strip configurations. The schematic of a double-sided double-metal sensor

[22] Common strip orientations are 90° or a small stereo angle like 0.1 – 2°.

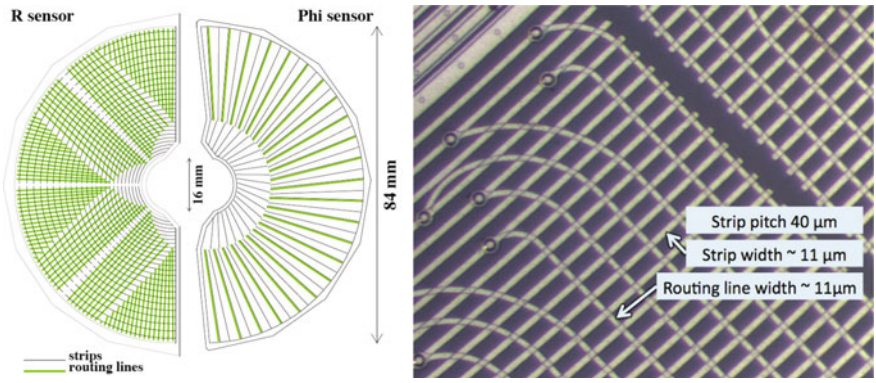


Fig. 1.31 The LHCb VELO (VErteX LOcator) sensor schematic is shown on the *left*. The *black lines* indicate the active strips in ϕ (radial) and r direction (circular). The *green lines* show the *routing lines* in the second metal plane routing the strips to the readout located at the outer fringes. The sensors surround the beam pipe directly in perpendicular geometry. On the *right*, a microscopic photo is displayed. More on the VELO detector and source of the figures can be found here [324, 325]

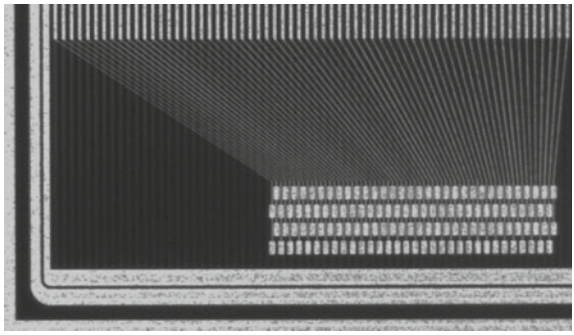


Fig. 1.32 The photo shows a pitch adapter PA integrated in the sensor itself. This can be realized in the first metal layer with some performance loss in the PA area or in a 2nd metal layer without any performance loss

is shown in Fig. 4.12. The insulation between metals layers consists mostly of SiO_2 or polyimide. It allows very complicated routing to connect to the strips, e.g. stereo angle, full orthogonal, crossing strips, curves having more or less dense connection areas.

This "simple" solution has some disadvantages or at least boundary conditions to pay attention to. The additional metal layer represents an additional parasitic capacitance to which signal can couple, and the channel capacitances increase thereby increasing the system noise. Due to these reasons the insulation layer should be as thick as reasonable achievable by the processing methods (decrease of capacitance and reduction of signal coupling). One to two micrometer are possible thicknesses for six inch wafers. Another problem is the mechanical stress created by an additional thick insulation layer on only one sensor face (stress of the crystal lattice and additional bowing of the wafer/sensor). This can be balanced by adding a similar

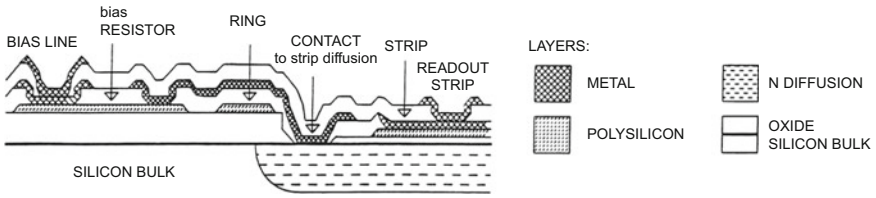


Fig. 1.33 Schematic of an AC-coupled strip plus a double metal layer on top. This configuration features an additional insulation layer most often of SiO_2 or polyimide where additional metal lines can route the signal to the electronics. Here, the readout strip runs orthogonal to the active strip implant; instead of having a direct simple AC-pad [231]

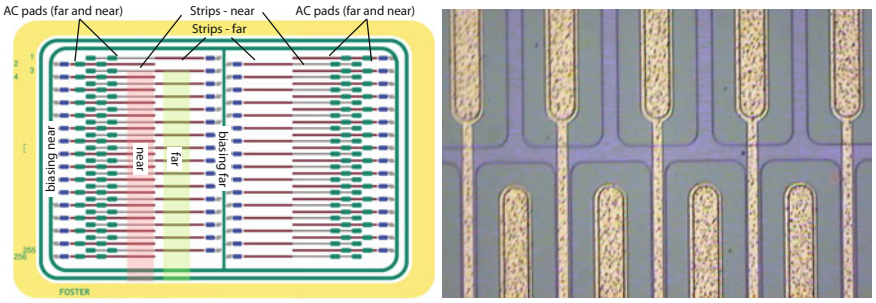


Fig. 1.34 FOSTER: A **fo**urfold **s**egmented silicon s**t**rip s**e**nsor with **r**eadout at the edges. A special configuration, where the connections are at the sensor ends the strips technically segmented in four. The concept works also for higher segmentation as long as there is space for the routing lines (larger pitch) [73]

thick layer on the other sensor face but then dedicated voltage contacts have to be established to the backplane (e.g. opening of insulation by an extra mask plus etching).

The additional insulator layer is also prone to surface radiation damage (see Sect. 2.3) and one has to expect changes with radiation.

Another solution would be to achieve the routing via external structures like flex hybrid, printed circuit boards or glass adapters, which reduce the intrinsic sensor complexity but add complexity and mass to the complete system.

Without crossings of channels, routing is also possible in the first layer; therefore without extra masks/processing. A special case study is being discussed in [73] where the granularity along the strip has been increased by a factor of two with all connections at the sensor ends. The routing of the inner (or here called far) strips happens on the first metal layer but additional implants below the metal lines were necessary to avoid high fields at the routing lines otherwise also they register part of the signal – see Fig. 1.34.

Such routing in the first metal layer is also being discussed and successfully realized for pixel geometries to allow for different pixel sensor cells than electronic cell geometries. The pads for the bumps of sensor and electronics have to match for a successful flip-chipping (see more in Sect. 1.12.1) and to achieve this the sensor contacts can be routed to the electronics pad location.

1.5 Noise Contributions

Even for a fully depleted sensor bulk volume, sources of statistical electron and hole fluctuations exist. These fluctuations, called noise, are counter-productive and compete with the signal from the traversing ionising particles. The sources are manifold and of different severity for the diverse operation conditions and readout strategies. A strong dependence on the peaking time t_p and operating temperature T exists. With **s**ignal-to-**n**oise (S/N) being the figure of merit of a sensor, noise must be minimized while signal can only be increased by increasing sensor thickness. The noise contributing elements are load capacity C_{tot}, leakage current I_L, parallel and series resistances R_P and R_S. A simplified noise scheme is shown in Fig. 1.35. The load capacitance C_{tot} is mostly comprised of the inter-strip C_{int} and strip to backplane capacitance C_{back}.

Noise is generally expressed as **E**quivalent **N**oise **C**harge ENC or Q_n, representing number of electrons contributing to the noise. The different contributions sum up quadratically:

$$ENC = \sqrt{ENC_C^2 + ENC_{I_L}^2 + ENC_{R_P}^2 + ENC_{R_S}^2} \tag{1.36}$$

with the *shot noise* term from leakage current

$$ENC_{I_L} = \frac{e}{2}\sqrt{\frac{\mathbf{I_L} \cdot \mathbf{t_p}}{q_e}} \approx 107\sqrt{\mathbf{I_L}/nA \cdot \mathbf{t_p}/\mu s} \tag{1.37}$$

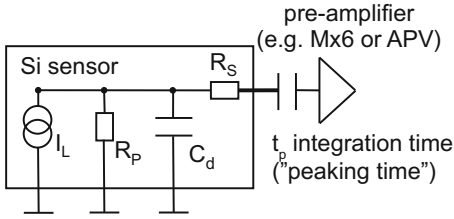


Fig. 1.35 A simplified equivalent network of a sensor together with its preamplifier. The main noise sources leakage current I_L, parallel and series resistance R_P, R_S and load capacitance C_d are shown. Such a diagram is valid per individual strip or pixel

the *parallel thermal noise* term from bias resistance

$$ENC_{R_P} = \frac{e}{q_e}\sqrt{\frac{k_B T \cdot \mathbf{t_p}}{2\mathbf{R_P}}} \approx 44.5\sqrt{\frac{T/K \cdot \mathbf{t_p}/\mu s}{\mathbf{R_P}/M\Omega}} \tag{1.38}$$

the *serial thermal noise* term from the metal strip resistance

$$ENC_{R_S} = \mathbf{C_d} \cdot \frac{e}{q_e}\sqrt{\frac{k_B T \cdot \mathbf{R_S}}{6\mathbf{t_p}}} \approx 0.025 \cdot \mathbf{C_d}/pF\sqrt{\frac{T/K \cdot \mathbf{R_S}/\Omega}{\mathbf{t_p}/\mu s}} \tag{1.39}$$

and the most significant contribution, the load capacitance

$$ENC_C = a + b \cdot \mathbf{C_d} \tag{1.40}$$

where e is the Euler number, k_B the Boltzmann constant, q_e the electron charge, t_p the peaking time and T the operating temperature. The constants a and b, in the ENC_C term, are preamplifier-specific parameters. The slope b is mostly defined by the intrinsic chip voltage noise $u_{n,amp}$ and the integration time ("peaking time") t_p together with a form factor arising from the pulse shape.

$$b \sim \frac{u_{n,amp}}{\mathbf{t_p}} \tag{1.41}$$

Vice versa the above-mentioned noise behaviours have direct implications on the design parameters of the sensors. To minimize noise the following sensor design criteria have to be fulfilled:

- small load capacitance $C_{tot} = C_{\text{strip}}$ ($\sim$strip length)
- low leakage current I_L
- high parallel resistance R_{bias}
- small series resistance R_{strip}

These design choices and parameters will be discussed in more detail in the next sections and chapters, e.g. strip length, choice of bias method, sensor quality, etc. Furthermore, the frequency dependence, hence readout method choice, significantly influences the design choices. For example, capacitance tuning is much more relevant for short peaking times ($\sim\sqrt{\frac{1}{t_p}}$), while for larger peaking times noise due to leakage current and bias resistance is more important ($\sim t_p$). Also the temperature is important for both ENC_{I_L} (*shot noise*) due to leakage current (current doubles for $\Delta T \approx 7°\text{C}$) and ENC_{R_P} and ENC_{R_S} (*thermal noise* $\sim k_B T$). Obviously, detectors at the future **I**nternational **L**inear **C**ollider ILC, with a longer shaping time will be sensitive to low resistor values and shot noise from leakage current, whereas detectors at the LHC with a very short bunch crossing use a short shaping time, where noise is maximally affected by the ENC_C term. To reduce the peaking time dependency one can use an analogue pulse shape processing step after the preamplifier stage and therefore

Table 1.2 The table shows the different noise contributions for a DELPHI and a CMS module. In the quadratic sum, the ENC_C term is the most significant. The higher noise of the CMS module is dominated by the faster peaking time of 50 ns, which mostly determines the slope b in the ENC_C term and the ENC_{R_S} term. To compensate the high noise for long CMS modules, the sensor thickness had to be increased to achieve a higher signal. The shot noise term ENC_{I_L} is suppressed ($I_L = 1\ nA$). The second ENC_{I_L} value of $756e^-$ for CMS reflects the current increase due to irradiation

Detector	t_P dep	DELPHI microvertex	CMS tracker	
Chip		MX6	APV peak	APV deconv
t_p		1.8 µs	50 ns	12 ns, effective
T		20°C	−10°C	
C_{tot}		2×6 cm → $C_{tot} = 9$ pF	2×10 cm strips → $C_d = 18$ pF	
		a = 340; b = 20	a = 270; b = 38	a = 400; b = 60
→ ENC_C	$\sim 1/t_P$	$520e^-$	$711e^-$	$1480e^-$
I_L		0.3 nA	1 nA to 1 µA	
→ ENC_{I_L}	$\sim \sqrt{t_P}$	$78e^-$	$24e^- - 756e^-$	×0.45
$R_P = R_{\text{bias}}$		36 MΩ	1.5 MΩ	
→ ENC_{R_P}	$\sim \sqrt{t_P}$	$169e^-$	$131e^-$	×0.45
$R_S = R_{\text{strip}}$		25 Ω	50 Ω	
→ ENC_{R_S}	$\sim \sqrt{1/t_P}$	$13e^-$	$230e^-$	×1.45
sum ENC		$552e^-$	$760e^- - 1073e^-$	$1518e^- - 1556e^-$

increase the integration time, but the pulse shape processing (deconvolution mode) itself will slightly change the different noise terms [156, 290, 299]. An example on how the different noise terms contribute for different modules and different peaking times is given in Table 1.2 including the modification factors for the deconvolution mode. Some examples of the ENC_C term for different chip versions of DELPHI, CDF and CMS, adapted for different load capacities and different timings, are given in Table 1.3. The three different values for the different stages of the DELPHI experiment represent different chip generations. The large dependence on capacitance for the CMS experiment presents the strong frequency dependence (peaking times: 1.8 µs for DELPHI, 132 ns for CDF and 50 ns for CMS). The C_{tot}-dependent value $b = 60\ e^-/\text{pF}$ for CMS seems to be high, but it is completely driven by the sampling time; compared to the MX3 value one should appreciate the development progress made between the year 1990 and 2000. Obviously the strip length is proportional to the load capacitance C_{tot}, being composed of the strip capacity to its neighbour and to the sensor backside. An example of linear noise dependence versus C_{tot} (strip length) is illustrated in Fig. 1.36.

Table 1.3 ENC values for the different chip generations. The bold numbers emphasize some important steps, e.g. load capacity adaptations (TRIPLEX) and reductions even for faster shaping times (CDF and CMS)

Detector; Chip	$ENC = a + b \cdot C_{tot}$
MARK II; Microplex	$ENC = 280e^- + 97e^-/\text{pF} \cdot C_{tot}$
DELPHI; MX3	$ENC = 670e^- + 55e^-/\text{pF} \cdot C_{tot}$
DELPHI; MX6	$ENC = 340e^- + 20e^-/\text{pF} \cdot C_{tot}$
DELPHI; TRIPLEX	$ENC = 283e^- + \mathbf{17}e^-/\text{pF} \cdot C_{tot}$
CDFII; SVX3D	$ENC = 500e^- + \mathbf{60}e^-/\text{pF} \cdot C_{tot}$
CMS; APV (Peak)	$ENC = 270e^- + \mathbf{38}e^-/\text{pF} \cdot C_{tot}$
CMS; APV (deconv)	$ENC = 400e^- + \mathbf{60}e^-/\text{pF} \cdot C_{tot}$

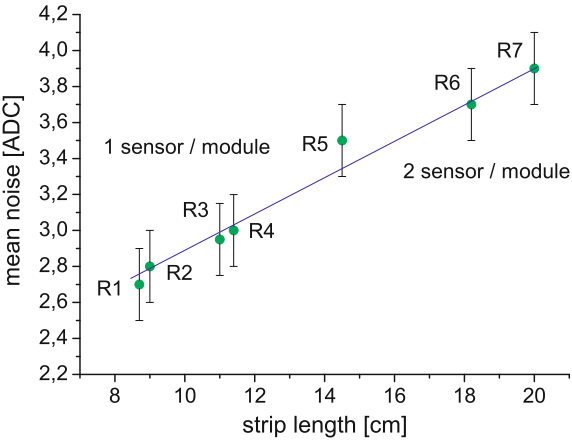


Fig. 1.36 Noise increases with strip length for the different CMS forward detector sensors and module geometries. R1–R7 labels the modules at different rings of the forward detector situated at different radii

1.6 Sensor Parameters

All parameters have to be tailored, specified in detail, validated and monitored according to their specific need, which will be discussed in the following sections. Certain interdependencies must be balanced to achieve best results. In the following, geometric structures are discussed, including their effects on the functionality like bias and guard ring, implant and Al together with p^+ strip widths and thickness. Furthermore, the fundamental electrical parameters are introduced to achieve the best operation conditions for the different needs, which are often substantially different for detectors of the last decades and their respective environments.

1.6.1 Global Parameters

The global sensor parameters, full depletion voltage V_{FD} and total leakage current, are discussed here.

Full Depletion Voltage ($V_{\text{depletion}} = V_{FD}$)

To ensure full charge collection in the detector, the silicon has to be depleted of free and mobile carriers by applying $V_{\text{bias}} \geq V_{FD}$ in reverse direction. The drift velocities and therefore the drift times and widths of the charge distribution are also affected by the bias voltage. Wider distributions lead to events where the charge is shared between multiple strips. For these events the centre-of-gravity calculation is possible, having an immediate positive consequence on the position resolution. Micro-discharge is another critical issue, which can be avoided by limiting V_{bias} and thus also V_{FD} safely below the micro-discharge voltage.

The depletion layer acts like a parallel plate capacitor, therefore the bulk capacitance *per unit area* is determined by the depth of the depletion layer w:

$$C_{\text{bulk}} = \frac{\varepsilon_0 \varepsilon_{\text{Si}}}{w} = \sqrt{\frac{\varepsilon_0 \varepsilon_{\text{Si}}}{2\mu\rho V_{\text{bias}}}} \tag{1.42}$$

where ε_{Si} is the permeability of the bulk silicon, μ the mobility and ρ the resistivity. Depleting the full sensor $w = D$, formula 1.42 resolves to V_{FD}

$$V_{FD} = \frac{q D^2 |N_{eff}|}{2\varepsilon_0 \varepsilon_{\text{Si}}} = \frac{D^2}{2\varepsilon_0 \varepsilon_{\text{Si}} \mu\rho} \tag{1.43}$$

where N_{eff} is the effective space charge density, which is the number of donors minus the number of acceptors, and q is the absolute value of the electron charge. After depleting the whole bulk, $w = D$, C_{bulk} will remain constant. This gives a characteristic behaviour C_{bulk} over voltage.

$$C_{\text{bulk}} = \begin{cases} \sqrt{\frac{q\varepsilon_0\varepsilon_{\text{Si}}|N_{eff}|}{2V_{\text{bias}}}} & V_{\text{bias}} \leq V_{FD} \\ \frac{\varepsilon_0 \varepsilon_{\text{Si}}}{D} & V_{\text{bias}} > V_{FD} \end{cases} \tag{1.44}$$

The value of V_{FD} is directly observable in a CV plot, where C_{bulk} becomes constant and the curve flattens. The absolute V_{FD} mainly depends on the resistivity ρ and therefore N_{eff} (see also Fig. 1.10 on p. 16). The applied voltage also influences the collection time of the electron–hole pairs, which furthermore defines the time of lateral diffusion. The diffusion determines the cluster size of the signal, see discussion on p. 23. A deviation of depletion voltage behaviour for segmented devices is briefly presented in Sect. 1.6.3.

Another criterion for a large silicon vertex detector is the ability to apply one single V_{bias}, therefore $V_{FD} \propto D^2 \cdot N_{eff}$ should not vary over the individual sensors. A good uniformity of wafer resistivity is very beneficial. Unfortunately, the direct

dependence of V_{FD} on the raw silicon resistivity results in a range of up to 200 V in the CMS experiment, especially for low-resistivity silicon sensors. In the ideal case, V_{FD} should be homogeneous, because several sensors are ganged to one module (e.g. see DELPHI Fig. 4.6 or CDF Fig. 5.14) and only one voltage can be applied, or even worse many modules are connected to one high voltage power supply channel. In addition, the Lorentz angle [74] is dependent on the operation voltage, which has to be taken into individually for position resolution and track reconstruction.

Total Leakage Current ($I_{leakage}$)

A low total leakage current is one of the major criteria for the quality and functionality of a microstrip detector. Leakage current is a source of noise in the final readout system, which should be kept as low as possible. Leakage current shows up in all detectors coming from generated electron–hole pairs in the silicon in the presence of an electric field. The leakage current is proportional to the depletion layer thickness w of the sensor which is proportional to $\sqrt{V_{\text{bias}}}$. It should be constant after reaching V_{FD}. Deviations from this rule are direct indications of sensor flaws, e.g. variations in processing. Bulk sensor currents are direct indications for bulk impurity concentrations, defining the charge carrier lifetime, and thus the current. The leakage current should not exceed a certain limit, excess leakage current results in noise. High current, especially a localized one, is also a defect signature. When measuring the global leakage current, bulk current must be distinguished from surface currents deriving from interface traps, which cannot be avoided and/or "bad" oxides, scratches and/or process errors. Bulk leakage currents are strongly temperature dependent (Sect. 1.1.2), while surface currents are much less temperature dependent. Finding the sources of high total leakage currents, can help to improve process methods (see Sect. 1.14).

The strong dependence of leakage currents to temperature leads to an additional critical effect – thermal runaway. Especially after irradiation, sensors have to be cooled to avoid a thermal runaway due to increase of leakage current and voltage – power. Heating power in the silicon can be expressed by

$$P_{\text{silicon}} = V_{\text{bias}} \cdot I_{\text{leakage}} \sim V_{\text{bias}} \cdot T_{\text{silicon}}^2 \cdot e^{-\frac{E_{eff}}{2 \cdot T_{\text{silicon}} \cdot k_B}} \tag{1.45}$$

with the effective energy E_{eff} describing the bandgap energy $E_{gap}(T)$ and its temperature dependence close to room temperature [64]. For the relevant temperature regime $E_{eff} = 1.21\,\text{eV}$ is used. The system is self-feeding; higher temperature introduces higher current, which vice versa increases power and temperature. Assuming a linear cooling power, a critical temperature limits the stable running. The system overheats, "runs away" above $T_{critical}$, while it stabilizes for $T < T_{critical}$ at T_{stable} (see also Fig. 1.37). For non-irradiated silicon $T_{critical}$ is well above room temperature.

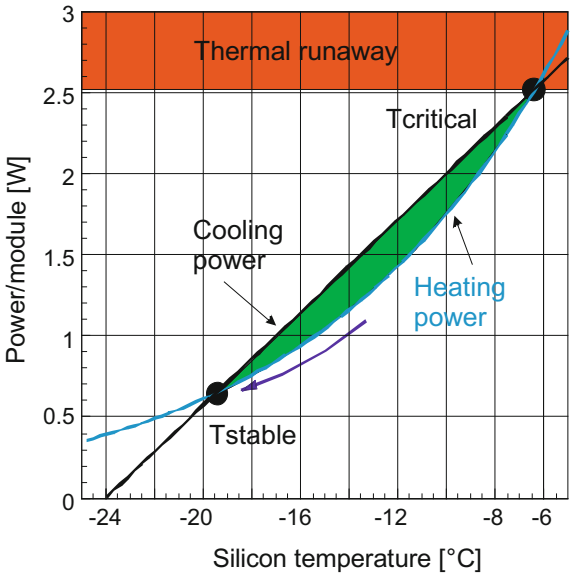


Fig. 1.37 Thermal runaway. For $T > T_{critical}$, the system will "thermally run away", while for temperatures below the critical one $T < T_{critical}$, the system stabilizes at T_{stable}

1.6.2 Bias-, Guard- and Outside Protection Rings

The ring structures surrounding the active area including the sensor outer edge are discussed in this subsection as well as passivation.

Passivation

Passivation is the final step to protect the sensor from the environment. It often consists of a crude form of SiO_2, sputtered on the sensor or a film of polyimide[23]. Only the bonding and testing areas (Al) are later opened, refer also Fig. 1.11. In contrast to the passivation needs in industry, the material for sensor protection has to be without contaminants. Already small charge concentration in the passivation can change the well-designed field configuration on the implants and $Si - SiO_2$ interface, e.g. negative charge concentration can evoke a surface inversion layer, opening a hole conduit from strip to strip decreasing inter-strip resistance. An example of this undesired effect is presented in Sect. 6.5.1.

Bias Ring

The bias ring and backplane are the main contacts to apply the bias voltage. The bias ring runs around the whole active area of the sensor to ensure a homogeneous potential for all strips. It either connects to all bias resistors, therefore supplying all implants with voltage, or via the punch-through effect or FOXFET described earlier. The bias ring is not ultimately necessary for DC-coupled sensors, where the potential

[23] Polyimide (sometimes abbreviated PI) is a polymer of imide monomers, it is resilient against ionising radiation.

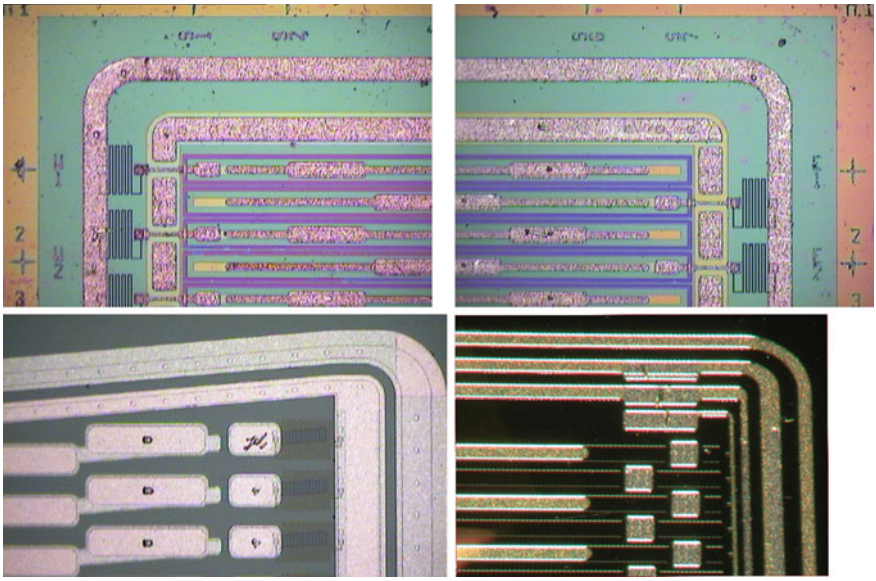


Fig. 1.38 The bias ring connects the strips to a defined potential, e.g. GND on a single-sided p-in-n sensor. The bias ring in the *first* two photos surrounds the guard ring and connects via polysilicon resistors to the individual strips – DELPHI n-side. On the *lower left*, a CMS sensor corner can be seen, with the bias ring inside the field shaping guard ring also providing bias potential via polysilicon resistors. On the *lower right*, a DELPHI Ministrip sensor corner is displayed, with (from *right* to *left*) two outer guard rings, a bias ring and a gate to control the FET channel /conductivity between bias ring and strips. Note the bond pads for one guard ring, the bias ring and the FOXFET gate

can be applied via the readout chip connection. A DELPHI and CMS example with polysilicon biasing plus a FOXFET bias example from DELPHI is displayed in Fig. 1.38. The different biasing schemes are explained in Fig. 1.27.

Guard Ring(s)

The guard ring shapes the field inside the sensitive area to minimize edge effects and guarantees a defined homogeneous potential for all strips, including the edge ones.

Two basic connection schemes can be applied. First, a direct connection of the guard ring to a certain potential, often GND, provides a drain for the leakage currents from the edges of the detector, e.g. applied for the DELPHI sensors. The second configuration uses one or more "floating" guard rings to discretely adapt the potential, especially for high voltages, where the voltage drops from outside in. The ultimate example of a single guard ring serving as multi-guard structure is described in Sect. 6.4.2. For the purpose of field shaping, an implant ring with the same type as the strip implant (p^+ / n^+ for p-in-n / n-in-p respectively) is located underneath the guard ring. The Al planes and implants are directly connected by vias.

Sensor Edges

Active Edge

Cracks caused during the cutting process at the edges introduce energy levels in the forbidden gap – large defect density at the sidewalls. The suppression of resulting leakage currents can be accomplished by avoiding high fields inside the cracked region. This can be achieved by introducing a broad implant at the edges of the sensor front face. This so-called 'edge field shaper' on the 'front-side' has the same polarity as the backside doping, n^{++} for p-in-n or p^{++} for an n-in-p sensor – the concept also works for double-sided sensors. This feature is graphically presented in the photo 1.25 and schematic 1.26 (p-in-n) and 1.28 (n-in-p). The edge implantations on both sensor faces then levels to the same potential, in most cases the HV potential. The edge therefore has a *defined* potential on the top. This design was first, very successfully, introduced for CDF RUN II (the n^{++} edge field shaper) and is consequently used in the current and future CMS sensor design. Without such a "protection" the depletion zone would reach the sidewalls creating a conductive surface and hence increase the leakage currents significantly.

This edge feature has historically, when first implemented, been called "active edge" but is today very common. The term "active edge" is therefore, today, mostly used for the "slim edge" concept described in the next paragraph.

Slim Edge

Due to very tight spacial constraints in a pixel detector, especially in the barrel section, pixel sensors cannot be staggered in the z direction, hence the inactive sensor edge defines a real inactive zone within a detector layer. The "active edge" (see last paragraph) is therefore uncomfortably large (500 – 1000 μm) and also multi-guard ring structures are large. There are additional ways to decrease the sensor edge but these come with additional processing and with steps unfortunately not available at all suppliers – the goal would be to go down to ∼50 – 100 μm to tile sensors seamlessly. There is also a high demand from medical applications. The **D**eep **R**eactive **I**on **E**tching DRIE, a 3D process, as described in Sect. 1.12.7 can be used to create very defined practically defect-less edges/sidewalls because they are etched instead of cut/sawed; one 'simply' etches a trench through the full sensor, itself resting on a handling wafer. One also often reads ICP-RIE (**I**on **C**oupled **P**lasma **R**eactive **I**on **E**tching.) or shortly ICP when discussing trench etching. This 3D process also allows to have a well-defined doping layer on the vertical 'sensor edge'. Even after the trench etching, the sensors rest on the handle wafer thus full wafer processing is possible hence doping via diffusion can be done for the sidewalls as is done for the pillars in a 3D sensor – Fig. 1.39 illustrates the additional steps. This allows to reduce the inactive distance between edge to active zone and have it at a well-defined potential.

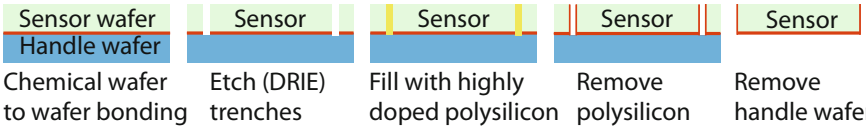


Fig. 1.39 Slim edge/active edge. The goal is to have a very smooth sensor sidewall featuring a very low defects concentration and which is homogeneously doped. A wafer with phosphorous-doped backside[24] is chemically bonded to a handle wafer. The trenches are created by DRIE. They are filled with highly phosphorous doped polysilicon to allow phosphorous diffusion into the sidewalls. The red lines symbolize the doping. Before the removal of the polysilicon all other processing steps happen (strips, polysilicon resistors, Al, etc.). Afterwards the polysilicon walls are being etched away by reactive ion etching and finally the support wafer is being removed

It is also possible to define the cut-edge region by doping or passivating with charged-up material,[25] after cutting by laser, diamond saw or even cleaving. This is only possible at FABs[26] which are able to process on individual structures. Most often processing is only done on full wafers, where the sensor edges are not accessible.

Another possibility described in [97] introduces an initial scribing perfectly along the crystal lattice orientation on the surface, e.g. via laser or DRIE and then cleave the sensor. The sidewall surfaces are then passivated[27] with dielectric with interface charge. This connects the dangling bonds and creates an appropriate interface charge with silicon. The sidewall becomes effectively resistive in a controlled way.

Slim edge design is being used in the LHC TOTEM detector where the sensors are perpendicularly placed as close as possible to the beam pipe [261].

1.6.3 Design of Strip Parameters

This section illustrates the basic design choices for aluminium readout and implant strips as well as the resulting electrical parameters, like coupling and inter-strip capacitances. Also inter-strip and bias resistances are discussed here. Common strip faults are introduced. The AC-coupling case is described. Figure 1.40 shows an electron microscope picture of a cut through an AC-coupled strip.

Layout of Implants and Aluminium Strips

To define the Al strip implant configuration several criteria have to be taken into account. In an AC-coupled sensor, the coupling capacity must be larger than the inter-strip capacitance and all parasitic capacitances. To first order the aluminium strip

[24] Phosphorous for n-bulk sensors; boron for p-bulk sensors.

[25] The reason to dope the vertical cut edges is to define a well known doping profile and potential, dominating any crystal defect from the cutting.

[26] FAB fabrication line of silicon devices, CMS sensors, etc.

[27] Negative charged passivation for p-type, e.g. Al_2O_3 and positive charged for n-type, e.g. SiO_2 or Si_3N_4.

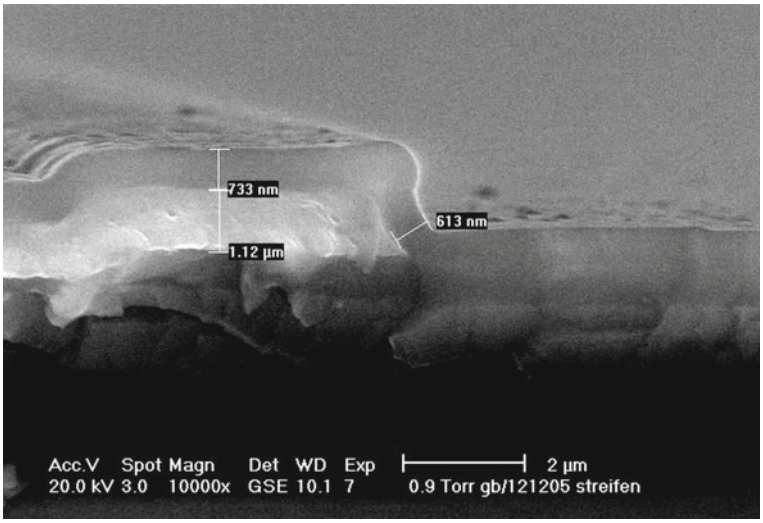


Fig. 1.40 Cut perpendicular through a single strip. The picture shows the *upper part*, especially the metal part and the passivation oxide protecting it from the environment

and the implant strip define parallel plate capacitors between themselves, between the Al strip implant complex and the backplane and the neighbours. The system is therefore described by the plate capacitor formula $C = \varepsilon_0 \varepsilon_r \frac{A}{d}$. The system requires a thin oxide d with a large width $\sim A$ to maximize the coupling capacitance. On the other hand, the width $\sim A$ must not be too large to minimize the inter-strip capacitance and capacitance to the backplane. Of course, here also the strip pitch plays an important role. The minimum oxide thickness is limited by technological application and insulation robustness considerations. A standard choice of oxide thickness is around 200 nm. The strip designs also represent a geometrical deviation from a full plate capacitor to the backplane. The full depletion voltage of a strip array is greater than the one of planar diodes. For junction-side, single-sided detectors, the increase in depletion voltage is caused by the fact that the equipotential lines are encircling the strips and are getting parallel to the surfaces only at the depth of about the pitch p. A semi-analytical solution of the Poisson equation is described in [22]. Using the strip pitch p, strip width w for p^+-doped strips, and diode thickness D, the following relation is given in [22] for single-sided sensors:

$$\frac{V_{FD,\text{strips}}}{V_{FD,\text{diode}}} = 1 + \frac{2p}{D}\left[f\left(\frac{\mathrm{w}}{p}\right)\right] \tag{1.46}$$

where $V_{FD,\text{diode}}$ is the full depletion voltage for planar diodes and the functional form f is given in [22] as

$$f(x) = 0.00111x^2 + 0.0586x^1 + 0.2400.651x + 0.355x^2 \tag{1.47}$$

For a double-sided sensor with strip width w′ for n^+ strips the function changes to:

$$\frac{V_{FD,\text{strips}}}{V_{FD,\text{diode}}} = 1 + \frac{2p}{D}\left[f\left(\frac{\mathrm{w}}{p}\right) + f\left(\frac{\mathrm{w}'}{p}\right)\right] \tag{1.48}$$

This deviation has to be taken into account when $N_{eff} \propto V_{FD}$ is extracted from a segmented sensor instead of a diode. In reality the situation is a bit more complicated, also slightly dependent on implant depth and metal to implant area (overhang or no overhang), and different for wedge-shape sensors. The equation is therefore a good starting point but a calibration measurement should be done for each new sensor type and its diode on the same wafer. The sensor capacitance to backplane per unit length follows a similar dependence:

$$\frac{C_{\text{back,strips}}}{C_{\text{back,diode}}} = \frac{1}{1 + \frac{p}{D}\left[f\left(\frac{\mathrm{w}}{p}\right)\right]} \tag{1.49}$$

For p-side strips, the inter-strip capacitance C_{int} and total strip capacitance $C_{tot} = C_{int} + C_{back}$ can be expressed as:

$$C_{tot}[\text{pF/cm}] \sim \left(a + b \cdot \frac{\mathrm{w}}{p}\right) \tag{1.50}$$

with $a < 1$, e.g. 0.75 and $1.4 < b < 2$, e.g. 1.8, depending on pitch and strip implementation details (overhang; depth). Clearly, wider strips have a higher C_{back} (area) and higher C_{int} and smaller pitchs have a higher C_{int} (distance). A real-life example is being presented in Sect. 6.4.2 on p. 245. In general n^+ strips have larger capacitance than p^+ strips since the charge accumulated between n^+ strips makes the n^+ cathode *effectively* larger. a, b in formula 1.50 are thus generally larger – outside the band given above, e.g. b can be larger than 2 or $a = 3$. This, though, can be mitigated by a decent design of p^+-stop and strip capacitance values for today's n-in-p sensors are again in the same regime as for p^+-strips.

In addition to the capacitance calculation, the detector design must take the potentials and resulting field configurations into account to guarantee a stable operation above the depletion voltage without any breakdown or micro-discharge. The Al strip implant configuration evolved over the last detector generations, going along with higher depletion voltages and potential configuration differences between single- and double-sided sensors. The breakdown field in intrinsic silicon is $V_{\text{break}} = \frac{\varepsilon E_g^2}{2qN_{eff}} \approx 30\,\text{V}/\mu\text{m}$, defining the maximum field allowed before breakdown. This dictates at least some limits on the implant profile, which should be thick and round to avoid edge or point-like structures with a high field, see Fig. 1.41a.

In the case of single-sided sensors, both the implant and Al strip can be at the same potential, most often GND. With double-sided sensors on one side, the implant

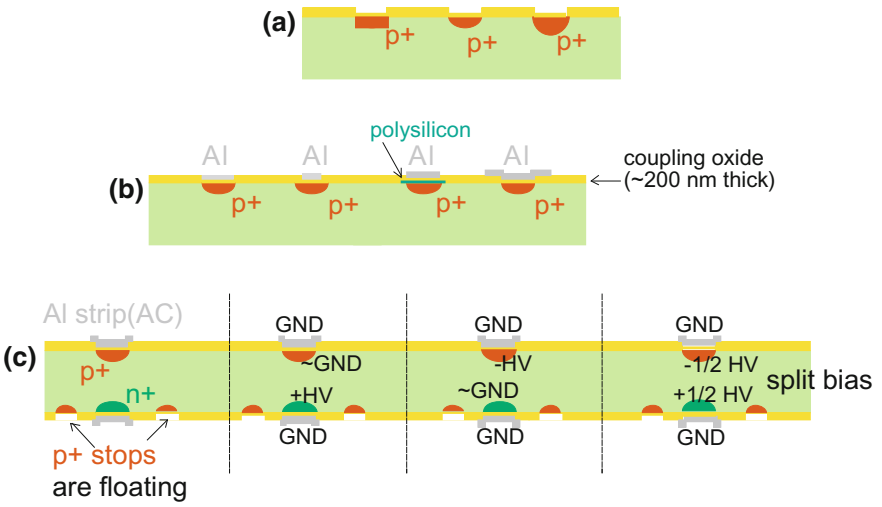


Fig. 1.41 Different strip designs. In first approximation, an AC-coupled sensor has a doped implant, a SiO_2 layer and an Al electrode connected to the readout. The common interest is to minimize the maximal E field at the *pn*-junction to avoid breakdown, by shaping the implant profile – avoid sharp edges and enlarge implant profile (this is demonstrated in part **a**). To increase coupling capacitance, the insulator oxide should be thin and Al and implant width large. Depending on the global sensor design and its operation environment, the strip layout has to be chosen very carefully to give the best possible field configuration. Double-sided sensors, having a voltage potential (see part **c**) between Al and implant, benefit from a narrower aluminium width with respect to implant. This ratio is very dependent on the needs, where a metal overhang is often sensitive to micro-discharge. In the single-sided CMS sensor, a large well-defined metal overhang was applied. With this configuration the electric field is drawn partially into the SiO_2, which has a higher breakdown voltage. A deep implementation of an additional polysilicon layer inside the oxide also moves high field into the oxide (see right complex in **b**)

to Al strip potential difference[28] equals V_{bias} (see Fig. 1.41c), resulting in a strong fringe field at the edge of the strip. This results in micro-discharges caused by a large local field. The pure bias voltage is no longer the only field defining entity; (MOS effect). In addition to this, charge-up in the oxide, which comes with poor oxide quality, poor passivation or charge-up during radiation, also influences fields locally and can lead to local micro-discharge. To avoid the MOS and oxide charge effect the Al electrode on a double-sided sensor should be narrower than the implant underneath to reduce the electric field strength through the bulk in the vicinity of the implant strip edges, since the electric field could cause the undesirable micro-discharge, a precursor to breakdown of the junction. This leads, for example in the CDF II case, to a 14 (or 16) µm Al electrode with a 18(or 22) µm implant for a sensor. Some possible configurations are displayed in Fig. 1.41b, where an implementation of doped polysilicon into the oxide was tried to shift the highest field into the oxide resulting in a 30 times higher breakdown voltage. The micro-discharge effect is

[28] Often split bias is used, where both coupling oxides carry $V_{\text{bias}}/2$ or the chip is kept floating.

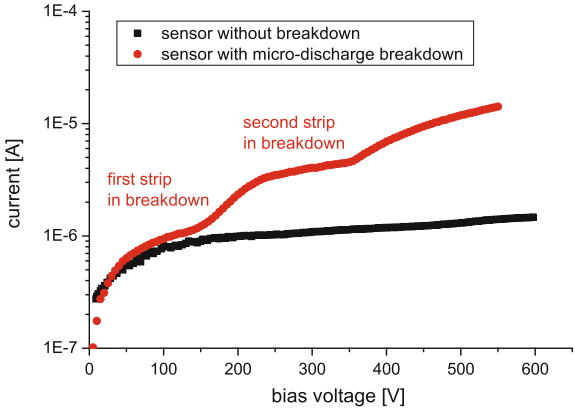


Fig. 1.42 An example of micro-discharge, the sensor shows a soft breakdown at a given voltage, the current increase is localized in a singe strip

described in detail in [225, 226]. Micro-discharge is a local problem, it also generates random pulse noise even before full breakdown. Figure 1.42 shows an example of current increase, localized in single strips at two different voltages. Of course the effects are all related to a high local field[29], which happen above V_{FD}. In a single-sided sensor, both aluminium strips and implants are at almost the same potential. The MOS effect is negligible and the Al electrode can be enlarged with respect to the implant to draw the high fields to the electrodes into the oxide. Another example of hand-tailored strip design is also described in [34, 47, 49, 77] and in Sect. 6.4.2 where the special needs of the CMS detector and its radiation tolerance needs are explained. CMS uses a 4 – 8 μm metal overhang to draw the maximum field into the SiO_2, refer to Fig. 1.41b. In short, each new geometry needs to be carefully simulated to ensure proper operation without breakdown in the given environment, while respecting needs of good coupling to inter-strip capacitance correlation; the above discussion can serve as a guideline.

In addition to the discussion on micro-discharge and capacitance correlation, a broader metal strip width can also act as an electron repellent field plate with an applied potential on the n^+-side instead of p^+-stops (see Figs. 1.29 and 4.12).

Sensor Strip Capacitances

Three different components define the strip capacitance network. The basic design strategy is to maximize charge coupling towards the readout electronic avoiding parasitic capacitances and to minimize capacitance load, to ultimately reduce noise.

Coupling Capacitance (C_{coupling})

The coupling capacitance is measured between the implant and the aluminium readout strip, which is directly connected to the readout electronics. Coupling capacitance should be large as they are directly proportional to the signal. A large capacitance means a thin isolation layer between the implant and Al strip, which is difficult to

[29] At least for p-in-n sensors, where the field builds up from the junction side.

achieve with a high reliability and without creating a short circuit between implant and readout strip, called a pinhole. Shorts or ohmic connections between Al and implant would violate the AC design, creating a DC coupling with a free bulk charge flow into the preamplifier. This limits the coupling capacitance to the manufacturing process reliability. DELPHI used integrated coupling capacitors, the value of 7 – 15 pF/cm was high enough with respect to the inter-strip and backplane capacity of 1.5 and 0.2 pF/cm, respectively. Using dry and wet oxidation (see Sect. 1.9.2 on p. 84), CDF increased the capacitance up to 140 pF per strip, a value of 20 pF/cm was specified for CDF to allow reliable production of two-layer capacitor dielectric (SiO_2 and Si_3N_4). With longer strips and broader metal strips CMS reached values between 300 and 600 pF depending on the sensor geometry – specification was $\frac{1\ \text{pF}}{\text{cm}\cdot\mu\text{m}}$. It has to be mentioned that reliable voltage resistant and thin, thus large, coupling capacitors were only technically realizable with additional nitrate layers.

Inter-strip Capacitance ($C_{\text{inter-strip}} = C_{int}$)

The inter-strip capacitance C_{int} is one of the major contributions to the capacitance load into the amplifier. It should have a certain value to share charge between two or more strips. Spatial resolution is improved by events distributing charge on two or three strips, for one could calculate the "centre of gravity". Therefore good charge sharing between direct neighbours is desired. Sometimes additional intermediate strips are introduced to have a better charge sharing, without large effect on C_{int}. In order to get most of the signal into the preamplifier, the inter-strip capacitance should be much smaller than the coupling capacitance. These two restrictions led to the specified value 1.0 pF/cm and the ratio $\frac{C_{int}}{C_{coupling}} = \frac{1.0}{20}$ for the CDF intermediate silicon layers sensors. For CMS the ratio stays between 1:30 and 1:60. This value is mainly determined by sensor layout and the strip width to pitch ratio.

Backplane Capacitance (C_{back})

To measure the strip to backplane capacitance C_{back}, the measurement instrument has to be spanned from strip to backplane, while bias voltage is applied.[30] This is a major effort for high V_{bias} because the LCR instruments are restricted to lower voltages (same as for global CV). In most cases C_{back} is approximated to $C_{\text{detector-total}}$/(number of strips) from the global CV measurement. But even without full or any bias voltage applied, the relative C_{back} measurement is an excellent identification tool for strip defects as there can be strip breaks, shorts between several strips and shorts to the implant[31]. A broken strip shows a lower capacitance, but the capacitances of both strip ends sum up to the nominal value. Shorted strips come in pairs or groups; the capacitance will then be N times the nominal capacitance. Pinholes, short-circuit readout strip to implant, result in a very high increase of capacitance. The capacitance measurement of pinholes is strongly frequency dependent and so a quasi-static instrument must be used. By measuring the values for every single strip, quality control can approximate the load uniformity of the preamplifier.

[30] Without full bias voltage, only the capacitance to the plane of the SCR would be measured.

[31] Known as a "pinhole" (see Sect. 1.6.3).

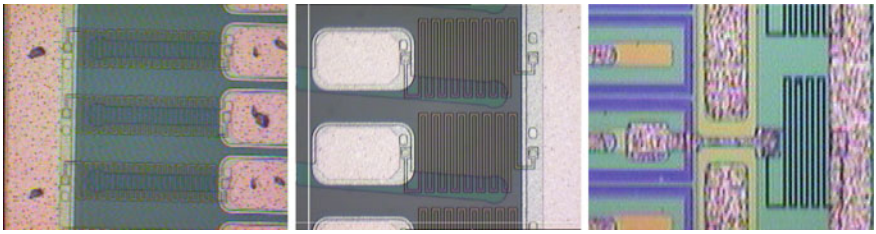


Fig. 1.43 A bias resistor made of polysilicon R_{poly}. The picture shows clearly the long and thin meander pattern of the polysilicon line, which ensures a high resistor value on a small area. The vias at the end of the polysilicon resistors to the p^+ strip as well as the vias to the bias ring are visible. Beneath the meander, the p^+ strip runs further to maximize the active area of the sensor. The pictures show a zoom of a CMS rectangular sensor, a CMS wedge-shaped sensor and a DELPHI *n*-side zoom

Strip Capacitance ($C_{\text{strip}} = C_{tot}$)

$C_{tot} = C_{int} + C_{\text{back}}$ includes all capacitances of a readout strip including inter-strip capacitances and strip to backplane capacitance. It reflects the total capacitance load to the chip amplifier. C_{tot} is a major noise source of a silicon detector. CMS limits C_{tot} to 1.2 pF/cm by defining width/pitch = w/p = 0.2, see Sect. 6.4.2.

Aluminium Strip Resistance

There is a correlation between the resistance of the metal strip and the signal pulse shape imposing a minimum acceptable resistance. Resistance R per unit length can be calculated from the width (w) and thickness (d) of the aluminium with resistivity ρ_{Al} by

$$R_{\text{strip}}/\text{cm} = \frac{\rho_{Al} = 2.824 \cdot 10^{-6}\,\Omega\text{cm}}{(d = 1.2\,\mu\text{m})(w = 8\,\mu\text{m})} \approx 30\,\Omega/\text{cm} \tag{1.51}$$

and thus relates to the specifications of the metal strip width and thickness. While the width as discussed earlier is defined by other constraints, the remaining adaptable parameter is the thickness. The example calculation in formula (1.51) shows the specification of the CDF II ISL sensors, requesting a minimum thickness of 1.2 μm. The aluminium resistance determining the strip resistance acts as a series resistor for the readout electronics. For short peaking times, high ρ_{Al} significantly influences the noise (see Sect. 1.5). It is worth noting that also DC-coupled sensors strips are fully covered by aluminium which reduces resistivity and thus noise.

Bias Resistors (R_{bias})

The bias voltage to the sensor is supplied through a set of polysilicon resistors $R_{\text{bias}} = R_{\text{poly}}$, connecting the implant strip to the bias ring, which is wire-bonded to the biasing terminal. Microscopic photos of polysilicon resistors are presented in Fig. 1.43. Noise contributions[32] from the bias resistors are proportional to $\sqrt{\frac{k_B T}{R}}$,

[32] Thermal noise.

where k_B is the Boltzmann constant, T the temperature and R the bias resistor. This dependency imposes the requirement for a high value of bias resistance. Referring to formula 1.39 on p. 41 the contribution is highly dependent on the readout frequency $\sim$ integration time t_p for experiments with long integration times, e.g. at LEP or ILC a high R_{bias} value is more important than for the LHC experiments. Variations of R_{bias} lead to voltage differences among individual strips and a non-uniform field distribution. Another upper limiting factor is the associated potential differences between neighbouring strips. Reliable polysilicon resistors with very high resistivity and only small deviations are difficult to process. The specification for the CDF II ISL sensors is $R_{\text{bias}} = 4.0 - 10\ \text{M}\Omega$ with $\pm 10\%$ variation within a single sensor. For CMS a lower value is possible due to the operation temperature of $-20°\text{C}$, the specifications are $1.0 - 2.0\ \text{M}\Omega$ with a maximum spread within one sensor of $0.3\ \text{M}\Omega$. To measure the real bias resistor, the sensor has to be at least partially depleted to have the final field configuration on the strip side, isolating the individual strips; otherwise several resistors are measured in parallel, resulting in a lower resistance. Vice versa a low bias resistance, together with a higher leakage current is an indication for a low inter-strip resistance.

With many bias resistors in parallel the global inversely summed up resistance is negligible compared to the bulk resistance (initially several GΩ and still many MΩ after radiation) – R_{bias} has no significant effect on V_{bias}. The individual bias resistor, though, can be determined by measuring the voltage drop ΔV over the resistor and the strip leakage current I_{strip}, and taking the bulk resistance into account (from an IV value above V_{FD}). But, in most quality assurance setups, a defined voltage V_{set} is imposed over the bias resistor, the current is measured and $R_{\text{bias}} = V_{set}/I_{measured}$ is calculated. Applying too high a voltage may disturb the fields, lowering inter-strip resistance and thus result in too low measured R_{bias} values – CMS had to change the measurement voltage from 10 to 2 V. For irradiated sensors the single-strip current, in the same order as the imposed current, has to be taken into account.

$$R_{\text{bias}} = \frac{V_{set}}{I_{\text{measured}} - I_{\text{strip}}} \tag{1.52}$$

The polysilicon resistor processing is explained in Sect. 1.9.2, especially Fig. 1.61. In reality the company establishes the final design, the customer defines value and size. With an estimate on the value of the sheet resistance[square][33], a meander structure can be designed.

To design a sensor, high values of R_{bias} are preferred for their low thermal noise, especially long ladders must have high value resistors because they all add up parallel, therefore the readout electronics see a lower value and therefore higher noise. On the other hand large resistors increase the RC time constant. For highly irradiated sensors, the formerly very high (order of GΩ) bulk resistivity drops and a good fraction of the applied voltage drops then on R_{poly}, an undesired effect, which would require a lower value. Here, R_{bias} also limits the strip current. With a high current,

[33] A square is a defined square size to allow normalization of a resistance per area.

e.g. after radiation, R_{bias} defines the potential $V = RI$ of the implant with respect to the aluminium strip. This should be taken into account when designing the readout electronics.

Inter-Strip Resistance $R_{\text{inter-strip}} = R_{int}$

A high inter-strip resistance is necessary in order to accomplish the isolation of all individual strips. R_{int} should be in the order of several GΩ. Obviously, this value is not applicable with doped silicon at non-cryogenic temperatures. It is only achievable with a fully depleted surface at the Si – SiO_2 interface reaching deep into the bulk volume. Inter-strip resistance is a very delicate parameter, sensitive to all kinds of oxide charge or even charge concentration in the passivation oxide spoiling the surface depletion to surface inversion or accumulation. Charge-up of the surface oxide will always affect R_{int}.

In Sect. 2.3, it is illustrated how radiation can decrease the inter-strip resistance ending up in values of several tens to hundreds of MΩ. The most complicated part is the R_{int} measurement in the presence of bias resistors, in the order of several MΩ, connecting all strips to the bias ring. This can be principally solved in applying a small voltage ramp on one implant strip (via the DC pad) and measuring the resulting current on the neighbouring DC pad, while both voltage supply and amperemeter share the same ground, namely the bias ring. Unfortunately only a small voltage on the order of about 1 V can be applied without disturbing the depletion layer configuration, thus short circuiting the strips. Applying 1 V over a resistance of 1 GΩ, the resulting current of 1 nA is in the order of the leakage current. With resistance values of 10 – 100 GΩ or high strip leakage current, e.g. after irradiation, the measurement current cannot be resolved any more, therefore the method can only give a lower limit for R_{int}.

Strip Faults

The most common strip faults are introduced, together with a measurement example and their effect on the readout electronics. There is always a compromise of defective strip fraction e.g. in CMS it was 1% per sensor to have a maximum of 2% bad strips on a two-sensor module. Long modules, with 10 sensors daisy-chained, need a much stricter criterion and vice versa a very simple and standard sensor design. A specified request of zero strip faults would decrease the numbers of accepted sensors, thus the yield. This would result in substantially higher cost. In the case of the daisy-chaining of several sensors, the sensor containing a strip fault should be placed as far from the readout hybrid as possible. This is especially true for breaks and pinholes, where the strip should be disconnected. The basic faults are illustrated in Fig. 1.44 as well as photos of shorts and breaks.

Pinholes

Pinholes are shorts or low ohmic connections between an aluminium readout strip and the implant. Charge can flow freely into the preamplifier, possibly overloading the amplifiers[34] of a large part of a chip or even of one whole chip (refer to [137] and

[34] Strong dependence on the ASIC design.

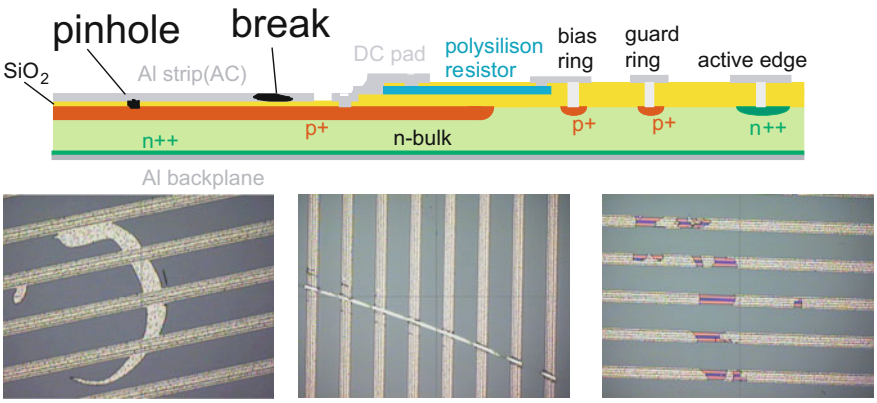


Fig. 1.44 The location of pinholes and breaks are shown in the sensor scheme. Shorts are displayed in the first two photos, one results from a processing fault, whereas the second came from wrong handling. The breaks are simple process faults, but are often also results from scratches

[136]). These strips must be identified if not avoided in the manufacturing process and remain disconnected from the amplifier. Commonly the following strategies are used to identify pinholes:

- applying several volts over the insulation between Al readout strip (AC pad) and the implant (DC pad). Simultaneous measurement of the current identifies all shorts and measures the resistance of low ohmic connections. Pinhole signatures are mA signals on a sub pA background for good strips.
- high C_{coupling} hints the presence of a pinhole but for an LCR device C_{coupling} measurement is also frequency dependent. With quasi-static devices[35], measuring $Q = CV$ in steps, C_{coupling} is out of range for pinholes.
- strip to backplane capacitance C_{back} also identifies pinholes quite easily. In this case the operation voltage is applied via the LCR metre, which cannot stand high voltages. Of course, a decoupling box could be applied but then the above-mentioned methods are as effective and easier. Therefore, the method is only easily applicable for low depleting sensors.

Finally in case of a pinhole, the wire connection between strip and preamplifier has to be pulled. In the case of the SVX4[36] chip, single channels can be even disconnected later, when configuring on chip level (refer p. 97).

Breaks

Breaks in the aluminium are no danger to the readout, but will disconnect the rest of the strip and therefore all strips of the next sensors in a line. Breaks in the implant, as

[35] A quasi-static capacitance measurement, where $Q = CV$ is measured for several values of V, works only with negligible current. In the case of a pinhole, the applied voltage V initiates a current driving the measurement value into overflow – pinhole signature.

[36] The SVX4, the successor of the CDF SVX3 chip.

long as they to not create a point-like defect increase I_{strip} are almost invisible within AC-coupled sensors since the charge will still fully couple to the Al-line. With a higher implant than Al resistance a complete implant line cannot compensate a break in the Al line. Sensors with Al breaks should be placed at the end of the ladder (like pinholes). Breaks are identified by comparing the current strip capacitance with the nominal strip capacitance. Strips with lower capacitance are candidates for breaks. For the DELPHI sensors, the capacitances of both strip ends were measured, where sum of the two values equals the nominal capacitance $C_{\text{left}} + C_{\text{right}} = C_{\text{nominal}}$. If the sum is equal to the nominal value the aluminium strip is broken.

Shorts

A short is given by a connection of one or more neighbouring strips. This is not a crucial fault, but will result in a higher load capacitance which associates a higher noise and also generates multi-strip hits, degrading the resolution. This kind of fault has no preferred position in the sensor row. Shorts are found by measuring C_{strip}. Capacitances which are N times the nominal capacitance should come in groups of N members. $C_{\text{strip}} = N \cdot C_{\text{nominal}}$ for N strips shorted together.

Leaky Strips

Strips with a high leakage current I_{strip} can increase the channel noise dramatically. In combination with a pinhole, a not so rare case, the current can even overload the whole chip. There are many reasons for leaky strips, e.g. defect bias resistors, inhomogeneity in the bulk silicon especially in the surface region, faulty strip implantation, non-depleted regions, etc. (refer Sect. 1.14). These strips often come in groups, which means a whole region has some defect. Very high currents increase the shot noise. In some cases, leaky strips even have a stochastic time structure, e.g. resulting from micro-discharge and therefore can produce noise deviating from standard shot noise. During quality control leakage current will always be measured at the maximum operation voltage, e.g. 80 V for CDF or 400 V for CMS; high current average or localized outliers are critical. Some examples of high strip leakage currents are presented in Sect. 1.14.

1.7 Practical Aspects of Handling and Testing Silicon Strip Devices

One reason to use silicon sensors is their robustness. SiO_2-passivated silicon sensors are impervious to their environment in some well-defined boundaries. Sensors do not degrade even after resting in cupboards for years. A clean room is sufficient for sensor testing and module assembly; no dedicated rated cleanroom of better than class 100,000 is necessary.

On the other hand sensors are thin, brittle and fragile. The passivation layer is only around 1 – 2 μm thick and provides almost no protection against mechanical damage. Scratches from bad handling or heavy dust during transportation degrade

sensor properties, refer to Sect. 1.14. Also probe needles with probe tips around 2 – 7 μm need to touch the aluminium pads and the risk of damaging the pad or even to pierce through the metal is not negligible. Sensors need to be handled with care and some healthy paranoia not to destroy them. Vacuum tweezers are standard tools to avoid any direct human contact. The cleanliness of the sensor surface is also very important for the later wire-bonding. Remnants, e.g. fat from a human fingerprint, can already significantly degrade bondability. Humidity levels in the laboratories have to be in well-defined ranges; very low humidity increases the risk of static electric discharge and also charge-up of the sensors; a high relative humidity level (>30%) can lead to increase of leakage current, as silicon is hygroscopic. A standard method is to allow a normal humidity (40 – 50%) in the laboratory while sensors are stored in a dry cabinet and sensor probe stations are flushed with dry air during testing. Charge-up of the passivation oxide can affect the Si – SiO_2 interface layer, refer to Sect. 1.14.

There are some dependencies, which can be explained more by "10-years-gut-feeling" and experience rather than by some mathematical formulas.

- sensors have to be stored in a dry and dark environment
- sensors have to be handled with vacuum tweezers
- face masks have to be worn, the human breath contains some acids, which can degrade sensor performance
- sensors are tested under low relative humidity (<30 – 40% RH)
- testing should be one's second nature but a certain paranoia helps to prevent stupid faults out of carelessness
- if a measurement looks "strange" repeat it; sometimes the sensor needs to be conditioned, e.g. some charge-up has to be drained off
- most often electric defects are reflected in optical defects
- strange results often reflect wrong measurement parameters, e.g. too high voltage detoriating surface depletion for inter-strip resistance measurement (Sect. 1.1.3)
- bad contacts are often related to insufficient opening of the aluminium pads and remains of SiO_2 are good insulators. In this case bonding also becomes more difficult
- a stabilized temperature is mandatory for measurements of irradiated sensors to
  - get reasonable, understandable and comparable results
  - avoid thermal runaway, which may even destroy the sensor
  - measure the α parameter (see Sect. 2.1), which is strongly temperature dependent.

1.7.1 What Is the Standard/Exhaustive Set of Quality Assurance Tests?

Sensors for HEP detectors are procured from industry. The quality assurance by the producer is very strict and part of the contract. A strip yield above 98 – 99% is a standard requirement. Furthermore, maximum leakage current and a range for the depletion voltage is requested. In addition all the above-discussed strip parameters,

like inter-strip capacitance, inter-strip resistance, bias resistance and coupling capacitances are specified. Furthermore more process-related parameters, like flat-band voltage, breakdown voltage, aluminium resistivity, dielectric breakdown voltage, etc. are specified. Process parameters are only tested on sample basis and mostly on dedicated structures.

Depending on the number of sensors, sample testing is appropriate after preproduction is sufficiently evaluated. For example, CMS asked for 5% of the full production before placing the final full contract. Without full testing there is always the chance to spoil the full module when assembling several sensors and the readout hybrid (each potentially faulty).

The minimum tests are

- global current–voltage characteristic (IV)
- full capacitance–voltage ramps (CV) to determine depletion voltage

Quality Assurance on Sensor Level

For strip testing, there are different more or less exhaustive strategies but some standard sets of tests are

- global tests:
  - IV scan $(0 - V_{IV_{max}})$, where $V_{IV_{max}}(\text{CDF}) = 100$ V and $V_{IV_{max}}(\text{CMS}) = 550$ V (or 800 V)
  - CV scan $(0 - V_{CV_{max}})$, where $V_{CV_{max}}(\text{CDF}) = 100$ V and $V_{CV_{max}}(\text{CMS}) = 350$ V
- common strip tests:
  - leakage current of each strip at V_{strip} (to determine noisy, leaky strips), where $V_{\text{strip}}(\text{CDF}) = 80\,\text{V}$ and $V_{\text{strip}}(\text{CMS}) = 400\,\text{V}$
  - bias resistance of each strip (R_{poly})
  - I_{diel} current over C_{AC}, where $V_{diel}(\text{CDF}) = 60\,\text{V}$ and $V_{diel}(\text{CMS}) = 10\,\text{V}$ to identify pinholes
  - coupling capacitance C_{coupling} with quasi-static C-metre or low LCR frequency ($\sim$100 Hz)
- measurements on sample, problem basis or on dedicated structures
  - inter-strip resistance
  - inter-strip capacitance at high LCR frequency ($\sim$1 MHz)

To guarantee the quality, every single strip has to be tested to identify the percentage of bad strips. For double-sided sensors, all tests have to be done on both sides.

Process Monitoring

Until the CMS experiment, process parameters were checked on first prototypes together with the producer or not at all. For the 24328 sensors of CMS, a new concept was developed to allow an easy measurement of silicon device parameters and to efficiently monitor the production process [35, 183]. A set of test structures was designed and placed on every wafer for all sensor types. Figure 1.45 shows the cutoff

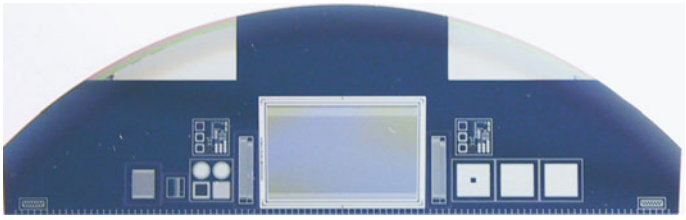


Fig. 1.45 Half-moon with test structures. *Left* to *right*, the structures with the following accessible process parameter are located (1) TS-CAP (C_{coupling}, V_{break}). (2) sheet ($\text{Resistivity}_{\text{Al}}$, $\text{Resistivity}_{\text{p+}}$, R_{poly}). (3) *G*ate *C*ontrolled *D*iode GCD (I_{surface}). (4) CAP-TS-AC ($C_{\text{inter-strip}}$). (5) full minisensor. (6) CAP-TS-DC ($R_{\text{inter-strip}}$). (7) Diode (V_{FD}). (8) MOS1 ($V_{\text{flat-band}}$). (9) MOS2 ($V_{\text{flat-band}}$)

from a 6 in. wafer with the test structures. These test structures also allow destructive tests to be done, like breakdown voltage of the coupling capacitors. A proposed set of measurements is leakage current and breakdown voltage measurement plus determination of depletion voltage on a minisensor comparable to the main sensor plus tests on standard diodes. Flat-band voltage and surface currents can be extracted on MOS and special structures called **G**ate **C**ontrolled **D**iodes GCD. Inter-strip parameters like resistance and capacitance plus aluminium and implant resistivity and bias resistance can be taken from individual dedicated structures. Of course, this effort is reasonable, if the measurements and analysis are fully automated. An example of parameter evolution versus time is presented in Fig. 1.88.

Irradiation Quality Control

With radiation becoming an issue for today's and future detectors in a collider experiment, radiation hardness studies are more and more important. For CMS a dedicated radiation hardness quality control concept was set in motion to irradiate around 5% of the above-described test structures plus about 0.5% of the received sensors and even a couple of fully assembled modules. Proposals for pre- and post-radiation tests at nominal operation temperature on sensors and minisensors are

- mandatory tests on sensors
  - IV and CV (determine V_{FD} and compare with model)
  - inter-strip capacitance and resistance
- additional measurements
  - coupling capacitance
  - bias resistance
  - strip leakage currents
- measurements on modules
  - noise
  - signal-to-noise

Unlike sensor quality control, measurement of all strips is not necessary, the difference of pre- and post-radiation values are important. To get an even deeper understanding, all parameters, also strip parameters, are measured for several bias voltage values to better understand the behaviour below and above depletion voltage. After irradiation often over-depletion is even more important than before. The CMS irradiation quality control example is described in detail in [80, 111].

Pixel Sensor/Module Quality Control

For pixel sensors with the much higher channel count and much lower cell values, e.g. load capacitances, the only reasonable test on sensor level is the global IV-scan, sometimes also a global CV-scan. All further quality assurance is being done after bump bonding to the readout chip. Often, so-called bare module tests are done connecting a probe card to the chips reading out the bare sensor-chip sandwich. With a reasonable yield of all components real quality assurance testing often only happens after full pixel module assembly with the final readout. These tests are then very thorough including additional IV-scans, full chip characteristics, thermal cycles, X-ray exposure and high rate tests.

1.8 R&D Methods and Tools: DLTS, TSC, TCT, Edge TCT, TPA-TCT, SIMS and Simulation

Several specialised expert methods complement the above described tools to have a more complete understanding of new materials or new sensor concepts on a more fundamental R&D level. These tools allow insight on fundamental damage mechanisms on crystal level and also give better judgement on dynamics of sensors before and after irradiation. The goal is to use the results to "*defect engineer*" future detectors.

The following paragraphs introduce the basics of these tools with enough detail to understand the usage and presented results. The real effort to finally exploit, setup, understand, debug and calibrate any of these expert systems is worth a Master or doctoral thesis. As a matter of fact the currently most detailed descriptions are to be found in doctoral theses.

Deep **L**evel **T**ransient **S**pectroscopy DLTS and **T**hermally **S**timulated **C**urrent TSC are tools to identify the defect levels in the band gap. DLTS has a higher resolution while TSC allows to probe higher defect concentration thus higher irradiated samples. The **T**ransient **C**urrent **T**echnique TCT gives information about the electric field profile therefore the effective space charge profile (and sign) given by initial doping or defect concentrations resulting from radiation; it is also a means to evaluate effective trapping times. **S**econdary **I**on **M**ass **S**pectrometry SIMS is a destructive way to determine concentration of dopants and impurities, a more standard tool used in industry and solid state labs.

These days, these tools are finding their way into evaluation of base material and sensor design or company qualification up to the mainstream standard quality

assurance or process monitoring. Dedicated examples and results will be presented in Sects. 2.2 and 7.1.

In addition to the tools described in the next paragraphs, there are:

- PITS (**P**hoto **I**nduced **T**ransient **S**pectroscopy)
- FTIR (**F**ourier **T**ransform **I**nfrared **S**pectroscopy)
- RL (**R**ecombination **L**ifetime Measurements)
- EPR (**E**lectron **P**aramagnetic **R**esonance) and ESR (**E**lectron **S**pin **R**esonance)
- PC (**P**hoto **C**onductivity Measurements)
- TDRC (**T**hermally **D**ielectric **R**elaxation **C**urrent technique) similar to TSC but used to study surface defects in the oxides.

1.8.1 Deep Level Transient Spectroscopy – DLTS

Deep **L**evel **T**ransient **S**pectroscopy DLTS is a widely used method to characterize deep level defects (energy levels) in semiconductors. After a filling of trap levels by light injection or by a voltage pulse, the following relaxation transients, either current (I-DLTS) or capacitance (C-DLTS) are measured. It is a very precise method and can derive the defect concentration N_t, their thermal activation energy E_a and their cross-sections σ_e and σ_h. The measurement process and data treatment is complicated; the basics will be described in a superficial way here and for a more detailed description, e.g. how the digital data is being refined, the reader is referred to [187] and [219]. We will concentrate on C-DLTS since later presented results have been obtained with this method. C-DLTS has been initially introduced by D.V. Lang [187] at the Bell Laboratories and has been patented by him. Today these systems are commercially available and in use in several labs dedicated to sensor R&D in HEP. The method has some limitations: It works only for trap densities $N_{trap} = N_t$ much lower than the shallow doping concentration $N_{Dopant} = N_D$ ($N_t \ll N_D$), respectively the method only works for low radiation fluences below $\Phi = 10^{11}$ to $10^{12}\, n_{1\mathrm{MeV}}/\mathrm{cm}^2$ for the high resistivity sensors used silicon detectors. Different to other methods, it is possible to distinguish between electron and hole traps directly by looking at the transient signal polarity and/or by different trap filling methods.

The basic principle is illustrated in Fig. 1.46 for a p-in-n diode. For better illustration we split the electron and hole trap configurations in two columns. Initially a reverse bias voltage V_R is applied to remove all free charge carriers from the depletion zone/volume $\boxed{1}$. In this case all majority carrier traps[37] in the depletion zone above the Fermi level E_F are empty while the minority carrier traps are still filled. The traps are then filled by pulsing the bias voltage $\boxed{2}$ thus changing their electric charge state. With a pulse still below 0 V only majority carrier traps (here electron capture) are filled while higher pulses respective forward polarity voltage pulses also fill the minority traps (here hole capture). The filled trap condition changes the space

[37] Reminder: Majority carriers for n-type bulk: electrons, for p-type bulk: holes.

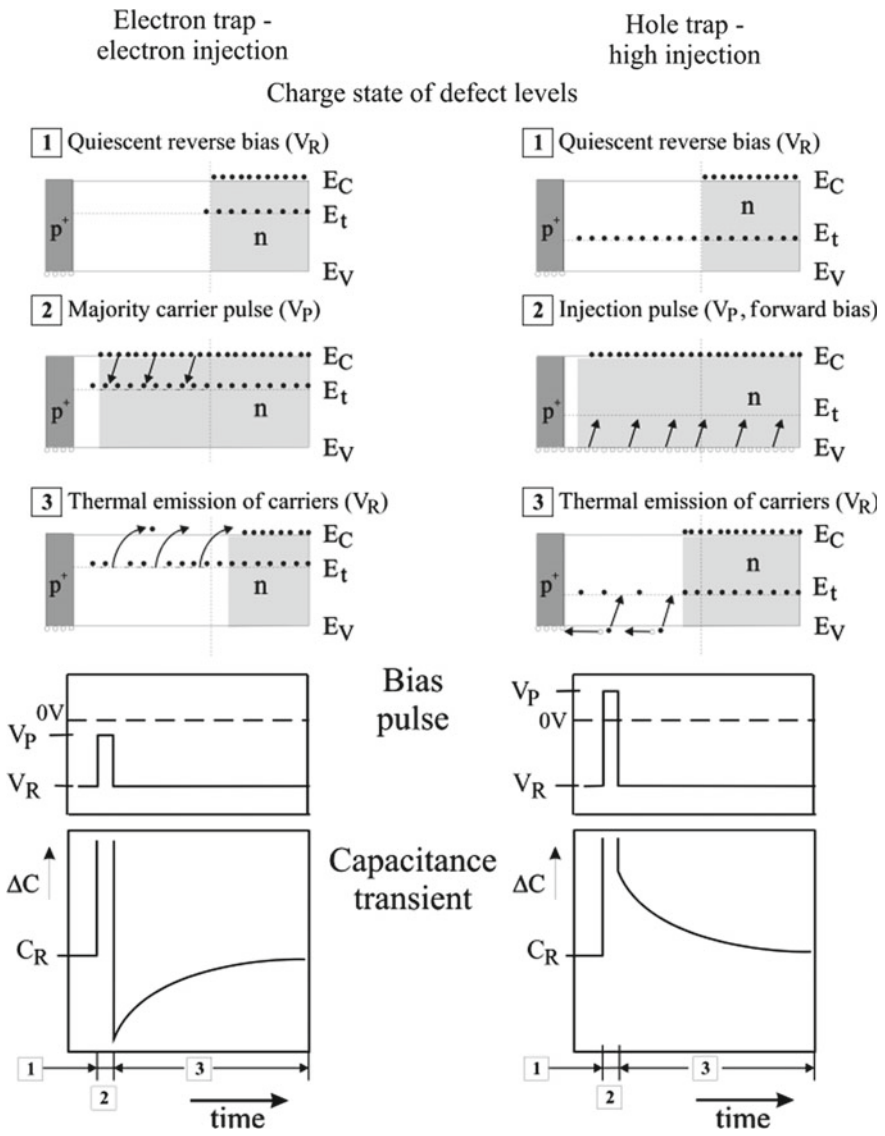


Fig. 1.46 DLTS – The Principle of operation for a *p*-in-*n* diode. The *left* column illustrates the case of electron traps, the *right* one for holes. 1 shows the reverse bias case with V_R applied and a depletion zone established. In 2 a voltage pulse is applied; *left* not passing 0 to only fill the majority carrier traps and *right* up to forward polarity also filling minority traps. The voltage and capacitance values are schematically shown in the lower plots timely for 1, 2 and 3. The measurement is taken after the voltage pulse 3 when the captured carriers are emitted due to thermal excitation energy and "being removed" by the re-applied reverse bias voltage. The information is in the temperature dependent exponential emission behaviour $\Delta C(t)$. Schemes from [219]

Table 1.4 Light penetration depths in silicon for different wavelengths at T = 300 k. A TCT setup normally features a red laser to deposit charges only on one sensor face plus an IR laser to penetrate the full volume. Electron–holes creation with an IR laser mimics closely the traversing of a minimum ionising particle. In some cases also a 980 nm laser is used, e.g. when a deep backside diffusion or a handle wafer has to be passed to deposit charges at the face of the effective active volume. Values from [120]

Light wavelength λ [nm]	405	640	980	1064	Below band gap energy
Penetration depth [μm]	0.1	3	100	1000	Meters (transparent)
	Surface	Near surface	μ-beam	MIP	No single photon absorption

charge of the system thereby decreasing and/or increasing the system capacitance for electron capture and/or hole capture, respectively.[38]

After the pulse with V_R applied again plus adequate thermal excitation energy, the filled traps are de-trapping [3]. The electron and/or hole emission results in a corresponding change of system capacitance – the exponential transient $\Delta C(t)$ we are recording. This transient is obviously temperature dependent. With higher trap concentrations than $N_t \ll N_D$ the ΔC transient would dominate the base value of C_R (at reverse bias) and not exhibit an exponential behaviour. Now the different polarity of capacitance change (rise or fall) shows the respective trap configurations. The information is in the exponential time constant of the transient which, again, is also temperature dependent. A main challenge is the measurement of small capacitance changes. Another way to fill traps in a controlled way is via light injection where an infrared pulse (light passes through the whole bulk) fills both kinds of traps.

It is also possible to fill only one type of traps, using a red or green laser (short wave-length) with a very shallow penetration depth (some µm, see Table 1.4 on p. 72). Depending on the face of illumination, one charge carrier type is swept away immediately while the others passes the whole bulk thereby filling the corresponding traps (reverse bias voltage applied). Shining on the junction side (p-electrode) holes are removed immediately and electrons fill the traps in the bulk. Shining on the ohmic side (n-electrode) holes are traversing the bulk filling hole traps exclusively. Now, how to measure the emission transient; how to evaluate it and how to extract the relevant information from it?

With some approximations the trap concentration N_t is basically proportional to the transient amplitude $\Delta C_0 = C(t = 0) - C_R$

$$N_t \approx 2N_D \frac{|\Delta C_0|}{C_R} \tag{1.53}$$

[38] In a graphic way one can say the additional space charge of the filled traps decreases/increases the depletion zone thus the full capacitance.

with N_D the doping concentration and C_R the global capacitance at full reverse bias voltage. The limit to detect trap concentrations therefore depends on the sensitivity and accuracy of the measurement device. For example, for

$$\Delta C_{0,min} \approx 5 \text{ fF}, C_R \approx 50 \text{ pF} \rightarrow (N_t/N_D) \approx 2(\Delta C_{0,min}/C_R) \approx 2 \cdot 10^{-4}$$

The time constant τ_t of the exponential decay is inversely proportional to the emission rate constant $e_{e,h} = \frac{R^{e,h}}{(n,p)_{dl}} = \tau_{e,h}^{-1}$ of electrons/holes of the defect levels (rf. formula 1.12). The transient itself follows the concentration of occupied traps $n(t)$, decreasing with carrier emission, and therefore has the form

$$\Delta C(t) = C(t) - C_R = \Delta C_0 \cdot e^{-\frac{t}{\tau_t}} \propto n_t(t) = N_t e^{-\frac{t}{\tau_t}} \tag{1.54}$$

from which $\tau_t = \tau_{e,h}$ can be derived.

With formula 1.12 from Sect. 1.1.1 we get the temperature dependence

$$\frac{1}{\tau_{e,h}} = e_{e,h} = \frac{R^{e,h}}{(n,\, p)_{dl}} = \sigma_{e,h} \cdot v_{th,e,h} \cdot N_{C,V} \cdot e^{-\frac{E_{a,e,h}}{k_B T}} \tag{1.55}$$

with $R^{e,h}$ full emission rate and $(n,\, p)_{dl}$ concentration of deep level (trap) states.

We get the activation energy $E_{a,e,h}$ as the inverse slope and the trap cross-section $\sigma_{e,h}$ as the x-intercept from the corresponding linear Arrhenius plot[39]

$$\ln \tau_t v_{th,e,h} N_{C,V} = -\ln \sigma_{e,h} + \frac{E_{a,e,h}}{k_B T} \tag{1.56}$$

The measurement cycle consists of a temperature scan with steps of about 1 K and smaller steps at interesting points of emission peaks. At each temperature point several fill and transient cycles are recorded (1 – 3). Several methods and/or correlations are being used to process the analogue signal often in combination. For example the double boxcar integrator used in the original system by Lang in 1974, simply measures the capacitances at two defined times t_1 and t_2 (a well-defined interval) giving two distinct capacitance values $C(t_1)$ and $C(t_2)$. Their difference $\Delta C_{1,2} = C(t_1) - C(t_2)$ is plotted versus temperature. At low temperatures, the emission is suppressed, thus the transient is too slow to record while at high temperatures, the emission is too fast to record and $\Delta C_{1,2}$ is zero/small. When the emission time at the corresponding temperature (corresponding time constant τ_t) matches the time window, $\Delta C_{1,2}$ exhibits a maximum. Another method is to vary the time window at a given temperature. Digital signal processing various correlator functions (e.g. sine) or Fast Fourier Transformation or Laplace Transformation are used.

An important example of an DLTS plot is shown in Fig. 2.13 on p. 152.

[39] Arrhenius plot: Plot the exponential temperature dependent decay $e^{-\frac{E_A}{k_B T}}$ in logarithmic scaling against the inverse temperature.

In a nutshell, in DLTS

- a temperature scan is being done
  - traps are filled and the emission is recorded several times at each temperature step (emission visible in the change of system capacitance)
- transient gives trap concentration N_t, its cross-sections σ_e and σ_h and activation energy $E_{a,e,h}$
- is very precise
- works for low concentration of traps up to radiations of $\Phi_{eq} = 10^{11}$ to $10^{12}\ n_{1\text{MeV}}/\text{cm}^2$
- trap filling can be steered (fill electron or hole or both traps)
- transient gives information of trap type (electron or hole)

With DLTS, even the N_t depth profile can be measured by variation of fill pulse or different starting reverse bias voltage levels (test different depletion voltage depths).

1.8.2 Thermally Stimulated Current – TSC

The **T**hermally **S**timulated **C**urrent technique TSC has been already applied to semiconductors the first time in 1968; further reading see [48, 353]. TSC works for higher defect concentration levels than DLTS and thus higher radiation levels. An important example is shown in Fig. 2.12 on p. 152. The basic steps of TSC are:

1. Cooling
   The sample is usually cooled to a low temperature while under reverse bias where traps stay empty.[40]
2. Filling
   At the low temperature, the traps are filled by forward bias pulses or light injection – same procedure as used in the DLTS system. As before, with IR light injection from one sensor face, only one trap type is filled.
3. Recording
   The emission current is then recorded while heating the sample with a constant heating rate resulting in the TSC spectrum. At specific temperatures, specific traps located at these energy levels emit their charges showing peaks – same as for the DLTS.

Measuring only currents, the TSC cannot distinguish between charge carriers. The difference between TSC and the DLTS method is that traps are filled only once at low temperature and the spectrum is recorded in a single temperature scan. DLTS and TSC, being similar in equipment needs, and are often combined in a single setup; just add an electrometer for the TSC to the capacitance meter being used for DLTS. With an applied constant heating rate of $\beta = \frac{dT}{dt}$, the temperature dependent emission rates $e_{e,h}$ are time dependent ($T \sim t$). Heating rates are about $\beta(t) = 0.1$ k/s.

[40] If cooled at zero bias voltage the majority carrier traps are already being filled during the process. In case of a very high majority trap level concentration with respect to the majority carriers itself, only the traps near midgap are being filled.

The measured current for a fully depleted volume is given by

$$I_{TSC}(t) = \frac{q_0 \cdot A \cdot D}{2} e_{e,h}(t) \cdot n_t(t) = \frac{q_0 \cdot A \cdot D}{2} e_{e,h}(t) \cdot n_{t,0} \cdot e^{\left(-\int_0^t e_{e,h}(t')dt'\right)} \tag{1.57}$$

with $n_t(t)$ also being time dependent [106]. A is the diode surface and D the diode thickness. The main interesting parameter is $n_{t,0}$ the absolute fraction of occupied traps during the filling process ($n_{t,0} = N_t$ if fully filled). Time integration[41] over a given peak gives $Q_t = \int I_{TSC}(t)dt$ and therefore

$$n_{t,0} = 2\frac{Q_t}{q_0 \cdot A \cdot D} = N_t \tag{1.58}$$

Therefore $n_{t,0}$ can be derived from the peak integral (it is also possible to derive it from the peak height). Also in the TSC case, to achieve a good resolution, the real measurement is more complicated than obvious at first glimpse.

For the plots in this book the delayed heating method has been used. A temperature starting point T_0 (near to a previously identified peak) is chosen, the traps are filled and the temperature ramp is then executed several times but with different delays before starting. With emission happening right away after filling, the peak heights (also integrated current during the ramp) are lower the longer the initial delay; or said differently the longer the delay the more traps emptied before the start of the ramp at $T = T_0$. Repeating the measurement for several T_0, one can then derive the trap cross-section $\sigma_{e,h}$ and activation energy $E_{a,e,h}$ from the Arrhenius plot of peak height versus delay time.

Other methods would be to vary the heating rate or in more complicated situations, where different peaks overlap more sophisticated deconvolution methods are applied, e.g. numerical least square fits on the whole spectrum.

In a nutshell, TSC:

- measurement cycle
  - device under test is cooled down to low temperatures; most often under reverse bias;
  - traps are filled by voltage pulse or light pulse;
  - traps are emptied during a temperature ramp with the global current as emission signal
- can be used at higher defect concentration thus higher radiation levels
- trap concentration, trap cross-section $\sigma_{e,h}$ and activation energy $E_{a,e,h}$ can be derived.

[41] For practical reasons with a constant heat rate time can also be substituted with temperature $T = T_0 + \beta(t)$.

1.8.3 Transient Current Technique – TCT

The first **T**ransient **C**urrent **T**echnique TCT measurements were made in the 60ties [14, 185], and since 1990 these systems are widely used in HEP to investigate irradiated sensors, see examples at [93–95, 108, 176]. Today TCT systems are commercially available for R&D but also for educational purposes. These systems are powerful expert tools able to derive the velocity *profile* and thus the electric field and effective space charge *profile* of the silicon device before and after radiation. In addition, depletion voltage, charge collection efficiency, space charge sign and effective trapping times can be derived. Again only the basics of the technique will be described underlined by examples allowing to understand results commonly presented. For more details and further examples the reader is referred to [80, 89, 92, 171, 172]. From my colleague and world expert Gregor Kramberger:

> What makes a Transient Current Technique such a powerful tool is its simplicity. Yet, it allows for so many different usages which make it unique in its reach. Virtually all detector and material properties can be explored with it. If you are looking for a single tool that will give you the insight into semiconductor detectors that is it.

In a time-resolved manner, as described in Sect. 1.3.2, signal formation follows the drift velocity of the charge. The Transient Current Technique exploits this behaviour, measuring the induced signal currents with a fast oscilloscope in a fine time resolved way and mapping them to the electric field using the relations below

$$I_{e,h}(t) = -N_{e,h} \cdot \boldsymbol{E}_W \cdot \boldsymbol{v}_{e,h}(t); \text{ with } I(t) = I_e(t) + I_h(t) \tag{1.59}$$

with $\boldsymbol{v}$ the charge velocity driven by the electric field, the weighting field $\boldsymbol{E}_W$ and $N_{e,h}$ the injected charge carriers. For a diode the weighting field $E_w = \frac{1}{D}$ is constant (similar to the electric field in a plate capacitor without space charge $E = \frac{V}{D}$) resulting in the following relation

$$I_{e,h}(t) = -N_{e,h}\frac{1}{D} \cdot \boldsymbol{v}_{e,h}(t) = \frac{N_{e,h}(0)}{D} \cdot \mu_0 \boldsymbol{E(r(t))} \tag{1.60}$$

with diode thickness[42] D and non constant electric field $\boldsymbol{E(r(t))}$. The current is directly proportional to the electric field. The maximal field itself $E = \frac{V_{\text{bias}}}{D}$ is varied by changing the external bias voltage (rf. Fig. 1.8 in Sect. 1.1.2) and is linear over the sensor depth.

The charges (electron–hole pairs) are generated in the active volume of the sensor by injection of light (laser). The electron and holes then drift to their corresponding electrodes. In principle charges can also be generated by α- or β-radiation but with a laser the exact timing (triggering) is known, the measurement can be repeated several times in a controlled way and also the power (amount of initial charges) can be tuned

[42] The relation is correct for a fully depleted diode otherwise D becomes w, the thickness of the depletion zone.

(controlled). Using different laser wavelengths the location of charge creation (penetration depth) can be selected as well. With a red laser all charges are created near the surface and only one type participates in the drift. With an IR-laser electron–hole pairs are created in the whole volume (for standard sensor thicknesses). Table 1.4 gives the different wavelengths versus the penetration distance in silicon. For example in a p-in-n diode, light is injected at the n-side, the electrons are immediately[43] collected by the n-electrode while the holes drift for several nanoseconds towards the p-electrode along a linear growing electric field therefore gaining velocity $\boldsymbol{v}_{\text{drift}} = \mu_0 \boldsymbol{E}$. How does this look like in reality and how to interpret the signals? The example of hole or electron 'generation'[44] via red laser and hole–electron 'generation' via IR-laser is illustrated in Fig. 1.47.

Figure 1.47a (hole 'generation') schematically shows the light injection on the n^+ face and the holes drift towards the p^+-electrode. The electric field configurations for bias voltages below and above depletion voltage are overlaid. On the right, the measured currents are displayed for different bias voltages. The initial steep rise reflects the light pulse. The current (charge carriers' velocity) increases until the holes reach the collecting electrode when the signal steeply falls off. The current slope reflects the electric field; higher voltages have steeper slopes and a shorter collection time. For low bias voltages holes are 'injected'[45] in a non-depleted region with no field present, thus take a long time moving via a diffusion mechanism into the field region – see long tail. As soon as the depletion voltage is reached the signal has a distinctive kink (here around 20 V) meaning V_{FD} is between 16 and 20 V.

Figures 1.47b (light injection at the p^+ face) depicts the electron 'generation'; with a three times higher mobility electrons are collected much faster than holes but they see the opposite field configuration thus slow down during their drift (opposite slope). The field driven drift (thus the induced signal) stops when the electrons reach the un-depleted zone for $V_{\text{bias}} < V_{FD}$ otherwise when they reach the n^+ electrode. The linearity of the field and the dependence on V_{bias} is clearly visible.

With a n-in-p diode the signal shapes would be reversed; e.g. electron 'generation' would exhibit a positive slope as in Fig. 1.47a, of course with faster rise time given the difference in mobility.

Figure 1.47c illustrates the 'generation' of holes and electrons with an IR laser in the whole volume; the signal corresponds to the movement (collection) of both charge carriers; the tails reflect the slower hole collection.

All in all, the full TCT system consists of a set of fast lasers with different wavelengths, an amplifier, some filtering circuits to disentangle bias voltage from the signal, a focussing system, a fast oscilloscope plus a computer to store and further analyse the signal. To measure irradiated sensors/diodes the whole system normally

[43] The induced signal ends so fast that is gets filtered by the bandwidth of the amplifier and/or scope (several GHz). In simulation, however, one sees the very fast and short amplitude.

[44] The jargon is to 'inject' electrons or to 'inject' holes.

[45] Again, this is jargon for generation of electron-holes with laser light, while one charge carrier sort is collected immediately at the near electrode, here electrons thus holes are 'injected'.

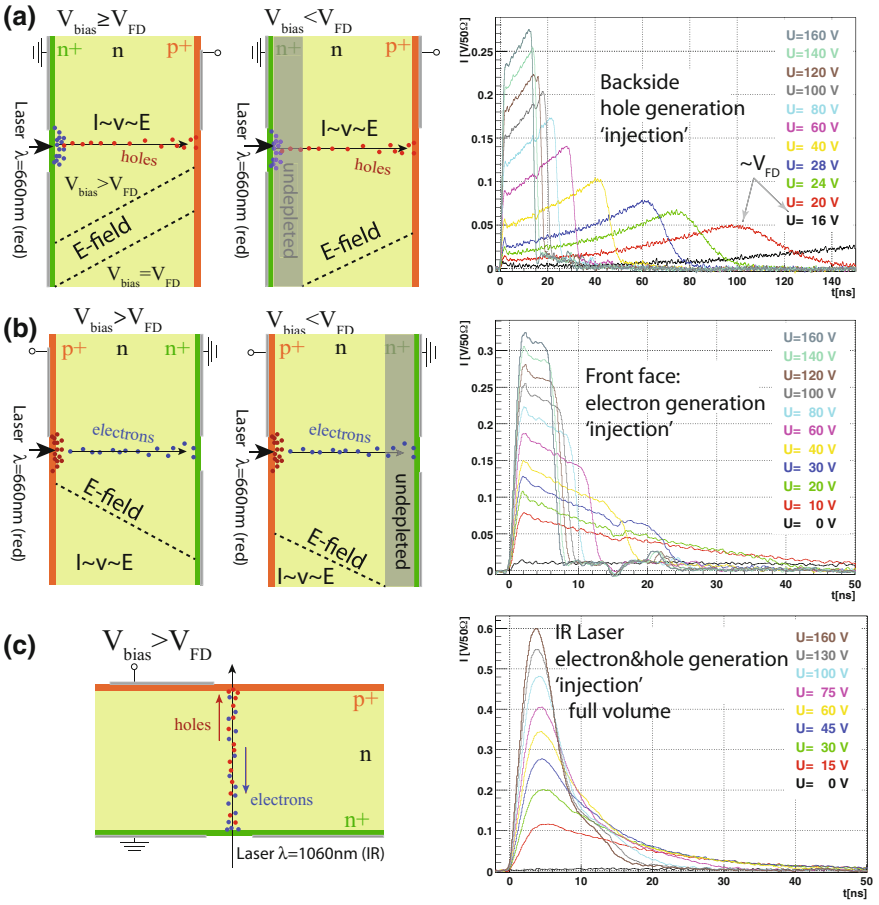


Fig. 1.47 The *left* part of the figure schematically illustrates the charge carrier 'generation' in a *p*-in-*n* sensor and the charges drift path without any information about the timely behaviour. The electric field configurations for voltages above and below depletion voltages V_{FD} are schematically overlaid. **(a)** Shows hole 'generation', **(b)** electron 'generation' and **(c)** both via IR laser. On the *right* the measured currents are depicted for the three different cases. The slope and height of the transient corresponds to the electric field resulting from the applied bias voltage. The signal drops sharply when the charges reach the electrode. The higher electron mobility is being reflected by the shortened time axis of (**b**). With $V_{\text{bias}} < V_{FD}$ holes are 'injected' into a non depleted zone without electric field thus do not drift thus do not feature the above described slope and drop; this is shown in (**a**) for voltage below 20 V. TCT plots from [171]

features a cold chuck and dry volume. Some systems can be set under vacuum to evaluate the *d*evice *u*nder *t*est DUT under cryogenic conditions. To resolve the signal and correctly extract the electric field the laser pulse has to be in the order of ∼100 ps and one needs a fast (GHz or better) oscilloscope; at high voltages the full electron signal stays below 6 ns for a 300 µm thick and even lower for thin sensors. In a real

system, special care has to be taken to eliminate any reflections in the cables and to fully understand the network of passive components and the diode itself plus the laser pulse shape – use a spice simulation and further process the signal accordingly, translating to the real current in the sensors/diodes.

With more complicated field configurations, a simulation of the whole system is compared in an iterative way to the measurement results, e.g. for strips or pixel systems where the weighting field is more complex or for highly irradiated samples. The main use case of TCT is to examine the behaviour of irradiated materials and specific sensor geometries. Several technical terms will be used here, which will only be fully introduced in Sect. 2.1 and more use cases will be given in Sect. 2.2.3. In the next paragraphs the "methods" to determine depletion voltage, to determine if the diode is type inverted and a way to evaluate effective trapping times τ_{eff} are introduced.

Full Depletion Voltage – QV-Method – CCE

The integrated signal, the collected charge $Q = \int I(t)dt$ versus the set voltage $Q(V)$ can be used to determine the depletion voltage. Comparing the collected charges Q for non-irradiated and irradiated DUTs at a given voltage, normally above V_{FD}, determines the **C**harge **C**ollection **E**fficiency CCE after radiation. For reasonably short integration times, only charge carriers created in the depleted zones of the device contribute to the integrated signal Q. This is due to the comparable long time ($\sim$50 ns) charges need to leave the un-depleted zone via the diffusion process and the additional electron-hole recombination during that time.

This can be exploited in two ways to measure the depletion voltage.

1. Inject red light at the un-depleted side (e.g. n^+ for p-in-n) for several bias voltage levels and record Q. A decent signal, with a kink, only starts for $V_{\text{bias}} \geq V_{FD}$. See also Fig. 1.47a, where the 'kinks' on the right indicate voltage values above depletion.
2. Using an IR laser depositing charge carriers throughout the volume, Q increases linearly with depletion depth reaching a plateau when the complete bulk is depleted (see also CV method in Fig. 1.10). TCT, though, does not fully replace the CV method but, by comparison, rather confirms the correct LCR frequency.

It is worthwhile to note that a similar method is used when the sensors are installed in in-accessible places. The signal is then scanned for different bias voltages and the signal plateau is determined (see also Sect. 6.6.2 later).

Space Charge Sign Inversion – SCSI – Double Junction

As will be explained in Chap. 2, radiation crystal defects can act as donors or acceptors. In an n-bulk acceptors can lead to the point where the sensor/diode bulk space charge changes *sign* (**S**pace **C**harge **S**ign **I**nversion SCSI or often briefly named "type inversion"), meaning, e.g. n-bulk changes to p-bulk. During this gradual change the depletion voltage changes and this can be measured with the CV or QV method but these measurements provide no information about the space charge sign. With bulk changing from n to p also the location of the main pn-junction changes from the

p- to the *n*-electrode thus also the electric field changes direction/*sign*. Measuring the N_{eff} and electric field profile with the TCT method gives a direct answer. This is illustrated in Fig. 1.48.

The feature of the double peak/double junction in Fig. 1.48 (lower right) will be further discussed in Sect. 2.2.3. With the onset of trapping, the second peak is often reduced/hidden and the signal needs further interpretation and corrections. A side remark, due to the double junction/double peak, experts have been discussing in the RD50 forum for some years if mCz material undergoes SCSI or not. TCT measurement were done from different faces (electron OR hole 'generation') and the slope of the first peak was interpreted as "the" slope while the second peak thus second slope was hidden due to charge trapping, where charges did not reach the electrodes.

Effective Trapping Time τ_{eff}

IR light, generating electrons+holes throughout the full volume is excellently suited for Charge Collection Efficiency measurements. The reduction of charge collection with radiation can be derived from the division of $CCE = \frac{Q_{irrad}}{Q_{non-irrad}}$ at voltages above depletion voltage. The difference of collected charge before and after irradiation $Q_{trapped} = Q_{non-irrad} - Q_{irrad}$ can be attributed to charge carrier trapping – with a certain trapping centre concentration, depending on the amount of radiation, and with a constant effective trapping time constant $\tau_{eff_{e,h}}$. Taking trapping mathematically into account, the full signal description of the TCT method is

$$I_{e,h}(t) = -e_0 \cdot N_{e,h}(0) \cdot e^{-\frac{t}{\tau_{eff_{e,h}}}} \cdot \boldsymbol{E}_W \cdot \boldsymbol{v}_{e,h}(t) \tag{1.61}$$

while the following part accounts for the signal reduction due to trapping

$$N_{e,h}(t) = N_{e,h}(0) e^{-\frac{t}{\tau_{eff_{e,h}}}} \tag{1.62}$$

Without trapping the measured integrated current signal $Q_m(V)$ for all voltages above depletion voltages *does not change*. *With trapping* the measured charge $Q_m(V)$ *still increases* with higher voltages thus higher field thus higher drift velocity (less time for trapping – less trapping) thus higher integrated signal[46]; basically more charges are being collected at the electrodes or better: more contribute with a higher drift velocity to the induced signal. Adding all up, the correlation between measured I_m and corrected I_c current is given by

$$I_c(t) = I_m(t) e^{\frac{t-t_0}{\tau_{eff_{e,h}}}} \tag{1.63}$$

[46]This is true for short current integration times as necessary for the LHC (bunch-crossings every 25 ns); with integration times longer than de-trapping times, the full charge would be measured.

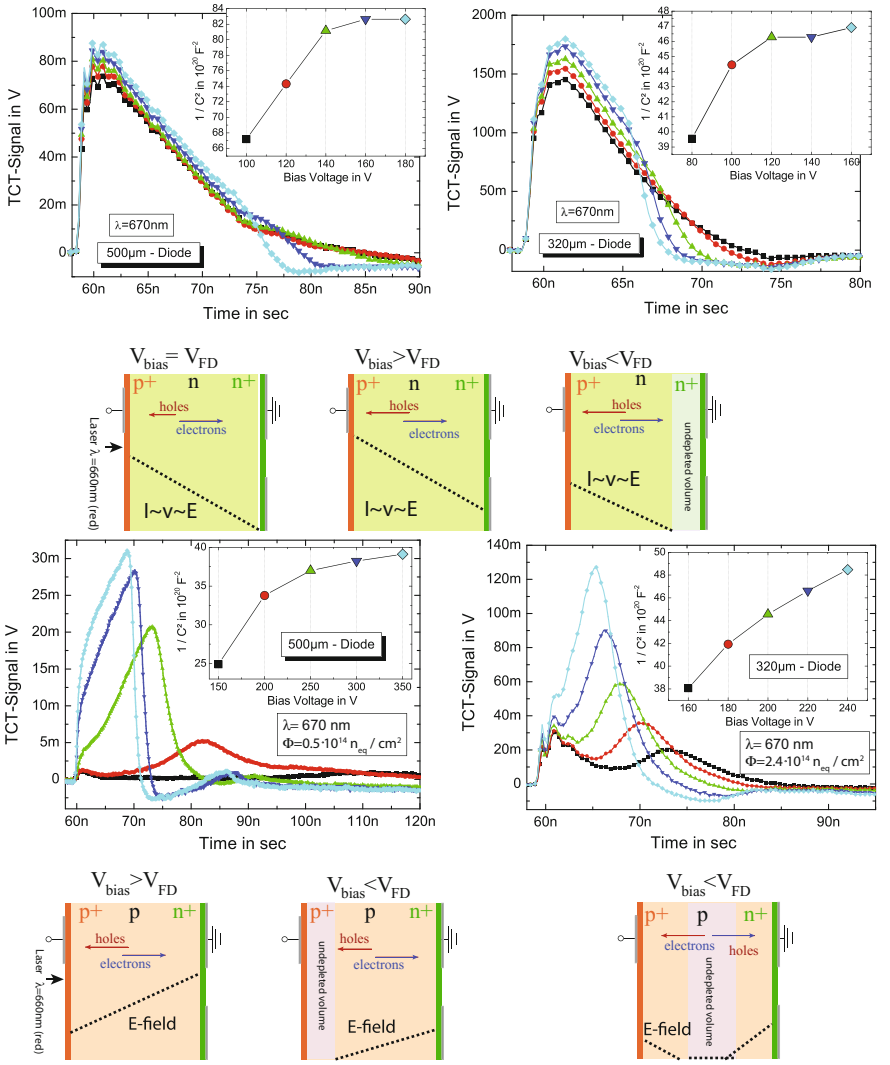


Fig. 1.48 The *upper part* shows the transient currents of *p*-in-*n* diodes of two different thicknesses before irradiation with electrons injected. The inlaid plots show the corresponding CV scans. The diode layout cartoons show the electric fields and depletion zones for the different voltages. In the *lower part*, the same measurements after radiation are shown. The change in slope direction of the *left* plot of the 500 µm thick diode clearly shows the SCSI. The 300 µm thick diode on the *right* shows an even more interesting feature namely a double junction with high field regions on both diode faces. Again the proportionalities from transient current to velocity to field to space charge give direct insight in the space charge. TCT measurement plots are from [80]

with t_0 as time of the laser pulse thus charge 'generation' and the fitting parameter τ_{tr}. **Selecting/Fitting**[47] $\tau_{tr} = \tau_{eff_{e,h}}$, the time integral of $I_c(t)$ gives the corrected collected charge $Q_c = \int I_c(t)dt$. Now Q_c will be constant for all voltages above depletion voltage. The difference between corrected and measured charge $Q_c - Q_m$ is equal to the charge lost due to trapping at a given voltage.

The described method to extract effective trapping time has been introduced in [171, 174]. It is also vital to correct for trapping to understand the real field and N_{eff} profile corresponding to the corrected $I_c(t)$ shape.

Varieties of the TCT-Systems

Several varieties of the TCT-system exists:
TCT, multi-channel-TCT, Scanning-TCT, edge-TCT and Two-Photon absorption TCT; the last two systems with additional parameter extraction capabilities, will be described with further detail in the next paragraphs. A multi-channel TCT is used for segmented sensors (pad or strips), where the different segments/channels are each connected to a fast oscilloscope (several channels simultaneously) this is often combined with a movable laser (point of light injection) and then called scanning TCT. The point of charge 'generation' can, for example, be moved along the pitch of a strip sensor where the signal formation properties depend on geometry thus electrical $\boldsymbol{E}$ and weighting field E_W of a multi-electrode system. It is thus possible to study electric field properties and charge sharing between strips in the presence of intermediate strips, routing lines, pads, or trapping.

Many CCE and depletion voltage results of Sect. 7.1.1 have been obtained with a TCT setup; especially for highly irradiated samples. It is worth to note that for new materials (mCz, Float zone with different oxygen content, Epitaxial silicon, …) to understand the radiation tolerance wrt. to bulk damage, the measurement on diodes with a simple geometry is fully sufficient!

In a nutshell, TCT

- is mainly used to understand changes after irradiation of different materials and sensor geometries
- is based on fast charge 'generation' ($\leq$100 ps) by shining focussed laser light and measurement of the effective transient current with timing/GHz resolution. The measured transient current depends on the electrons/holes velocities in turn dependent on the electric field
  - red laser on *p*/*n* electrode face injects electrons/holes, respectively
  - IR light injects both electrons and holes in the whole volume
- the **electric field profile** and thus the **effective space charge profile** can be derived – the slope of the TCT signal is directly linked to effective space charge concentration
- depletion voltage, space charge sign, charge collection and effective de-trapping times can also be determined
- the results need to be corrected for trapping for full exploitation.

[47] If $1/\tau_{tr} > 1/\tau_{eff_{e,h}}$, Q_C would be too high or vice versa $1/\tau_{tr} < 1/\tau_{eff_{e,h}}$ too low.

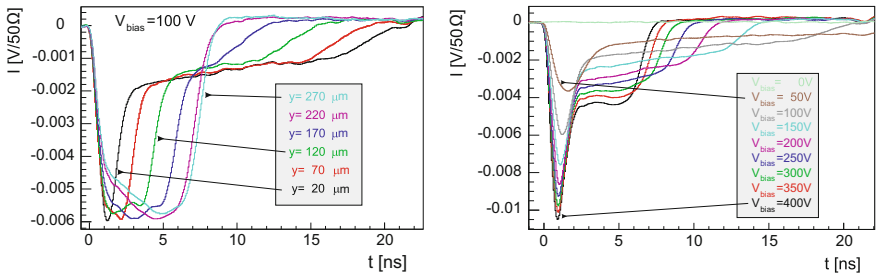


Fig. 1.49 Edge TCT, a novel tool to achieve a deep understanding of charge propagation was developed in the Jožef Stefan Institute, Ljubljana, Slovenia. Infrared laser light shines from the side allowing dedicated charge deposition per unit depth. The *left* figure shows the different TCT signals per depth in a non-irradiated n-in-p FZ sensor ($V_{\text{bias}} = 100\,\text{V}$); $y = 270\ \mu\text{m}$ is situated near the backplane and $y = 20\,\mu\text{m}$ near the strip region. The initial peak represents the collected electrons and the long tail comes from the drift of holes. The shortest signal can be seen for $y = 220\,\mu\text{m}$, where electron and holes have an equal drift time. In the *right* figure, a bias scan at $y = 20\,\mu\text{m}$ has been done. The second peak in the induced current is getting shorter with voltage as well as the electron peak is getting higher. TCT plots from [178]

Edge-TCT

A new tool developed in the Jožef Stefan Institute in Ljubljana sheds new light on dynamics in strip sensors at dedicated locations inside the device's bulk – Edge TCT [178]. Combining the TCT method with the "grazing" signal method, infrared laser light shines into the sensor from the side (edge). After preparation (cutting, polishing) of the side of a sensor, light can be injected perpendicular to the strips therefore shining into defined regions with respect to volume depth and illuminating homogeneously several strips. In this configuration the normal weighting field for strips/pixel (rf. Sect. 1.3.2) does NOT need to be taken into account; strips and neighbours experience the same charge without disturbing the real field configuration. The effective factor corresponds for the weighting field of a diode $1/D$ simplifying the analysis. In reality one has to take into account the growing beam spot with distance in the volume due to Gaussian propagation. Figure 1.49 shows a charge deposition versus depth scan. The method is very effective and offers more information than front or back face light injection only. The detailed description of this impressive tool and the possible analyses is beyond the scope of this document.

The system allows very detailed studies for example the determination of the "velocity profile" where the IR laser scans over the sensor thickness (example in Fig. 2.21 on p. 159), "trapping time","electric field" and "charge collection profile"; these are described in [178]. An example how to investigate 3D sensors with edge-TCT is described in [278].

Two Photon Absorption – TCT; TPA-TCT

A new **T**wo-**P**hoton-**A**bsorption-**TCT** TPA-TCT concept allows 'generation' of electron–holes in a "spotlike" fashion – in reality an ellipsoid, couple of micrometers wide times a couple of tens of micrometers long, somewhere targeted inside

the sensor/diode volume! Geometry of the electric fields, localized detector response can be evaluated in a full 3-dimensional and real-time fashion. This allows direct investigation of interesting regions, e.g. near a strip or pixel cell or a dedicated deep well in HV-CMOS structures.

With photon energies (wavelengths) below the bandgap energy ($E_{\text{photon}}, E_{gap} \sim 1.1\,\text{eV}$, two photons are necessary simultaneously (within ∼100 attoseconds) to create an electron–hole pair. The first photon creates a virtual state for about $\tau_{\text{virtual}} \sim \frac{\hbar}{E_{gap}/2} \sim 0.1 \cdot 10^{-15}$ s; adding the second photon totals to enough energy to cross the band gap creating an electron–hole pair. The trick is to focus a single laser with λ= 1300 nm or 1500 nm[48] at the spot of interest or better move the focused spot by having the device under test DUT or laser optics on a moving stage. A femtosecond pulse laser[49] in mode-locked mode enhances the probability of TPA ensuring the timely coincidental presence of two photons for the same intensity. Special optics are necessary to work for these high wavelengths. The injection can be done from top, bottom or edge, thereby directing the ellipsoid (longer in beam direction); as for the normal TCT metal obstructs the laser beam.

Another interesting testing mode is to keep the focus outside the sensor volume and scan irradiated samples. For non-irradiated samples, with the focus outside, **S**ingle **P**hoton **A**bsorption SPA is not possible due to the low absorption at this wavelength. For irradiated DUTs trap assisted SPA can be observed; another way to check relative trap concentration. In irradiated samples the trap assisted SPA also gives an offset along the full laser line to the TPA signal – which can be corrected for by measuring first with the focus outside (subtract I(t) from trap assisted SPA). The trap-assisted SPA being linear with intensity is more suppressed with respect to TPA, which is non-linear with intensity for $\lambda = 1500\,\text{nm}$ compared to $\lambda = 1300\,\text{nm}$. For two use-cases in HEP, the reader is referred to [113, 114].

1.8.4 Secondary Ion Mass Spectrometry – SIMS

Secondary **I**on **M**ass **S**pectrometry – SIMS is a well know method to analyse the composition of 'surfaces' with a depth profile of several micrometers. It is commercially available and is used commonly in any material science and material quality assurance lab. The surface of the specimen is subjected to a focused ion beam sputtering small pieces from it. The pieces are then analysed by a mass spectrometer thus determining their elemental, isotopic or molecular composition. Depending on the machine sensitivities of parts per million up to parts per billion can be achieved.

For silicon sensors, SIMS is being exploited to understand impurity levels but also the general composition of the different areas, like bulk, p- or n- implants, SiO_2-passivation or thermal grown oxide and metal. Dedicated "SIMS" fields are

[48] Any wavelength above 1150 nm would do the trick; the absorption maximum is at about λ= 1300 nm. λ= 1500 nm lasers are commercially available from telecommunication industry.

[49] Typical pulses are 100 fs wide.

placed on the half-moons of the wafers with all the above items in a representative way (see also Sect. 1.7.1). SIMS is often used during initial R&D or early material/sensor evaluation of first production pieces. The method allows depth profiling from several Angstroms to tens of micrometer. Levelling or even cutting a sensor and polishing the edges can even give insights throughout the full thickness of the individual layers and their molecular composition.

1.8.5 Simulation

Software simulation packages are becoming more and more powerful. The tools are also often adapted and improved to reflect specific needs. The main ones are Synopsys TCAD [282], Silvaco TCAD [273] and Cogenda [66] plus many non-commercial tools.

The four main usages are (a) device simulation, (b) dynamic system simulation, (c) simulation of radiation effects and (d) simulation of the complex system to improve tracking in the final experiment.

Simulations are necessary to understand and optimize the breakdown voltages, leakage currents and the capacitance network; minimize load and parasitic capacitances. New sensor designs are simulated by the designers and often again by the producers. Strip width to pitch ratios, metal overhangs, p^+-stop implant widths and distance between implants are some examples, others are the optimization of doping concentrations of p^+-stops, strips and backplane implants or p^+-spray doses, implant depths, etc. Basically all parameters discussed in Sect. 1.6 and also routing schemes as described in Sect. 1.4. From this field configuration, mobilities, drift velocities etc. can be derived. For example, the simulation presented in Fig. 1.19 on p. 26 illustrates quite nicely field, mobilities and the dynamic charge carrier movement in a strip sensor.

Another important aspect is the simulation of complicated dynamic systems as the Transient Current Technique TCT (see Sect. 1.8.3) or the grazing angle method in conjunction with radiation (see Sect. 2.2.3). In these cases the full system is simulated, compared with the measurement, parameters are adapted in an iterative way and in the case of measurement to simulation match, the right parameters have been derived.

During the design stage, the full system – sensor plus front-end readout – is simulated to tune the electronics design (amplifier, shaper, etc.).

The simulation of radiation effects is becoming more and more predictive up to radiation levels expected for the experiments at the future **H**igh-**L**uminosity **L**arge **H**adron **C**ollider HL-LHC. One basically exploits the understanding which defect energy levels and defect concentrations are being introduced in the different materials with the different particle radiation fluences – energies and particles see Fig. 2.11 on p. 151. Technically the multitude of defects are approximated by 2, 3 or 4 effective levels representing donor and acceptor levels plus trapping and current generators. An example of such models and how they are deployed in simulation can be found in [37]. This allows a decent projection of evolution of leakage currents, space charge

hence electric field configuration and trapping. The main bottlenecks are still the input parameters derived from microscopic measurements; e.g. introduction rates of defects, cross-sections, etc. Also evolution of strip/pixel parameters with surface damage is being simulated, mainly the change of inter-strip capacitance and inter-strip resistance and the influence with respect to breakdown voltages.

Unfortunately, no general parametrization, as the HH-model (rf. Chap. 2) for *n*-type Float-Zone sensors, exists yet for CCE, trapping and depletion voltage for the next generation of sensors. Full simulations with trap models are useable tools.

Another example is the PIXELAV [280, 281] simulation, used in the full CMS pixel detector, producing so-called pixel templates "knowing" a priori the expected pixel cluster shape depending on the spatial placement with respect to the primary interaction zone of the pixel module; e.g. "knowing" the principle incident angle of the particle but also taking into account the Lorentz angle, the operation voltage (thus field configuration) and the possible partial depletion configuration, etc. all following the evolution with radiation. This improves the spacial resolution per se but also allows prediction of future behaviour with radiation. The tool is also being used to optimize future pixel designs (pixel cell geometry, sensor thickness, etc.).

1.9 Production of Silicon Sensors

The basic material for the manufacturers of silicon sensors is sand SiO_2 quartzite, existing in abundance. It is cheap, the price of a sensor is completely driven by the number of processing steps and the required quality. The success of the different companies is due to a deep understanding of the different processes and required recipes. The basic principles are known to everybody but the tiny details are always "company secrets". The subject of silicon processing could easily fill a unique book. The next sections focus on the following steps:

- silicon crystal ("ingot") production
- slicing into wafers, lapping, etching and polishing wafer
- processing
  - oxidation
  - photolithography in-between almost all the steps
  - doping by ion implantation
  - annealing
  - metallization
- dicing (cutting) the sensor out of a wafer
- testing

This sequence grandly simplifies the real process with nearly 60 – 100 steps, e.g. some additional annealing for re-diffusion of previously introduced dopants or removal of defects. In addition, in between each major step there are extensive cleaning cycles, lots of acidic chemicals as agents are used as well as very pure de-ionised water.

1.9.1 From Pure Sand to Detector Grade Silicon

The starting material is simple mined quartz sand (SiO_2), undergoing extensive purification and distillation. The first step is to produce *metallurgical grade silicon*, 95 – 99% pure silicon (MG-Si) in a series of melting processes at temperatures from 1500 to 2000°C, while applying carbon in the form of coke or coal. The oxygen then combines to carbon monoxide, leaving chips of MG-Si ($Si\,O_2 + 3C \rightarrow Si\,C + 2C\,O$ then $2Si\,C + Si\,O_2 \rightarrow 3Si + 2C\,O$ at temperatures around 1500 – 2000°C).

In the next purifying step MG-Si is subjected to HCl in a fluidized-bed reactor to form trichlorsilane ($Si + 3H\,Cl \rightarrow Si\,H\,Cl_3 + H_2$ at temperatures around 300°C); in this course many impurities are removed. The next purification step to accomplish an impurity concentration of less than 1 ppba[50] is achieved by distillation.

The hyper-pure $Si\,H\,Cl_3$ conversion to hyper-pure silicon is finally accomplished by **C**hemical **V**apour **D**eposition (CVD), where $Si\,H\,Cl_3$ is vapourized with highly pure hydrogen additive in a deposition reactor ($Si\,H\,Cl_3 + H_2 \rightarrow Si + 3H\,Cl$ at temperatures around 900 – 1100°C). The final silicon then slowly grows on ∼1000°C hot slimrods in a few days to **E**lectronic **G**rade **S**ilicon (EGS) and has achieved the commonly known "eleven nines" (99.999999999%) < 1/100 ppb) purity. The slim-rods are then broken to pieces, dopants are added and the whole melange heated up to a melt in a quartz crucible. The real processing starts by growing a single crystal ingot from the polysilicon including the needed dopant additive. The most common method of wafer development is the **Cz**ochralski ("Cz") growth: The silicon melt is held just a few degrees above the melting point and a single rotating crystal as seed initiates the growth at the top. The slow retraction allows the melt to solidify in perfect crystal orientation at the boundary. The growth is controlled by the pull rate, the melt temperature and the rotation. Intrinsically for this method, oxygen dissolves into the melt during the process, enriching the final crystal homogeneously with oxygen.

To get *n*- or *p*-type base materials, dopant atoms such as boron or phosphorus can be added to the molten silicon in precise amounts to dope the silicon.

It has been shown, in a further step, that the homogeneity can largely be improved by conducting the whole process into a magnetic field; the resulting material is then called **m**agnetic **Cz**ochralski or simply "mCz" – more of its benefit in Sect. 2.2 and Chap. 7.

The second method is the **F**loat **Z**one Crystal technique ("FZ"), where a polysilicon rod is brought into contact with a seed crystal and the rod is then locally melted with RF heating. The RF heater and therefore the melted zone moves along the rod. The impurities have a different diffusion constant and tend to stay in the liquid, leaving a pure single crystal. The ingot growing apparatuses of the FZ and CZ techniques are displayed in Fig. 1.50. In addition the silicon rod has *no* contact to any melt or crucible. Oxygen concentrations are inherently very low.

[50] 1-ppba: part per billion active (impurities).

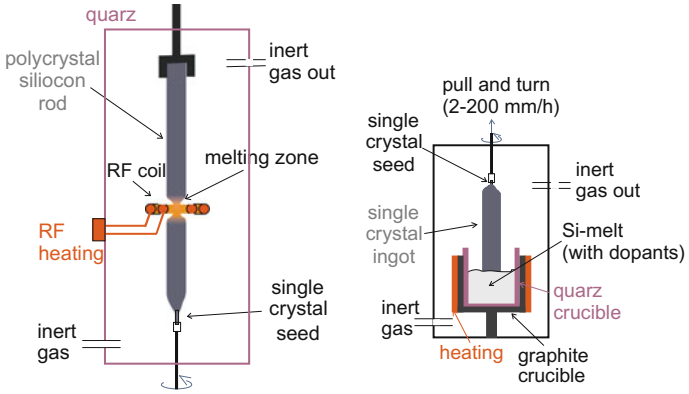


Fig. 1.50 The *left* picture shows the *F*loat-*Z*one method FZ, where a polycrystalline silicon cylinder from a mold with a single crystal seed on one edge is subjected to RF heating, melting the cylinder to form a single crystal or "*ingot*". The impurities have a higher solubility, therefore impurities diffuse to the boundaries further purifying the crystal. The *right* picture shows the **Cz**ochralski method CZ, where the single crystal "*ingot*" is directly drawn out of the melt using a single crystal as seed. The Czochralski method is most common globally, it additionally allows easy adding of dopants directly into the melt

FZ ingots are doped by gas diffusion to achieve n- or p-type base materials. n-type can also be achieved by **N**eutron **T**ransmutation **D**oping NTD, where with $^{30}\text{Si+n} \rightarrow {}^{31}\text{Si} \rightarrow {}^{31}\text{P}+\beta^-$ one achieves better uniformity.

The typical length of a single crystal ingot is 1 – 2 m. The ingots are normally not fully round and are then grinded to real cylinders, also a flat or a notch is applied to indicate the crystal orientation. Very early[51] sensors have been produced on 2 in. diameter wafers while today 4 – 6 in. wafers are common and 8 in. wafers are candidates for some HL-LHC detectors – many examples in the next chapters. Today's electronic ASIC industry works mainly with 8 – 12 in.

For several years the FZ technique was superior to the CZ method with respect to purity (e.g. oxygen content) and was therefore the only possibility to achieve "sensor grade" silicon. Also high resistivity material was not available due to the lack of commercial applications. With a better purity and therefore a longer charge carrier lifetime it was also the preferred material for solar cells. Nowadays, the CZ also realizes "sensor grade" material and probably the high oxygen content makes it interesting for the HL-LHC project, see Sect. 7.1.1. On the other hand, since higher oxygen concentration seems to improve radiation tolerance, oxygen is often added to FZ in a later step by diffusion – DOFZ **D**iffusion **O**xygenated **FZ**.

The silicon ingot now needs to be cut into thin disks, called "wafers". The actual thickness is still not final but already below 1 mm. The cutting is achieved by a disk with a diamond blade or for multiply synchronous dicing a **M**ulti **W**ire **S**aw MWS,

[51] 1955, early silicon ingots if 3/4 to 1 in. diameter were fabricated by Montecatini.

Fig. 1.51 A silicon ingot, source of about 200 wafers. The picture shows raw silicon chips, a full size silicon ingot and sliced silicon wafers of different size. [Copyright Siltronic AG]

where diamond-coated wires are arranged in parallel to cut several disks at once. The sliced wafers have to be cleaned before processing, which is achieved by

- lapping: grinding away large imperfections, also wafer edges are shaped
- etching "chemical polishing": full removal of imperfections, micro-protrusions and surface impurities, wafers in special chemicals are agitated, rinsed, heated, dried, like porcelain dishes in a home dishwasher
- polishing: final polish[52] is needed to ensure a minimum number of defects and guarantee the specified thickness of the silicon wafers

An ingot and cut wafers are shown in Fig. 1.51.

1.9.2 Processing

The processing steps are not different from standard integrated circuit chip processing (memory chips, processors). The photolithography step, explained just once, happens basically in between all further steps. As an initial step, a high resistivity polished silicon wafer (1 – 10 kΩcm) is covered with SiO_2.

Silicon oxide growth is a key process step in the manufacture of all silicon devices. Outer layers of silicon wafers in a heated quartz tube, a furnace (see Figs. 1.52 and 1.53), are transformed to SiO_2 by an atmosphere of O_2, H_2O steam or N_2O. In the

[52] To process a double-sided sensor, the wafer needs to be polished on both sides. This step, as easy as it sounds, is not available on both sides from all manufacturers.

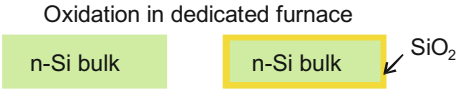


Fig. 1.52 Silicon oxidation: to enable segmentation and structuring, the raw silicon wafer is first subjected to an oxygen or water atmosphere at 800 – 1200°C in a furnace (see Fig. 1.53). The wafer is afterwards fully embedded in a SiO_2 layer

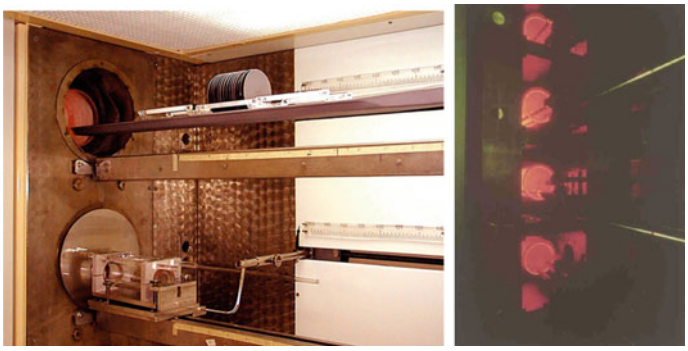


Fig. 1.53 Two quartz tubes in the furnace are visible on the *left* photo including the loader of several wafers. Each individual furnace has a well-defined atmosphere and always remains at the same temperature and content mixture. There is one furnace for each individual process. On the *right*, several burning furnaces are shown. Courtesy of ITE Warsaw [347]

"dry" process, pure dry oxygen is employed, the oxide grows very slowly realizing very uniform layers with high quality and few defects at the oxide – silicon interface. This process is mostly used for dielectrics for coupling capacitances (∼100 – 200 nm) of AC-coupled sensors and for MOS (∼100 Å) transistors.

In the "wet" process, H_2O steam is injected, the oxide grows fast. The additional hydrogen atoms may degrade the oxide quality, the process is mostly used for thick field oxide or maybe passivation oxide (∼μm). The processes are

$$Si + O_2 \rightarrow SiO_2 \quad \text{"dry"}$$
$$Si + 2H_2O \rightarrow SiO_2 + 2H_2 \quad \text{"wet"}$$

For SiO_2 used as insulator, e.g. in-between metal layers, SiO_2 obviously cannot be grown, in these cases sputtering is used – deposited oxide.

Figure 1.54 describes the photolithography and etching process. First, for each layer, a mask[53] is needed, which is normally a chromium pattern on a glass plate. Then, the wafer is covered with UV light-sensitive photoresist, which is spilled on the fast rotating wafer in the spinner (see Fig. 1.55). UV light exposure of the resist through the mask transfers the patterns to the wafer. A final development step, like in every photo studio, is also necessary. Figure 1.56 shows a picture of the mask-aligner

[53] The mask itself must be very precise and is very expensive.

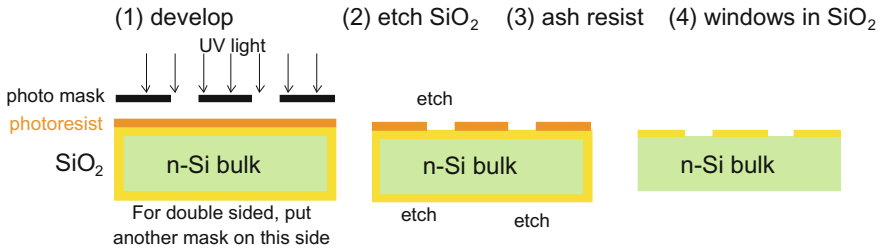


Fig. 1.54 First segmentation step: The oxidized wafer is now covered with photoresist, parts of the resist is then masked and subjected to UV light; after final development, the sensor is etched. As a final step the resist is etched away leaving defined windows in the SiO_2 layer

Fig. 1.55 Photoresist is roughly spilled in the middle of the wafer, centrifugal forces, during fast rotation, homogeneously distributes the resist on the wafer. Courtesy of ITE Warsaw [347]

Fig. 1.56 On the *left* the mask-aligner is printing the mask onto the photoresist with UV light, opening the windows for the next etch process step. On the *right*, the spinner and developing machine can be seen. Courtesy of ITE Warsaw [347]

and the spinner, this part of the clean room area needs to be especially clean. Any dust particle during lithography will end up as a sensor defect later. Additionally, in some cases, the photoresist undergoes a heating cycle for hardening purpose to make it more resilient for different processing steps.

The pattern allows a controlled removal/etching of SiO_2 on the sensor to enable a controlled implantation later, the SiO_2 serves as a screener.
There are two basic etching methods:

- **wet etching** uses liquid etchants, where wafers are immersed in the solution. Wet etches are selective, isotropic and fast. Time, concentration and applied temperature of the immersion are important. Also crystal orientation has to be taken into account, e.g. $\langle 100 \rangle\ Si$ can be etched 100 times faster than $\langle 111 \rangle\ Si$. Wet etching is also used for SiO_2, silicon nitride, polysilicon, photoresist, etc.
- **dry or plasma etching** uses gas phase etchants in a plasma. Directional etching is possible by applying electric fields to direct ionic species in the plasma, there are reactive and ionic components. The etching is also very selective. Plasma etching stresses the object and needs an annealing step at higher temperature afterwards, it is therefore not applicable after metallization.

In Fig. 1.57 both plasma and wet etching laboratories are shown. The full etching recipes are always adapted to the dedicated device and are an important know-how of the manufacturer. Basically *Si* is etched in a two-step process

$$\text{Oxidation } Si + 2HNO_3 \rightarrow SiO_2 + 2HNO_2$$
$$\text{Reduction } SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$$

while etching of SiO_2 is a one step process

$$\text{Etching } SiO_2 + 6HF \rightarrow H_2SiF_6 + 2H_2O$$

Fig. 1.57 On the *left*, the plasma etching laboratory of a silicon plant in an institute is shown. On the *right*, the wet etching facility is displayed. The secret and art lies in the special recipes to use the right acid, the right duration and time on the different structures and materials, like Si, SiO_2 and Al. Courtesy of ITE Warsaw [347]

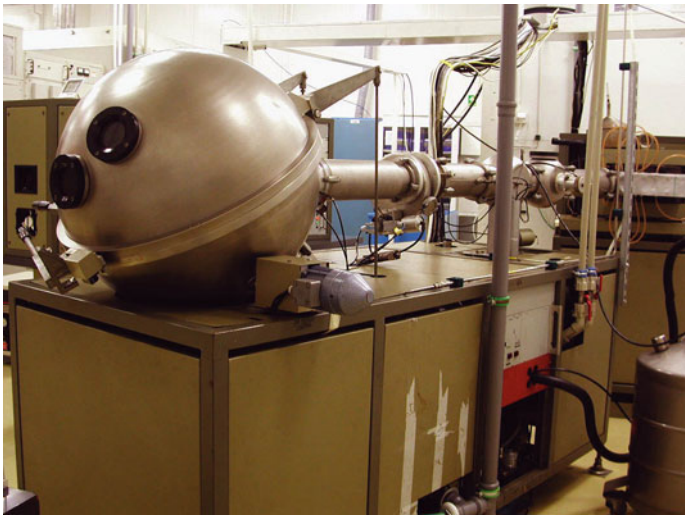


Fig. 1.58 The picture shows an ion implanter. Ions are accelerated and guided onto the target inside the sphere, where several wafers are rotated and homogeneously hit. The ions do not penetrate through the masking layer (SiO_2 or photoresist), therefore well-defined patterns, e.g. strips/pixel can be implemented on the wafer. Courtesy of ITE Warsaw [208, 347]

Aluminium etches in water, phosphoric, nitric and acetic acid mixtures.

And finally the photoresist has to resist all the above etchants to serve as a mask. After the etching, the remaining resist will be etched away, last portions of the resist will be dissolved by oxygen or fluorine plasma, mostly this goes along with another wet cleaning step. In the next step the strip implants or homogeneous backplanes are realized by ion implanting or diffusion. An older version of an ion implanter can be seen in Fig. 1.58. A couple of patterned wafers are placed in a rotating disk to guarantee a homogeneous irradiation. Ions are then accelerated towards the wafer and are finally implanted, e.g. a 15 keV B^+ beam results in a shallow doping or a 30 keV As^+ beam for a deeper backplane doping. Typical doping concentrations, depending on the beam current/intensity, would be $N_{\text{acceptors}} \approx 5 \cdot 10^{18}/\text{cm}^2$ strips or $N_{\text{donors}} \approx 5 \cdot 10^{19}/\text{cm}^2$ for the backplane. Doping penetration for energies below 300 keV stays well below 1 µm. The SiO_2 or photoresist acts as a mask. The standard example, also cartooned in Fig. 1.59, is the implantation of small p^+ lines into the n silicon bulk, creating many local pn-junctions – the strip and ring pattern. The ion implantation results in a very shallow doping. After the implantation, a high-temperature step is necessary to stimulate dopant diffusion, driving dopants further into the volume and smoothing edges. It also anneals structural damages. This high-temperature process also 'activates' the dopants.

Full plane doping, like the backside, is often realized via diffusion. Diffusion dopant sources can be gases, doped amorphous silicon or dopant-enriched glasses, like boron or phosphoric glasses, which are deposited on the surface and etched away

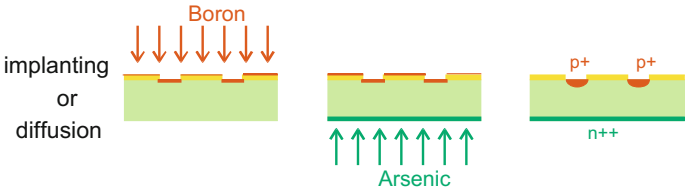


Fig. 1.59 Ion implantation or diffusion is utilized to realize the pn-junction with B on the front side and avoid a Schottky contact on the backside with arsenic As or phosphorus P (here higher doping concentration e.g. n^{++} for a p-in-n sensor). SiO_2 is used as a mask

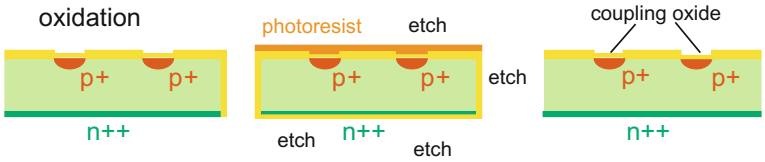


Fig. 1.60 To capacitively isolate the p^+ from the aluminium readout strip, a layer of high-quality SiO_2 is introduced by thermal growth. As for all thermal oxidation the oxidation is applied to the full wafer. A photoresist step on the front and an etching step is needed to allow backside oxide removal. In most real structures, patterns of silicon oxide and silicon nitride are applied for the strip pattern

after the diffusion process. During one side processing the other side is protected by a SiO_2 layer.

For an AC-coupled sensor, several steps are needed to build high-quality thermal grown oxide, where the basic steps are shown in Fig. 1.60. The full wafer again undergoes one to several high-temperature steps in the oven with steam, oxygen or nitrogen atmosphere to achieve a sophisticated, thin and voltage-stable oxide; sometimes also layers of polyimide are applied. The result is a thin isolating oxide above the strips and thicker oxide in the strip intermediate areas. Quite often additional oxide deposition is applied beneath the final pads for micro-bonding to increase thickness and strengthen the structures. To remove the oxide from the back, photoresist is spilled on the front without mask, and the back oxide is etched away.

For the basic strip concept all steps but the final metlization and passivation are described so far. Besides the strip and ring structure processing, several additional steps are needed to apply the bias to the p^+ strips. The most complicated and most common case, the polysilicon resistor biasing, is introduced briefly in Fig. 1.61.

Finally the back side will be fully metallized for simple electric contact, while the front side must be patterned and needs additional steps (Figs. 1.62 and 1.63). For metallization, most often aluminium Al with some percentage of Si is used, due to its good electric conductivity ($\approx 3.7 \cdot 10^7$ S/m), its good connectivity to silicon and finally its easy shaping by photolithography. Aluminium coverage is achieved by **C**hemical **V**apour **D**eposition CVD or a sputtering process. For the front side another lithography step deposits photoresist, in a later development step windows

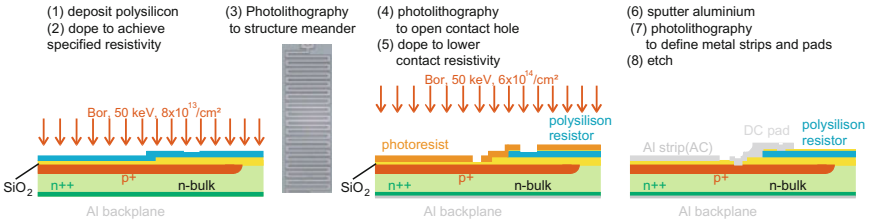


Fig. 1.61 The request of a polysilicon biasing comes along with several additional process steps. First, the sensor is fully oxidized and then covered with a CVD-deposited layer of polysilicon. Resistivity is defined by a controlled doping by ion implanting, with, e.g. boron. As done initially, the wafer is oxidized, structured by a lithography process where only the final bias meander is covered with photoresist, and most of the poly is etched away. The absolute resistance is now defined by initial resistivity, length and width of the polysilicon resistor. Another lithography step is needed to open small windows for the metal contact to lower the contact resistivity of the polysilicon locally. Final opening of the contacts into the oxide to the p^+-strips, the poly meander and the bias ring allows final metallization. The cartoon is presented as a side view of the sensor

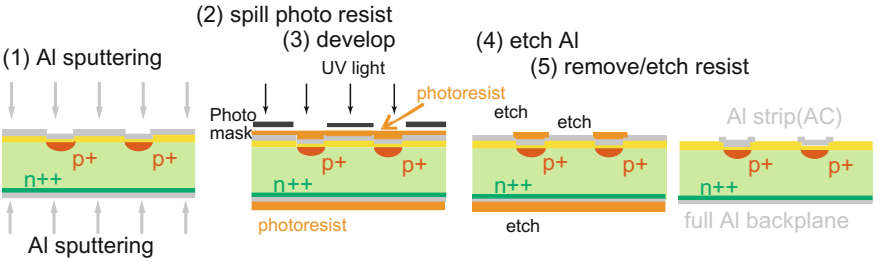


Fig. 1.62 As for the oxide, the full surface is metallized by a CVD sputtering process. Another lithography and etching step later, aluminium strips, pads and rings are formed. For a good electric contact, the back side is fully covered with aluminium

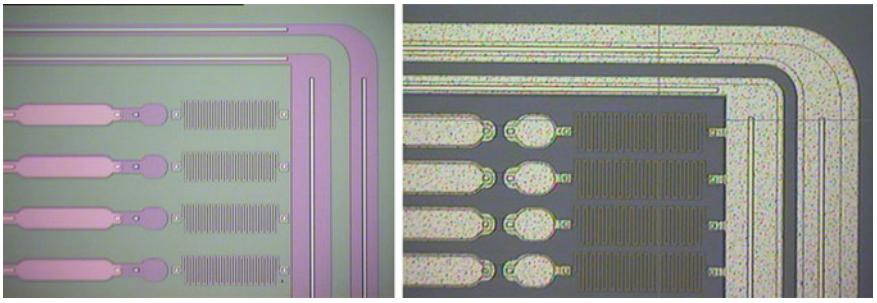


Fig. 1.63 The *left* shows all the p^+ structures and polysilicon resistors before the final aluminium is applied in comparison to the final step on the *right*. Bias- and guard ring implants are connected using electric "vias" to the metal layer above; here in the configuration of a *continuous* via contact, while also discrete "vias" as used for DC pads are common

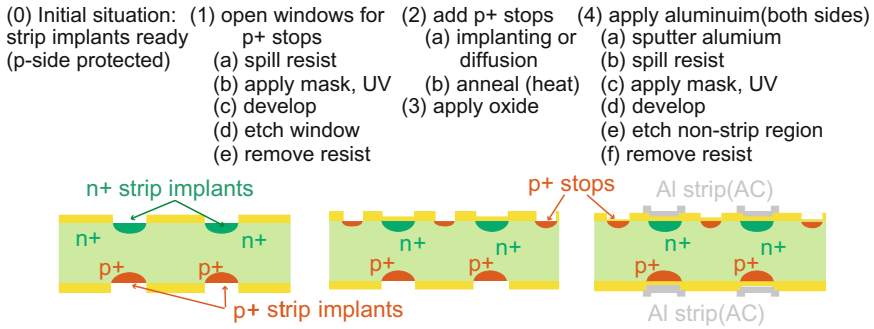


Fig. 1.64 The cartoon illustrates the many additional steps needed to structure the ohmic side. n^+ strip segmentation on the ohmic is implemented analogue to the p+ strips technique on the junction side described earlier. Additional p^+-stops necessary for n^+ strip isolation also need lithography, implantation and annealing steps analogue to the strip processing. Coupling oxide is applied via thermal growth (this step is not illustrated). Finally on top of the n^+ strips aluminium (with some Si content) strips are formed (for use case rf. Sect. 1.4)

are opened and aluminium is etched away. Additionally, there are a lot of testing steps in-between all these processes to guarantee quality and to fully identify problems early in the process chain.

At this stage, a single-sided DC-coupled sensor is completely processed waiting for its final cleaning, passivating and cutting step. For the second face, the ohmic side processing, only the combined layer processes are briefly described; the detailed steps, like lithography and etchings, etc., are the same as for the junction side and are not elaborated in detail. The different steps are illustrated in Fig. 1.64. One of the complications is the necessary protection of the other side. Double-sided sensor processing is only possible on substrates that are polished on both sides. A strip-like segmented n^+ layer replaces the formerly described full area n^+ implantation. To establish strip isolation additional p^+-stops are needed. And finally, aluminium lines are needed on top of the n^+ strip above the coupling oxide, as well as all pads, bias and guard rings.

In a lot of cases, the processing does not end here. Another isolation layer (of SiO_2 or polyimide) is needed, pierced by vias, connecting to aluminium routing strips above the isolation.

In principle none of the above-described steps are problematic if isolated. In the full combination with the necessary purities, one has to fear pinholes, failed vias, increased capacitance, cross-talk between channels, tiny contamination in some oxide layers, failures affecting neighbour or other side strips, etc. Keep in mind that different from standard chips, one single sensor covers the **full** wafer, therefore one single defect may be enough to reject a processed whole wafer. Finally the full wafer is passivated most often with a rough layer of sputtered SiO_2 but also commonly with polyimide. The passivation is for protection, otherwise silicon would react with all contaminants in the environment. A last etching step is required to make the required

Fig. 1.65 The cutting of chips with a diamond saw. Note the high water flux to avoid heating up due to the high saw rotation speed. Courtesy of ITE Warsaw [347]

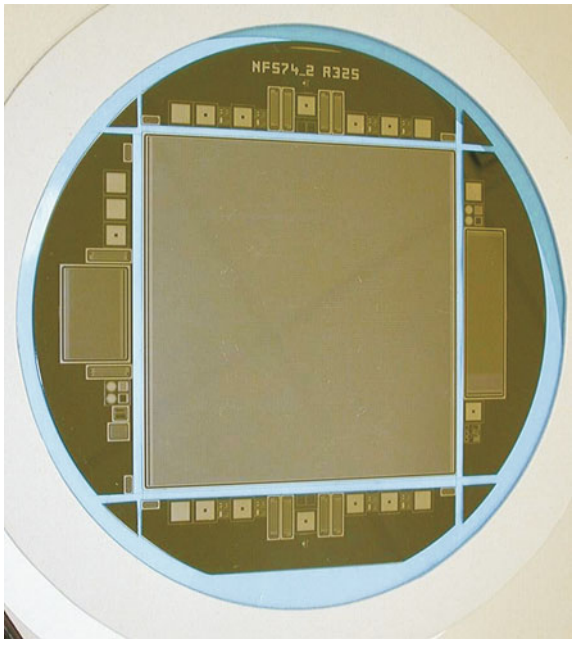


Fig. 1.66 A CMS sensor cut from its wafer. The blue foil, the wafer rests on, was not cut

metal pads accessible for micro-bonding (see Fig. 1.11), of course another mask and etching step is needed.

Sensors and chips are always processed on the full wafer, the last step before testing and packaging is the precision cutting of the sensor out of the wafer, see Figs. 1.65 and 1.66. Sensor or chip cutting is normally done by a diamond saw and continuous water flushing. Experience tells that even for experienced companies, breaks and burns are common. Meticulous optical checking during quality control

is very important. To have more complicated cuts than straight lines, laser cutting is the state of the art, used for LHCb VELO[54] sensors.

Another way to 'cut' the sensor out of the wafer is **D**eep **R**eactive **I**on **E**tching DRIE. Trenches are etched through the full sensor thickness. The method is often used for 3D sensors, described in Sect. 1.12.7. To apply this method a handle wafer is necessary to avoid the sensor falling down at the end of the etching process. Since this method gives a very precise edge without any chips as resulting from diamond saw cutting the number of crystal defects is very low and the distance to the active volume can be reduced. The inactive sensor edge can be made smaller. This is interesting for pixel sensors, where staggering in z is not possible due to tight spacial constraints. It is often called slim edge or active edge. Another cutting method would be by CLEAVING. Cleaving and DRIE is also discussed in Sect. 1.6.2.

An example summary of all main process steps needed to process a double sided sensor is shown in Table 1.5, a real case.

1.9.3 Thinning

Important sensor parameters are (a) its *physical* and (b) its *active* thickness. Having a reduced "*physical*" thickness is to minimize mass and therefore multiple scattering in the final tracking system. This is especially interesting for smaller detectors where electronics and services are located fully out of the fiducial volume; e.g. NA11 and DELPHI. Processing physically thin sensors is a challenge for the manufacturer but also to finally assemble modules. The system becomes more fragile. The standard is a thickness of 300 μm but also 100 – 200 μm physical thick sensors have been produced.[55] Some companies leave a temporary thicker ring around the whole wafer for holding purposes during processing steps while the real sensor area is already thinned.

A reduced "*active*" thickness is to minimize the operation voltage, leakage current, thus power and to decrease drift paths; interesting for highly irradiated sensors to mitigate trapping (rf. Sect. 2.2.4).

Thinning is complicating the process at the manufacturers and there are limits to what can be handled in the full automatic processing lines. Thin wafers tend to break and hence decrease yield.

The different options of thinning are illustrated in Fig. 1.67.
(a) shows the standard ∼300 μm thick sensor for comparison;
(b) shows the same with a reduced thickness of 200 μm with full processing happening on thin wafers;
(c) shows a thin active wafer chemically bonded to a low resistivity handle wafer where the active thickness is reduced but not the physical one (mind the backside

[54] LHCb is one of the four detectors at the LHC, the sensors for the **VE**rtex **LO**cator VELO detector are shaped roughly as a half-moon and the round cutting was achieved by laser cutting.

[55] Recently 200 μm on an 8 in. wafer.

Table 1.5 The different processing steps of a real double-sided sensor are listed as an example

Step	Processing description	Days	Notes
1	Silicon selection and scribing		
2	Field oxide		Thermal at 1000°C; 0.9 μm dry+wet+dry, Ar/N
3	Image main implants	9	p^+ on junction side, $n+$ on ohmic side. Deposit resist, open windows, etch in BOE[a] 10 min (should reach silicon), strip resist
4	Reoxidation (cap oxide)	5	Thermal at 900°C; 0.2 μm wet+dry
5	Main implants		p^+ junction side: boron 80 keV 2E15; n^+ ohmic side: phosphorus 150 keV 5E15
6	Image secondary implants	8	p^+-stop on ohmic side (isolation), image n implant on junction side Etch in BOE 10 min (should reach silicon), back B^+; front P^+ at same time
7	Secondary implants	2	p^+ implant on junction side: boron 20 keV 5E14 n^+ implant on junction side: phosphorus 40 keV 5E14
8	Oxide partial strip	8	Remove 0.1 μm of SiO_2 in BOE sulphuric and rinse, then thin down
9	Reoxidation		Thermal at 900°C; 0.1 μm wet+dry Total time at 900°C; 1 h, in steam for 20 min Three steps: Ar, O_2, H_2O
10	Poly deposition		0.7 μm 2 h at 600°C
11	Poly implant	2	B^+ 1.3E14, 80 keV, both sides Boron resistivity about 30 $\frac{\Omega}{cm^2}$.
12	Silox deposition	8	0.8 μm
13	Image silox for poly contacts		High-dose implant, both sides or one at a time
14	Poly high-dose implant	4	Boron resistivity about 20–100 $\frac{\Omega}{cm^2}$ on both sides
15	Image of poly, plasma etch	2	Plasma etching 10 min
16	Anneal	3	30 min at 900°C; ∼3000 Å
17	PECVD[b]	6	0.1μm adjusted to have 0.3 μm over implant
18	Contact (via) opening		Wet etching BOE about 2.5 min
19	Wafer cleaning		Remove about 0.01 μm of oxide
20	Metal deposition	4	Sputtered Al—Si
21	Metal imaging, wet etching	4	
22	Passivation silox deposition	3	350°C thickness 1 μm
23	Passivation window opening		Wet etch that stops on metal
24	Cutting	5	Cut and wash with de-ionised water
25	Probing		

[a]BOE stands for **b**uffered **o**xide **e**tch. It is hydrofluoric acid (HF) with ammonium fluoride (NH_4F) and water to "buffer" the etching effect. Regular HF would etch too fast. The HF solution etches only SiO_2, not silicon.
[b]PECVD: Plasma Enhanced Chemical Vapour Deposition

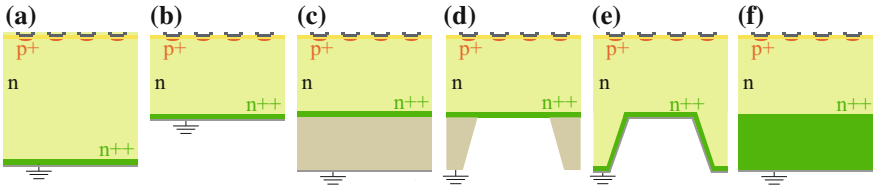


Fig. 1.67 Thin Sensors **(a)** Standard 300 μm thick sensor. **(b)** Standard, processed directly on a thinner wafer. **(c)** Thinner active thickness sensor processed on a handle wafer. **(d)** Same as (**c**) but handle wafer partially etched down later. **(e)** Front side processing up to metal layer on full wafer, then thinning of islands; then backside implant and last metallization steps. **(f)** Deep diffusion wafer – no advantage for physical thickness but active thickness can be chosen – processing on standard thickness

implantation happens before the wafer to wafer bonding); also an Epitaxially grown sensor grown on a handle wafer looks the same but even lower active thicknesses (down to 20 μm) can be done.
(d) shows how the handle wafer could as a late step be partially thinned down by an etching process to allow manufacturing on thick wafers but achieve low mass and good stability in the end – very interesting. The handle wafer can also be fully removed.
(e) is a variation of the (d) but without handle wafer. All steps on the front side up to the metallization are done prior to etching. The backside implantation is done after etching with the metallization as final step top and bottom. The difficulty lies in the placing and focusing for the metal process with thin and thick sensor areas.
(f) is a so-called **d**eep **d**iffusion *dd* sensor, which will be further discussed in Chap. 7 (very high doping concentration – very low resistivity – for a large part of the backside volume, i.e. n^{++} for an p-in-n sensor). The active thickness of the initial wafers can be chosen consciously; e.g. 300 μm physical thickness with 100, 200 or 250 μm active thickness interesting for the outer HL-LHC detector parts; or 200 μm physical with 100 or 150 μm active thickness interesting for the inner pixel layers. These wafers are procured from the wafer suppliers. Due to the deep diffusion process they come with a natural high oxygen concentration and are used only for FZ and not for Cz material. The full processing then happens on the standard wafer thickness comfortably for the manufacturer.

1.10 Readout Electronics – Strip ASICs

As will be discussed throughout the next paragraphs of this book, not only are the silicon sensors evolving, but the involvement and evolution of the front-end electronics is probably even more impressive. Initially in the 1980s the silicon crystals were held by a frame and signals from each strip were fanned out to amplifiers held by bulky printed circuit boards (see Sect. 3.2). In a collider experiment, where

space constraints in the vertex regions are very tight, a similar assembly is certainly not applicable. Fortunately, in the 1980s another development in industry and also in the universities, the *integrated circuit revolution*, made custom **A**pplication **s**pecific **I**ntegrated **C**ircuits ASICs broadly available. This initiated the opportunity for amplifier miniaturization and optimization. Without ASIC development, silicon microvertex detectors as we know them today would not have been possible. Small electronic chips around $5 \times 5\,\text{mm}^2$ were developed serving mostly 128 sensor channels. They were placed together with some passive components on small platforms of ceramic, beryllium-oxide or flex (polyimide). These platforms, called hybrids, were then attached on the end of daisy-chained sensors or sometimes even on top of the sensors itself. Photos of such assemblies can be seen in Fig. 4.7 (DELPHI hybrid), Fig. 4.6 (DELPHI module), Fig. 4.10 (DELPHI Ministrip module and pixel module) and in Fig. 6.33 (CMS components of a module).

Chip Features

Low noise and large gain were one of the most critical requirements for the readout electronics. Signal amplitudes only rise to around 3.6 fC for a minimum ionising particle in 300 μm thick silicon. Most chips work in the **D**ouble **C**orrelated **S**ampling[56] mode DCS. Small readout pitches of 25 – 100 μm require high connection density on sensor and chip side. The connection to the silicon sensors is being achieved by ultrasonic wire-bonding. An electron microscope picture of a wire-bond foot attached to a silicon sensor pad is presented in Fig. 1.68. Connection pictures in the DELPHI case are presented in Fig. 4.15 on p. 188, while CMS bonding is shown in Fig. 6.35 on p. 254.

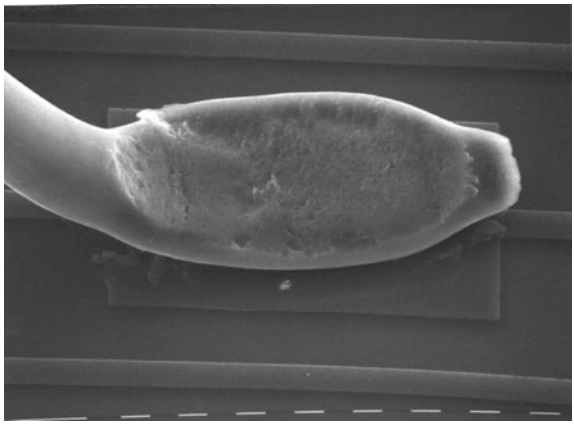


Fig. 1.68 An electron microscope picture of a bond foot. These connections are achieved via a thin wire (17 or 25 μm is standard width) bonded via ultrasonic power

[56] Double correlated sampling: Charge is integrated on a capacitor C_1 with signal absent, while the charge of the next time slot is stored in capacitor C_2 with potentially signal present. The charge difference between both capacitors represents the SIGNAL. This technique reduces leakage current effects and low frequency noise.

In addition the high number of readout channels of the chips requires multiplexing and in some cases the chip readout is only active, when a signal is present (*pulsed power mode*) to dissipate as little power as possible in the detector volume. In many cases channel signals are only read out when charges are above a certain threshold to avoid wasting bandwidth with zero information (*sparsification*). In the hadron colliders, especially the LHC and the future HL-LHC, radiation hardness of the readout chips was one of the most difficult challenges in the last millennium. Most chips are processing pure analogue signals, but some are already implementing the digitization in the front-end chip. Table 1.6 summarizes different choices and time evolutions.

Application Specific Integrated Circuits ASICs History

The next lines illustrate the historical evolution, improving noise values, adding features, increased readout speed, decreased feature size, always fighting against high capacitance and high-radiation environment.

DELPHI – LEP

The first **V**ery **L**arge **S**cale **I**ntegration VLSI[57] chip, the Microplex designed in Rutherford Appleton Laboratories and fabricated in Stanford University [349] was used for the MARK II [193] experiment at SLAC. It was a NMOS chip, the predecessor of all MX [19, 157] and Triplex chips used in the DELPHI (MX3, MX6 and TRIPLEX) and Opal (MX5) experiments at LEP produced in CMOS technology. These chips contained 128 channel charge-sensitive amplifiers, low noise, high gain chips for medium capacitance values with multiplex feature and pulsed power mode. In the last DELPHI upgrade stage, the TRIPLEX chip was introduced, being optimized for higher load capacitances, required for longer modules.

CDF – TEVATRON

The early MX chip version was also evaluated to be used in the CDF experiment at Fermilab but longer silicon modules, with higher load capacitor values initiated a new development in Berkeley, the SVX [122] chip series, which was used from 1992 to 2008 and beyond. The SVX was also the first chip designed for a high-rate environment. All SVX versions have the sparsification feature and use double correlated sampling.[58] For the later CDF and D0 operating since 2000, a new dead-timeless readout chip with analogue-to-digital 7-bit conversion on chip level, the SVX2/SVX3 [360] with a 46 capacitor deep storage pipeline plus one reference cell, was introduced. A total of 42 cells were used for storage, 4 as actual flagged readout buffer plus 1, which is finally utilized in the digitization process. The chip is purely digital and can be programmed. It can operate in the following modes: readout

[57] **V**ery **L**arge **S**cale **I**ntegration VLSI stands for the process of creating **I**ntegrated **C**ircuits ICs by combining thousands of transistor-based circuits into a single chip.

[58] In the first version, with DC coupling, even quadruple correlated sampling was used to suppress the direct current flowing into the amplifier.

Table 1.6 VLSI chips for strip detectors along the years

	Chip	Year	Feature size (μm)	Channels	Rad tolerant	Pulsed power mode	Bunch crossing	Output	Sparsification	Pipeline depth
DELPHI	MX1	1989 – 1990	5	128	∼10 krad	Yes	22 μs	Analogue	No	0
DELPHI	MX3	1990 – 1993	3	128	∼10 krad	Yes	22 μs	Analogue	No	0
DELPHI	MX6	1994 – 2000	3	128	∼10 krad	Yes	22 μs	Analogue	No	0
DELPHI	Triplex	1996 – 2000	1.2	128	∼10 krad	Yes	22 μs	Analogue	No	0
CDF	SVXD	1986 – 1989	3	128	∼12 – 20 krad	No	3 μs	Analogue	No	0
CDF	SVXH	1993 – 1996	1.2	128	1 Mrad	No	132/396 ns	7bit	No	32
CDF	SVX3	1995 – 1998	0.8	128	10 Mrad	No	132/396 ns	7bit	Yes	42
—	SVX4	2000 – 2003	0.25	128	50 Mrad	No	–	7bit	Yes	42
CMS	APV25	2000	0.25	128	20 Mrad	No	25 ns	Analogue	Yes	192
HL-CMS	CBC3	2016	0.13	254	> 100 Mrad	No	25 ns	Binary	Yes	512
HL-CMS	SSA	2017	0.065	120	> 100 Mrad	No	25 ns	Binary	Yes	512
HL-CMS	MPA	2017	0.065	16 · 120	> 100 Mrad	No	25 ns	Binary	Yes	512

This table cannot be exhaustive, it holds chip versions of the detectors discussed in this monograph. There are a multitude of other chips, e.g. the Viking or VA [224] is used on multiple occasions, the CAMEX64A used in ALEPH, the HELIX for LHCb, the ABCD for ATLAS and all of this does not even take into account the multiple pixel chips like the SP8 for DELPHI and the PSI46 for the CMS experiment. The field and developments are extensive with many special necessary solutions but also with many common aspects. The main "environmental" parameters are shaping time and load capacitance and the use of analogue or digital circuitry complemented by the feature needs like sparsification, dead-timeless readout, etc

"hit channel" or the "hit channel plus neighbours" or "all channels". The SVX3 is optimized for 396 and 132 ns beam crossing intervals. It is composed of a front-end (analogue) part separated from the back-end (digital) part.[59] The current SVX chip version, the SVX4[60] [116], is produced in 0.25 μm radiation-tolerant technology and with about 30% less noise than SVX3. The SVX4 has the feature to decouple individual channels via a program mask, this feature is especially profitable when a channel is noisy or connected to a pinhole.

CMS – LHC

For the LHC, with much shorter bunch crossing intervals, a low-noise analogue chip, the **A**nalog **P**ipeline **V**oltage chip APV25 [109, 125, 156] was developed. The APV25 was fabricated in 0.25 μm CMOS technology – Quarter Micron Technology – which is radiation tolerant, when obeying some special design rules, namely enclosed transistor layout techniques where thick transistor-relevant oxide structures are avoided. A photo is being presented in Fig. 1.69. This feature will be further discussed in Sect. 6.4.1 on p. 238. The chip is optimized for the 25 ns bunch crossing at LHC. As the name says, the chip produces analogue output, converted via analogue-opto-hybrids to laser signals and routed via ∼100 m optical fibre links to the CMS ADC (**A**nalogue to **D**igital **C**onverter) modules located in the adjacent service cavern. With about 25,000 electron–hole pairs per MIP in 300 μm thick silicon around 110 mV/MIP is expected at a sampling frequency of 40.08 MHz. The chip stores charges in a 192 elements deep pipeline to cope with the 4.8 μs trigger latency. The pipeline cells are connected to the **A**nalog **P**ulse **S**hape **P**rocessor APSP, which works in *Peak, Deconvolution or Multimode*. Despite bunch crossings every 25 ns, the APV peaking times t_P are 50 ns long, optimized for low-noise performance. The APSP discriminates signals from consecutive bunches. Overlapping signals are *deconvoluted* by adding three consecutive peak mode samples with correcting weights [125]. This deconvolution reduces the pulse width to 25 ns. In peak mode, each signal is read out without weighting and multimode corresponds to three consecutive peak mode signals. The signals of the 128 channels are multiplexed. A new important feature is the common mode[61] noise suppression. The APV can work with both polarities, meaning the amplifier works for electrons and holes, because it was originally designed for *p*-in-*n* silicon and **M**icro **S**trip **G**as **C**hamber MSGCs.

CMS for the HL-LHC

The full description of the CMS Tracker design for the high luminosity phase (HL-LHC) is described in Sect. 7.1. The main new challenge is the selective readout at 40 MHz for track trigger data (rf. Sect. 7.1) while, with same bunch crossing times integrating/shaping times do not really change. Due to the higher luminosity the sensor cell sizes are reduced thus **C**harge **S**ensitive **A**mplifiers CSA are optimised for

[59] Initially, these parts were even produced separately, only later a monolithic chip was produced.

[60] It was designed for the planned RUNIIb of the TEVATRON, which was finally abandoned.

[61] Common mode: Baseline fluctuations common for all channels of a chip.

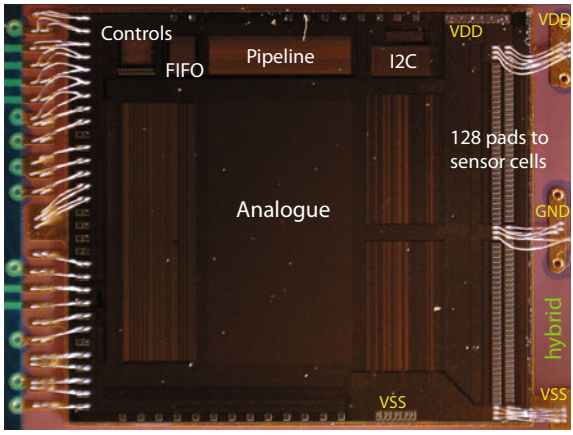


Fig. 1.69 Photo of the CMS APV chip. Wire-bonds are connected to the surrounding hybrid. VSS and VDD pads are to connect analogue and digital voltages. [Courtesy CMS, IC]

the new lower capacitance values and for electron readout as foreseen with n-in-p sensors. The **C**MS **B**inary **C**hip CBC is a consequent further development of the APV. It features a synchronous but only binary readout with an internal threshold comparator. The amplifier pulse shape peaking-time should be $<$20 ns with a return to the baseline within 50 ns – no deconvolution mode is necessary as for the APV. The noise target of the pre-amplifier and post-amplifier combined is $<$1000 electrons for 5 cm long strips with a leakage current up to 1 μA. A 130 nm process has been chosen and the chip will be bump bonded to the flex hybrid connecting power-, GND-lines, control communication, trigger and clock signals, data-out, and all sensor channel connections. The chip can accommodate 256 channels where the odd and even channels are alternatingly connected to two different sensors. Individual channels can be masked. The chip is able to correlate particle hits from one sensor (in the odd channel groups) to the other sensor hits (in the even channel group). The acceptance window sizes (number of channels) can be programmed on the upper sensor and also the centre of the window[62] can be freely programmed. In case of a hit in the lower sensor and a corresponding one in the upper sensor acceptance window the chip identifies a so-called 'stub' (close-by hits represent high momentum tracks, less bending in the magnetic field). Stub information relevant for the trigger is sent out at 40 MHz. The normal readout data, to be sent after a trigger decision, can be stored for a latency of 12.8 μs in dual-port SRAM cells with 512 cells plus 32 buffers where data selected after a L1-accept can be stored. A readout rate for the full data set (Trigger Level-1 accept rate) is foreseen to be 750 kHz. The CBC will readout the so-called 2S-module consisting of two strip sensors.

At lower radii PS-modules with $\sim$1.5 mm long macro-pixel- plus strip-sensors will be installed. The **M**acro-**P**ixel-**A**SIC MPA and **S**hort **S**trip **A**SIC SSA are processed

[62] The chip allows to offset the centre of the acceptance window to accommodate for tracks with an angle depending on the 3D location of the module in the Tracker.

in 65 nm technology.[63] The MPA, featuring the same 'stub'-finding logic as described above, has similar integration times and a binary readout (electron polarity) with internal threshold comparator. Obviously number of buffers for a latency of 12.8 μs and readout rate (40 MHz for stubs and 750 kHz for full data set) are the same as the CBC. The SSA features 120 channels and will also be bumb bonded to the flex hybrid. The MPA will be directly bump bonded on to the sensor with standard (not fine-) pitch industrial bump bonding, less expensive than fine pitch BB necessary for micro-pixel HAPS. The MPA has $16 \cdot 120$ channels. The SSA transfers the hit information to the MPA where the 'stub'-finding logic is located. With the macro-pixel granularity 2D points will be generated. PS plus 2S 'stubs' will allow high precision tracking and vertex identification with about 1 mm precision for tracks above 2 GeV at trigger Level-1.

Current Designs and Future Developments

With the smaller feature size, hence the possibility to place a larger number of transistors per area, it is possible to implement many more features directly on the chip level. Noise values of the different chip generations are presented in Table 1.3 on p. 44 where the noise–load capacitance correlation is discussed.

The initial chips were produced in a 5 μm CMOS process, followed by 3, 1.2, 0.8, 0.5 and 0.25 μm in the current collider experiments. Today, there are prototype chip runs in 130 nm CMOS and 65 nm for the future HL-LHC upgrades [205]. Dedicated exploration runs with 28 nm feature size are also envisaged.

As an interesting fact, it can be noted that already in the time of the SVX, the so-called "multiproject wafer" system was used. To reduce cost only design files are sent to companies, which then combine projects from different groups and subcontracts a number of different vendors for mask and final chip fabrication. This is a common practice today to enable the HEP community to have several design runs for a chip at reasonable costs. As an example the CMS HL-LHC strip chips (MPA and SSA) share a common engineering run with the common ATLAS/CMS RD53-pixel chip prototype.

For the LHC detectors, quarter micron technology is radiation tolerant, when utilizing special design rules/designs (enclosed geometry, rf. Fig. 6.23 on p. 240). In 2008, with first prototypes of feature size 130 nm available, the community anticipated that these advanced technology nodes (feature size 130 nm and lower) would be intrinsically radiation tolerant without enclosed geometry. Extensive tests have shown that minimal size transistors, used in high density digital logic blocks, are too sensitive to radiation and therefore each new technology must be certified for radiation tolerance.

Today, with adequately sized transistors, no special cell design is necessary.

[63] Early prototyping is done together with the future pixel chip from RD53 (see later) to save engineering cost.

1.11 Readout Electronics – Pixel Readout Chips – ROCs

Pixel ASICs often called ROCs **R**ead**o**ut **C**hips underwent a very similar evolution. Hybrid Pixel Detectors will be described in Sect. 1.12.1 with examples in Chap. 4 and Sect. 6.1. With a full 2D pixelated sensor the ROCs cover the full passive sensor and are bump bonded (BB)[64] to it (see e.g. Sect. 4.2 or 6.1). Metal balls (bumps) are 'placed' on the sensor or chip cells, e.g. by direct placement (jetting of individual balls or with a stencil holding multiple balls) or more conventionally by lithography processes (full metal coverage – photo-lithography – etching leaving only metal-pads (e.g. cylinder) – reflow to balls). The chip is then flipped and placed on the sensor, final connection is established by pressure and thermal soldering (also called reflow soldering) – more in Sect. 1.12.1 and photos in Figs. 1.70 and 1.73 on p. 108.

OMEGA –> CDFSP8/8b 06/95 for DELPHI

The first pixel chip installed in a HEP collider was the SP8 chip for the DELPHI **V**ery **F**orward **T**racker VFT (rf. Chap. 4) developed by RD-19 and the DELPHI collaboration. It was a true 2D chip processed in 3 μm CMOS feature size with zero-suppressed binary readout called **S**parse **D**ata **S**can SDS at the time. It combined charge amplifier, discriminator, leakage current compensation, shift register output and had a tunable delay. 8 chips have been connected together via a Kapton bus glued on top. Signals down to 5000 electrons could be cleanly measured. Many connection schemes have been tried in house due to the high cost of fine pitch BB at industry, while in the end, the 'cheap' flip-chip C4-process (**C**ontrolled **C**ollapse **C**hip **C**onnection – CCCC) was done in industry [216]. This was possible due to the relatively large bumps of 100 μm diameter[65] on $330 \cdot 330$ μm^2 pixels with a 80 μm diameter bond pad. The ROC area was $0.8 \cdot 0.8\,$cm^2 with $24 \cdot 24$ pixel cells or $0.8 \cdot 0.6\,$cm^2 with $16 \cdot 24$ pixel cells and 10+6 chips were placed on a pixel sensor totalling 8064 pixel cells. A photo of DELPHI SP8 chip section with bump balls is presented in Fig. 1.70. A photo of a module is displayed in Fig. 4.9 on p. 181. More details can be found at [144, 253, 254, 264].

PSI46, PSIDIG and PROC600 for CMS

For the LHC, fast and radiation tolerant pixel chips are necessary, even more important as for the strip system located at higher radii. The CMS pixel chips have been realized in quarter-micron technology (CMOS 250 nm) with enclosed geometry (rf. Fig. 6.23). Version PSI46 has served CMS well until the detector replacement in 2017 when Layers 2 – 4 have been equipped with the PSIDIG chip and an improved high-rate version for Layer 1, the PROC600. The first CMS pixel detector is described Sect. 6.1 and its upgrade in Sect. 6.2. The full details on the chip architecture are beyond the scope of this paragraph and can be found at [54, 149, 163]; the final application is described in Sect. 6.1. The chips have a zero-suppression mechanism

[64] Often also called flip-chip.

[65] A previous version had a bump of 150 μm which was finally decreased to reduce the capacitive load.

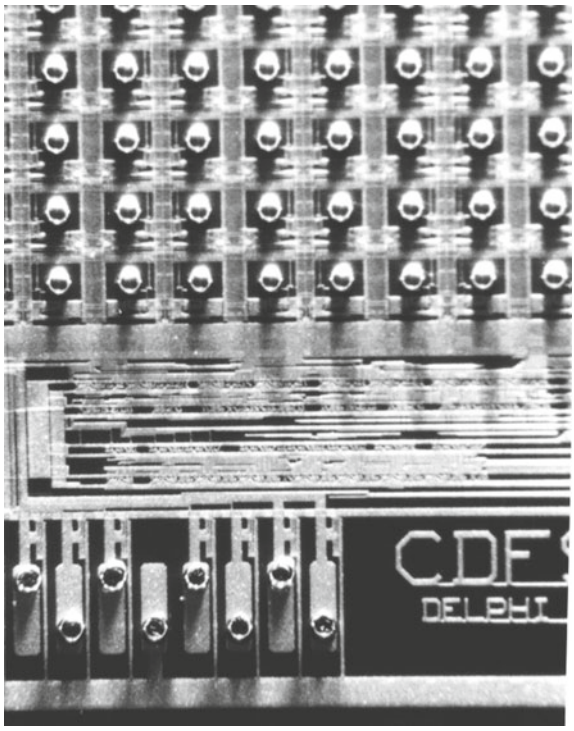


Fig. 1.70 SP8 pixel chip with bump balls in the pixel array and chip pads [144]

and the full pulse height information of a hit channel is passed to the subsequent readout boards. In addition, individual pixels can be masked and thresholds trimmed. The PSI46 transfers the analogue information (pulse height) while the PSIDIG chip has an on-chip 8-bit ADC and transfers the pulse height as digital information at much higher bandwidth. All chips have 52 columns by 80 rows serving $100 \cdot 150$ μm^2 pixel cells. The PSI46 readout chip uses a two stage preamplifier and shaper before the signal is passed to the comparator. When a pixel is hit (above threshold), the hit information is stored and the periphery is 'notified'. A 'token' is sent out and the information of a double-column is then transferred asynchronously to the periphery and being buffered there during the latency time until the trigger arrives. This simplifies the individual cell circuitry. The mechanism is called 'double column-drain'. Thresholds of about 3200 electrons have been achieved for the analogue chip PSI46.

For the CMS Phase I pixel upgrade new chips have been developed to cope with the beyond-design luminosities – higher pixel hit rates. Thresholds of about 1600 electrons have been achieved for the new chip – PSIDIG. It works well up to a hit rate of 120 MHz/cm^2 compared to 80 MHz/cm^2 for PSI46. The main changes are adjustment of buffer size to store more events during the trigger latency and implementation of an additional readout buffer stage to reduce dead time during the column readout. Each digital ROC then outputs data at 160 MHz to the **T**oken-**B**it-**M**anager chip TBM. On the TBM, 2 ROC streams are merged, hence the 320 MHz,

and a parity bit is added. So the data stream leaving the TBM is 400 MHz. The digital readout allows for much higher data transfer (320 MHz) to cope with the much higher numbers of channels fed through the same numbers of optical fibres. More details can be found at [164, 212]. One evolution further, the PROC600 changes the column-drain architecture itself allowing even higher rates up to 600 MHz/cm^2 hence the name **PSI ROC 600** MHz/cm^2 – more at [277]. Radiation tolerance has been improved from 20 MRad (PSI46) to 100 MRad (PSIDIG) to 600 MRad (PROC600) by clever architecture. The chips are flip-chipped to the sensor and wire-bonded to a flex hybrid where then a **T**oken-**B**it-**M**anager chip TBM sends 'token' to several chips, collects the data and then sends it to the high speed optical links.

All chips are thinned before BB to minimise material budget.

The RD53 Chip for the ATLAS and CMS in the HL-LHC Phase

For the High Luminosity LHC operation a new R&D collaboration RD53 [337] has been formed to realise a first version of a pixel chip. ATLAS and CMS might then adapt for a final version. The RD53 ROC will be the only active element on the CMS Phase 2 pixel modules! The chip has to cope with the following challenges:

- Extreme hit rates: up to 3 GHz/cm^2 (inner layer, Pile-up = 200)
  - Hit loss (dead time + buffer losses) <1% at 3 GHz/cm^2
- Extreme radiation tolerance: 1 Grad of TID; $2 \cdot 10^{16}$ $n_{1\text{MeV}}/\text{cm}^2$ fluence (inner layer after 3 ab^{-1}). Can be reduced to 0.5 Grad with one Layer-1 exchange.
- Small pixels $50 \cdot 50$ μm^2, real challenge to implement all features – 144,000 pixels per chip
- Threshold: 600 to 1200 e$^-$ and Time Walk: <25 ns
- Very long trigger latency: 12.8 μs
  - Hit buffering (increased by factor ∼100) – high density technology
- High trigger rate – full readout rate: 750 kHz equal to ∼1 TBytes/s
- Large pixel chip around $22 \cdot 18.4$ mm^2 with roughly 1 billion transistors
- Deal with serial powering
- Electrical links transmitting over long distance of 1 – 2 m to the optical links, located at higher radius (1 – 4 links of 1.28 GBits/s per chip)
- Low power density (∼1 W/cm^2 for the inner layers)
- Flexible enough to work with diamond, silicon planar, or 3D silicon sensor with different sensor cell capacitances.
- Fine pitch bump bonding

To match all the requirements on such a small 2500 μm^2 pixel cell, a very small CMOS feature size is necessary. The chip will therefore be realized in the 65 nm technology without enclosed geometry, but avoiding minimum transistor sizes in critical high speed digital functions. First measurements showed that the smallest transistors are less radiation tolerant. The small feature size is also necessary to minimise the power consumption and realise the necessary speed.

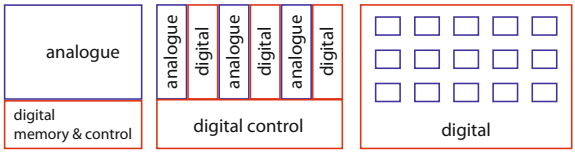


Fig. 1.71 *Left*, main part of the chip consist of analogue cells. Hit information is shifted to the periphery to be buffered and then transmitted. *Middle*, some logic is already implemented near the cells in a column like architecture including buffering. *Right*, cartoon-layout for the RD53 chip where the main real estate consists of digital cell with some distributed analogue cells. Speed requirements do not allow transport to the periphery any more. Hit information are stored locally

The 65 nm process is a strong technology node extensively used in industry, e.g. automotive. It will be available long-term and has significantly increased density, speed and complexity compared to 250 or 130 nm. Design tools and libraries exist to allow collaborative design with a larger HEP community. IP blocks can be developed at different places, distributed and later merged.

The chip will have a $\leq$4-bit (maybe 5) digital readout ($\leq$8-bit for low occupancy operation) and an on-chip data compression circuit to decrease transmission rate (still in the regime of high-speed memory chips) via electrical links to the optical links located at higher radii, due to radiation tolerance, at a distance of about 1 m. Also a shunt-LDO[66] will be implemented to work with serial powering. Figure 1.71 illustrates the evolution of chip layouts where hit information is not transmitted to the periphery and stored there but buffered and processed close to the analogue amplifier cells.

For the RD53 chip, 2 by 2-analogue cells[67] share buffers, hit processing like **T**ime-**o**ver-**T**hreshold ToT counting and management, **A**nalogue-to-**D**igital **C**onversion ADC, configuration logic – a 'simple' processing core. 8 by 8 of these pixel regions then share clock and trigger distribution plus control input and data output circuits. A $2.2 \times 1.84\,\text{cm}^2$ chip with $2.2{\cdot}1.64\,\text{cm}^2$ active area consists of 144,000 pixels sharing common periphery control, chip monitoring, power handling, data compression and control input and data output. The chip contains a differential cable driver with signal emphasis matching equalizer filters in the receiver end. At higher radii with lower hit rates, data from several chips are combined to a single electrical link. The data aggregation circuit is part of the chip itself; thus several chips feed data to a single chip, sending merged data with 1.28 GBits/s to the opto-conversion module operating at 10 GBits/s.

A correct layout is crucial to avoid digital interferences in the low-noise analogue front-end.

[66] The shunt-LDO regulator is a new regulator concept which combines a shunt and a Low Drop-Out LDO regulator, it is also implemented in the ATLAS FE-I4A ASIC [119].

[67] Maybe 4 by 4.

1.11.1 Chip Developments for the Future

Long term R&D is ongoing in industry and in the HEP community to realise a 3D chip where different layers have different functions and short vertical connections (Through Silicon Vias TSV). The reasons to tile chip functionality vertically are:

- reduce interconnect length
  - improve speed
  - reduce interconnect power
- reduce cross-talk (disentangle e.g. analogue and digital circuits)
- reduce chip footprint size

The industry target would be to have up to five layers with the following functionalities: Layer 1 digital electronics (ICs); Layer 2 Memory (RAM); Layer 3 Analogue electronics; Layer 4 Optoelectronic devices and waveguides; Layer 5 RF-MEMS + integrated antenna. This would allow dedicated processing in one layer and data in/out (I/O) in another with short path data transfer. One could even think about several digital layers.

The HEP dream would be a sensor layer connected (wafer to wafer bonding with TSVs) to a full analogue layer then connected to a digital processing layer and maybe to another one handling I/O. This would allow full optimization of all circuits individually without cross-talk and short paths. This technology is under discussion for the future linear collider. Of course all layers would be thinned to minimise material budget. The layers are often called 'Tiers'. First prototypes in the HEP community have been realised, like the **V**ertical **I**ntegrated **P**ixel VIP with three Tiers for the **I**nternational **L**inear **C**ollider ILC. Cost and yield remains a major challenge for such 3D integrated circuits.

1.12 Other Silicon Detector Types

In this section different silicon-based detectors are shortly described. Pixel devices HAPS, CMOS, HV(HR)-CMOS, and SOI sensors are introduced. Silicon drift detectors are used as an alternative 2D position sensing device. DEPFETs are developments where charge amplification is implemented with a different concept from standard devices. The 3D detectors are deviations from standard planar processing mainly developed for harsh radiation environments. LGADs also provide fast timing information.

1.12.1 Hybrid Pixels – An Alternative with a High Number of Channels

The full name is **H**ybrid **A**ctive **P**ixel **S**ensor HAPS. The technological difference between silicon strip sensors and silicon hybrid pixel sensors is marginal. The strips are segmented further down to a pixel structure (e.g. strips $100 \cdot 150\,\mu\text{m}$), the readout is adapted to small capacitances and is designed to handle DC coupling to the pixel implants. Consequently, the connection of the small structures to readout electronics is not possible with the standard wire-bonding anymore. The solution is a readout chip the size of the sensor with the same channel "pixellation" as the sensor which is placed in a sandwich configuration on top of the sensor. It is "bump bonded" or "flip-chip-bonded" to the sensor. The pads on the sensor and readout chip are treated with a special under-bump metallization. A lithography step opens holes in the applied photoresist over the pads and a dedicated bump metal is filled in the holes. After removing the resist, the bumps are shortly heated to reflow into balls. Finally the readout chip is flipped, aligned and pressed onto the sensor. Another short heating step connects readout to sensor. There are several recipes for a sturdy bump connection, e.g. Cu as bump on a PbSn surface or an indium bump on a **U**nder **B**ump **M**etallization "UBM" (sandwich of Ti, Ni, Au). A cartoon of the bump bonding processing steps and a HAPS scheme can be found in Fig. 1.72 while a picture of actual bumps and bump bond pads can be seen in Figs. 1.73 and 1.70 on page 103.

The obvious benefit of 2D information from pixel detector comes with the prize of a huge number of channels to be handled. The $\sim$1.75 m^2 pixel detector of CMS has about 127 million channels, about 13 times the channel number of the 206 m^2 strip detector. HAPS always comes with sparse readout, where only pixels with signal are read out. Due to the DC coupling, a current compensation circuit to counter the dark current is needed in addition to the pixel amplifier. The granularity is limited by the minimum size of the bumps but still suits all the needs of HEP detectors.

HAPS were first used in the DELPHI experiment in the forward region, the example is described in Sect. 4.2 with a pixel module picture, refer to Fig. 4.10. In the LHC area full silicon pixel detectors of several barrel layers and forward disks are used in ALICE, ATLAS and CMS (see Sects. 6.1 and 6.2).

1.12.2 CMOS Detectors – Monolithic Active Pixels – MAPS

Monolithic **C**omplementary **M**etal-**O**xide **S**emiconductor sensors or abbreviated CMOS or MAPS (Monolithic Active Pixels) are pixellated sensors favourably used in industry and medical applications to detect visible light – it is the successor of the **C**harge-**C**oupled **D**evices CCD sensors. CMOS devices are equipped as digital camera in every modern mobile phone. It works as an ionising particle sensor as well as for visible light. CMOS sensors fully integrate electronics and sensor volume in one wafer, better yet there is a thin epitaxial layer serving as solid state ionisation volume

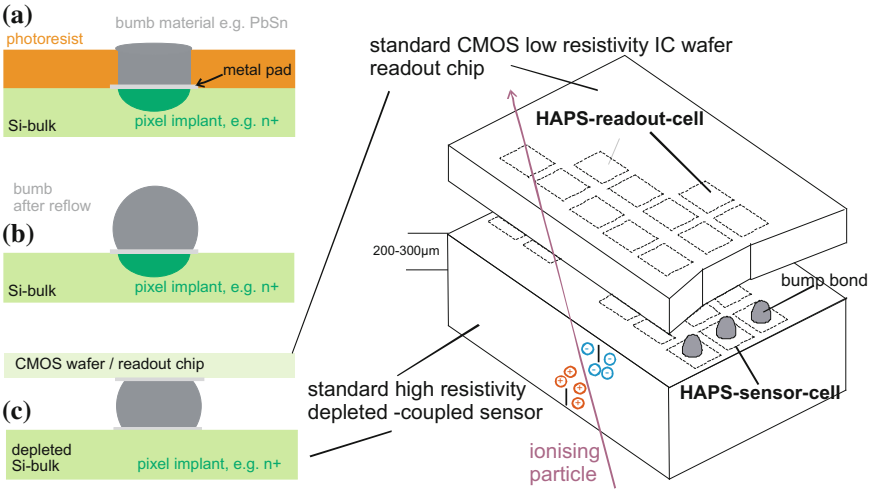


Fig. 1.72 Scheme of a **H**ybrid **A**ctive **P**ixel **S**ensor HAPS. A HAPS is a sandwich of a silicon sensor and a standard CMOS readout chip. The sensor is of the high resistivity-depleted DC-coupled type processed as described in Sect. 1.9.2. The readout chip is realized in standard CMOS technology on a low-resistivity wafer, the same size as the sensor, and its readout cells are distributed in the same "pixellated" way as the sensor pixels. The merging is realized via so-called "bump bonding" or "flip-chip-bonding". After preparing the pads with a dedicated under-bump metallization a further lithography step opens holes on each pad to place the bump metal (**a**), e.g. Cu or In. After removing/etching the photoresist the metal undergoes another temperature step, the so-called reflow to form balls of metal (**b**). The chip is then "flipped", aligned and pressed onto the sensor, warmed up for reflow, connecting sensor channels to readout cells (**c**)

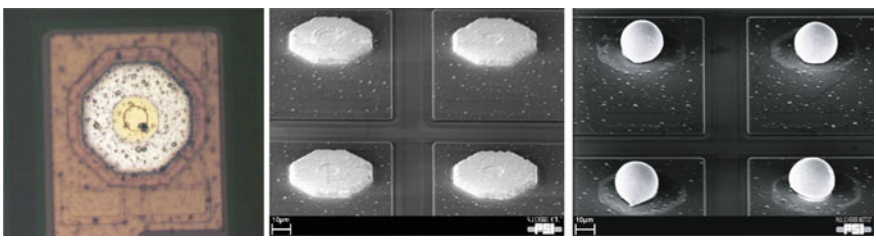


Fig. 1.73 Bump bonding at PSI for the CMS pixel detector. The *left* shows a bare contact on the pixel silicon sensor. In the *middle* part, an electron microscope picture of the structured indium bumps before the reflow process is shown. On the *right*, the bump ball after reflow is shown. The distance between bumps is 100 µm, the deposited indium is 50 µm wide while the reflowed bump is only 20 µm wide [51, 259]

with standard NMOS[68] electronics on top. The main difference to other solid state sensors used in HEP is the lack of any bias (depletion) voltage. A shallow depletion layer created purely by the *pn*-junction collects charges right away. Charges created in the EPI layer[69] are collected via diffusion to the *N*-wells. To achieve reasonable signal values, also true for light detection, EPI layers are rather "thick", in the order of 15 – 20 μm. The material must guarantee a long lifetime of the charge carriers to ultimately avoid trapping during the random walk. The still low signal is compensated by very low capacitance values, hence very small noise, finally resulting in reasonable signal-to-noise values up to S/N = 30. The concept only works for pixels and not for strips where capacitance values are much too high. Figure 1.74 gives a cross-section of a CMOS pixel cell and indicates the charge collection path. The NMOS electronic implementation directly in each pixel cell represents the first amplification step. The possibility of active electronic implementation directly in the sensor readout cells leads to a large variety of detector types. In-cell amplification, shaping, storage, comparator logic, sparsification are examples, and for the future linear collider several types are listed in Sect. 8.1.

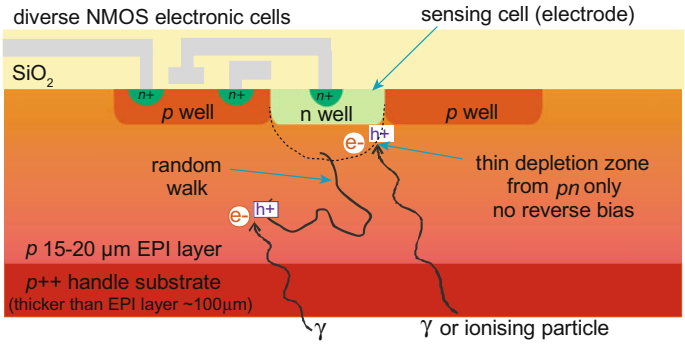


Fig. 1.74 Cross-section of a CMOS sensor, one pixel. The scheme nicely depicts an example of NMOS transistors and the *N*-well to collect electrons from ionisation or photo-effect. Electrons created inside the shallow depletion zones are fully collected while electrons from the EPI layer randomly walk towards the *N*-well and with an excellent lifetime behaviour, only some of them will be trapped. Nevertheless, CMOS devices have an excellent signal-to-noise ratio due to their very small capacitances and low currents, therefore the low noise compensates for the low signal

The ALICE detector at the LHC will completely replace its inner tracking system with MAPS sensors [284] based on triple-well technology thus able to use NMOS and CMOS. A 10 m^2 detector is planned consisting of thin (50 μm) monolithic

[68] or PMOS in *n*-bulk devices. Today, there are also more complicated designs/processes with "nested wells"/tripple well technology, allowing both NMOS and PMOS.

[69] EPI layer: Created through epitaxy or epitaxial growth: A thin (0.5 – 20 μm) layer of single crystal silicon is grown over a single crystal base substrate through chemical vapour deposition CVD. For HEP sensors thickness up to 60 μm are processed. These devices are different to the ingot growth and laser cutting described in Sect. 1.9.1.

active silicon pixel sensors with pixel cells of $30 \cdot 30\ \mu m^2$ and a moderate radiation tolerance of $10^{13}\ n_{1MeV}/cm^2$ sufficient for the expected fluences at ALICE. Several volts will be applied to the sensor part to increase the depletion zone but not to fully deplete the bulk. The electronic design is based on the so-called "Rolling Shutter" readout – a serial, row-by-row readout fast enough for ALICE but not for the other LHC experiments. The system is a perfect fit for ALICE. A further evolutionary step of this technology is described in Sect. 1.12.4.

1.12.3 Silicon on Insulator Detector – SOI

Silicon **o**n **I**nsulator (SOI[70]) technology, tested first in the SUCIMA [209] collaboration, has the potential to develop into the future star of sensor technology – see next section about HV-CMOS, where a variety is a plausible candidate for a full monolithic detector. The main difference between CMOS technology and fully depleted sensors is bulk resistivity and volume. A monolithic combination of electronics and depleted sensor would be very useful. For Hybrid pixels HAPS readout ASICs and sensors are processed separately and then merged by bump bonding. In SOI, the high resistivity, thick "sensor" silicon wafer with a SiO_2 surface will be chemically bonded to the low resistivity, thin "CMOS" silicon wafer. The processing then starts from the "CMOS" side where cavities are etched to the "sensor" part, to enable sensor side processing in a pixellated way and start CMOS processing afterwards in standard commercial IC technology. Basically it is a combination of the CMOS detectors and the hybrid pixels, as a monolithic device without bump bonding. With in-pixel processing it is possible to reduce power and increase speed. A challenge of the 'basic' technology is the so-called "back-gating" effect where the high operation voltage of the sensor part[71] affects the analogue transistor functionality operating at much lower voltages. Especially with thin oxide, transistor threshold V_{th} values are changed up to the point of non-functionality with full bias voltage applied. There are solutions to shield the CMOS circuits, e.g. by implementing an additional deep P-well 'through' the oxide insulation underneath all electronics cells. Another solution would a double-SOI sensor with "CMOS layer – burried oxide – silicon – burried oxide – depletion sensor zone".

A scheme of a basic SOI cell is shown in Fig. 1.75. Different to standard CMOS sensors SOI sensors are operated fully depleted and can accommodate NMOS and PMOS.

[70] Generally, the term SOI describes the basic wafer type *"silicon chemically bonded to insulator to silicon"* which is being used in other processes as well, e.g. use one silicon face as handle wafer. Thus, there is a potential mix-up with the term. Today 2017, SOI sensors are included in the HV-CMOS family, see next section.

[71] Voltage to fully deplete the sensor.

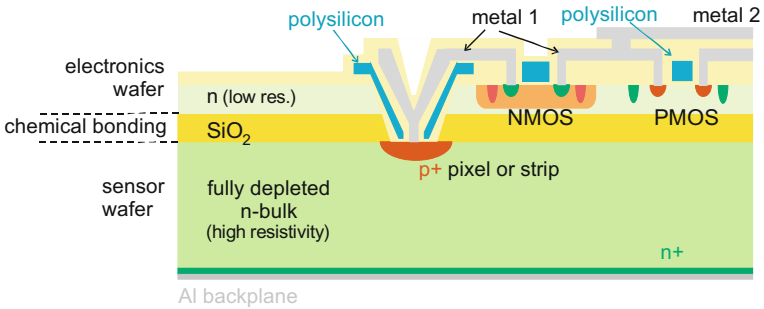


Fig. 1.75 Scheme of a silicon on insulator sensor. The scheme shows the basics of a SOI sensor. Passing charged particles create electron–holes pairs moving to the electrodes in a fully depleted high resistivity n-type sensor while the electronics are realized in a low resistivity n-type base material, separated by a layer of SiO_2. The connection of both parts is realized by etching while the electronics processing follows standard IC methods. In difference to CMOS devices the sensor wafer can be thick, of high resistivity and depletion is possible. NMOS and PMOS transistors are possible to be processed on the electronics wafer. In principle SOI can also be done with p-type sensor and electronics bulk

The basic production steps as used for SUCIMA are

1. bond[72] a high resistivity (sensor-grade silicon) wafer with a low resistivity **e**lectronics **g**rade **s**ensor EGS wafer with a SiO_2 layer by thermochemical bonding of the SiO_2 parts. One is called electronics volume, the other sensor volume
2. the low resistivity silicon wafer (electronics volume) side must be thinned down to about 1 μm thickness
3. etch a cavity in the electronics volume for the later pn-junction processing
4. create P-well on the CMOS side (later NMOS), add the p^+ channel stopper to P-well and n^+ channel stoppers to substrate (later PMOS) (not all litho steps are listed here)
5. create pn-junction: cover CMOS part with SiO_2, open SiO_2 between electronics and "sensor" volume, process p^+ on "sensor" (here implantation via diffusion)
6. apply an oxidized layer of polysilicon to protect the "sensor" side from contamination of the following processes
7. apply first metal layer, connecting pn-junction to CMOS and all necessary CMOS to CMOS connections
8. sputter SiO_2, open CMOS contact opening and apply a second metal layer to route CMOS contacts

The technology is finally on the way to maturity; it is currently under discussion as a candidate for the ILC detectors. As of today, SOI sensor technology is not a HEP standard.

[72] The bonding of the two wafers is done at the SOI wafer producer.

1.12.4 HV – CMOS/HR – CMOS

HV-CMOS (**H**igh **V**oltage CMOS) is probably the current most interesting concept with huge potential for the future, especially in full monolithic realization looking towards the ILC and FCC (see Chap. 8). The idea/hope is to exploit the large-scale production lines of CMOS technologies also on larger wafers (8 or even 12 in.) at lower cost and also avoid expensive fine pitch bump bonding as in HAPS, all along with in-pixel amplification. The pace of the development is high these days and this section is therefore to be considered a snapshot. The holy grail would be a full monolithic, low mass, radiation tolerant, small-cell pixel sensor at reasonable power, adequate speed …and all at reasonable/low cost!

The key to radiation tolerance and speed is to establish, like in conventional passive sensors, a depletion zone in ASICs technology to establish an electric field and thus make use of electron/hole drift instead of random walk as in previously described MAPS. With depletion depths of $w = 50 - 100\,\mu\text{m}$, reasonable signals of more 4000 electrons are achievable to be further amplified in simple circuits (or ∼1000 electrons with thin EPI on a handle wafer). This enables a fast and efficient charge collection within LHC time frames (25 ns) and a device with short drift paths is less vulnerable against charge trapping after radiation. Samples have shown radiation tolerance of several hundreds of Mrad and fluences of 10^{15} to $10^{16}\ n_{1\text{MeV}}/\text{cm}^2$.

Several different variants are under development/evaluation. Combinations of simple 'in-sensor' circuits plus standard ASICs similar to HAPS up to full monolithic designs and even a 'strip'-sensor version are being proposed. Full monolithic HV–, HR– CMOS (**H**igh **V**oltage–, **H**igh **R**esistivity CMOS) sensors are often also called **d**epleted **MAPS** DMAPS. The goal is to increase bulk resistivity ρ and applied voltage V_{bias} since the depletion depth w depends on them according to

$$w \sim \sqrt{\rho \cdot V} \tag{1.64}$$

The enabling technologies, different to standard CMOS processes, are:

- HV-NMOS, HV-PMOS, DMOS circuits as used in automotive and power management allowing for a moderate **'high voltage'** (∼120 V) – *High Voltage – HV-CMOS*
- 'high resistive' wafers, high compared to standard CMOS circuitry but still low compared to the standard HEP passive sensors, e.g. 1 – 2 kΩcm – *High Resistivity – HR-CMOS*
- Radiation tolerant circuitry, e.g. by utilizing small feature size (130 to 180 nm or smaller) and/or enclosed geometries
- Backside processing to allow for a backside 'HV' contact
- Stitching (more later) to allow for larger sensors, e.g. in the strip or large (quad) pixel case, as used in photo image sensor processes

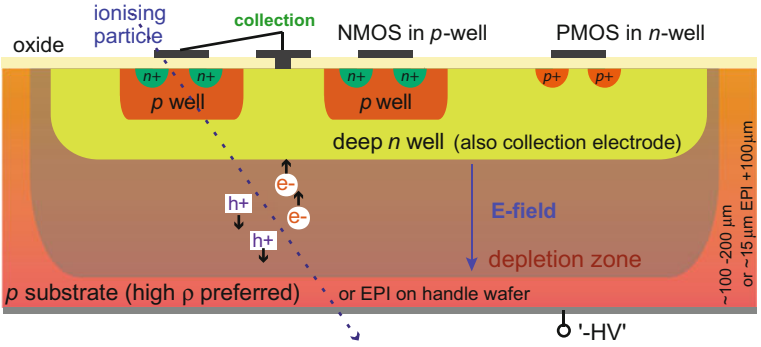


Fig. 1.76 Cross-section of a HV-CMOS sensor, one pixel cell. The picture shows the very basic version of a HV-CMOS sensor cell – a triple well configuration. The main feature is the **deep** *N*-well isolating/shielding the CMOS electronic cells from the 'high voltage'. It allows for PMOS and NMOS circuitry plus the application of a relatively low 'high voltage'. Many of the exemplarily shown CMOS cells form a full circuitry embedded in the deep well, e.g. a **c**harge **s**ensitive **a**mplifier CSA, etc. The sensor electrode is connected to the CSA. Higher circuitry complexity increases the cell size thus cell capacitance. More design variants and use cases are explained in the text

Voltages up to 120 V have been exploited, still only giving a 10 – 15 µm depletion depth for a standard CMOS resistivity of 20 Ωcm. Best are combinations of the above items. These technologies are not available in every FAB line.

Figure 1.76 illustrates the basic principle which comes in a large variety of implementations. The NMOS and PMOS cells are embedded in deep wells thus isolated/shielded from the depleted sensor bulk with 'high voltage' applied. The picture shows a deep *N*-well with embedded PMOS while the NMOS circuitry is furthermore confined in its own *P*-well (as in the previous MAPS Sect. 1.12.2). In this case, the *N*-well serves as the PMOS substrate but also as collection electrode. This is the so-called triple-well process. In principle, deeper nesting is possible allowing larger fill factors (see later), e.g. a quadruple-well process.

In other designs, the full CMOS circuitry resides fully in a deep *P*-well (embedded inside the *p*-substrate) plus a smaller dedicated *N*-contact in the *p*-substrate serving as sensor electrode or with the CMOS embedded/shielded in a **b**uried **ox**ide BOX – both variants are illustrated in Fig. 1.77. The point is to embed full CMOS circuitry (NMOS & PMOS) allowing reasonable complexity but 'protect' it against the higher, otherwise incompatible, voltages. The deep well (or deep well + electrode[73]) covers one pixel cell. One of the enabling technologies, the nesting of several deep wells is not available at every CMOS process. The first example (Fig. 1.76) allows homogeneous depletion zone growth and short drift paths but has a large capacitance compared to the second example with the dedicated cell+electrode concept,

[73] To limit the pixel capacitance, designs exists that split-up the deep *N*-well into several connected blocks.

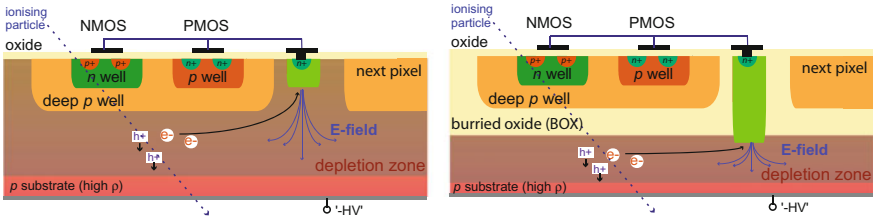


Fig. 1.77 Cross-section of a HV-CMOS sensor with extra electrode, one pixel. The full CMOS circuitry is embedded in a deep *P*-well (*left*) or *P*-well + BOX (*right*). The BOX concept is basically the same as SOI (see Sect. 1.12.3). The electrode is separate and connected to the CSA in the *P*-well. The cell capacitance is much lower thus noise can be lower and speed can be higher, but the drift path is longer and the depletion zone grows more lateral. Higher circuitry complexity decreases the fill factor and increases the drift path length. This design aims more towards full monolithic implementation

for which the depletion zone grows more laterally (Fig. 1.77). Higher circuitry complexity increases the cell size thus cell capacitance of example one (well = electrode), or decreases the fill factor (electrode/area) and increases the path length in example two (well + electrode).

It has to be mentioned that the input capacitance to the CSA is, in general, larger by about a factor two compared to a passive sensor in the HAPS concept due to the additional capacitances in-between the deep wells. Also the capacitance to backplane is high for thin devices.

HV-CMOS as CCPD (Capacitive Coupled Pixel Detector) or Strip Detector

The most simple circuit to be implemented in the individual pixel cells would be a **c**harge **s**ensitive **a**mplifier CSA, pushing the amplification stage or at least part of it into the 'sensor' – **C**apacitively **C**oupled **P**ixel **D**etector, CCPD more at [235, 237, 238]. The amplified signal is large enough to allow capacitive charge coupling and no direct ohmic bump bonding sensor-to-readout cell connection is necessary – fine pitch bump bonding (BB) is a cost driver. BB also limits the minimal pixel size! This opens the possibility to 'connect' the pixel cells capacitively with a simple glue connection[74] to a standard pixel HAPS readout ASIC (ROC – Readout Chip) responsible for the more complex processing like clustering, ADC, buffering, trigger processing, fast data transmission, sparsification, etc. A certain number of bump connections are still necessary to provide power. **T**hrough **S**ilicon **V**ias TSVs for power connection could be another solution. With the amplification in the sensor cell, the ROC is not impacted by the sensor cell capacity thus low threshold values should be achievable (e.g. 1200 electrons or lower[75]).

One can also daisy-chain many HV-CMOS pixel cells into a strip-like configuration with conventional wire-bonding at the sensor end to a standard strip readout-ASIC – in principle any arbitrary pattern is possible – more at [236]. This concept

[74]The glue has to be very thin, in the order of ten micrometers only, to achieve a good coupling.

[75]Threshold values well below 1000 electrons have been achieved.

is also called "**P**ixelated **S**trip **D**etector" PSD. With this, the former large area 'passive' strip sensor could be replaced by an active HV-CMOS sensor with the caveat of dissipating more power over the full surface. The cell location within the 'strip' can, in principle, be encoded in the CSA pulse amplitude or pulse length and with a dedicated analogue readout ASIC be decoded again. This feature would allow 2D information in a single strip within a single readout channel at the sensor end.

An even more interesting use-case would be for the (innermost) pixel layers, namely the increase of pixel granularity; e.g. four HV-CMOS cells of different pulse amplitude of $25 \times 25\ \mu m^2$ area could feed a single pixel readout ASIC cell of $50 \times 50\ \mu m^2$ achieving the position resolution of a $25 \times 25\ \mu m^2$ cell by decoding the different amplitudes.

Without 'stitching', the strip length or pixel sensor size, is limited by the maximum CMOS reticle[76] size, which is about $25 \times 25\ mm^2$. Some companies can overcome this limit. Function blocks (lithography photo masks, reticle-size of smaller) can be 'stitched' together by transferring these blocks into the photoresist on the wafer. Basically parts of strips, composed of small HV-CMOS pixel cells, can be tiled/stitched together. Other lithography mask blocks could then form the sensor edges. This technology is mainly being used for large photon imaging detectors where not only number of pixels but also sensor size per se matters.

The combination where the HV-CMOS (CCPD or strips) simply replaces the passive sensor is appealing because no substantial system design changes are necessary. The full complex CMOS logic stays in the readout ASIC with some amplification stage in the HV-CMOS sensor.

HV-CMOS full monolithic and its challenges

The real goal for the future is to establish a full monolithic sensor with all necessary logic implemented – see also [140, 234]. This could dramatically decrease material budget and has the potential to simplify the system level complexity. Lots or progress happened on this front, the situation is fluid, stay alert.

Here, the real estate of electronics and readout electrode has to be shared. The whole electronic circuit itself needs to be radiation hard thus of small feature size. Readout speed must be increased[77] and buffering for long trigger latency be provided to the level of current 'pure' ROCs. Cells of $25 \cdot 25\ \mu m^2$ seem out of reach with today feature sizes accessible to HEP and cells $50 \cdot 50\ \mu m^2$ would be very challenging.[78] This would largely relax the digital complexity of the HV-/HR-CMOS sensor. With larger electronics cells, the electrode thus capacitance would increase (rf. Fig. 1.76) or the fill factor sensor electrode to readout would shrink (rf. Fig. 1.77). Another concept would be to read out all events at full speed and outsource caching, trigger handling, etc. to an additional ASIC.

[76] Defined by the field of view of the lithographic equipment; normally featuring one or more individual electronic chips.

[77] The 'rolling shutter' concept as used in the ALICE upgrade is not adequate.

[78] It is already a challenge to implement all logic into an 65 nm ASIC for a $50 \cdot 50\ \mu m^2$ cell without sensor electrode and special deep well design rules – see Sect. 1.11.

Still, despite the numerous obstacles and challenges, the technology has a huge potential for future detectors to provide high granularity, low material budget and large surfaces at affordable cost. Next generation devices could be the good fit for the future linear collider, while more advancements and probably some steps down in feature size would be necessary to accommodate the needs for the **F**uture **C**ircular **C**ollider FCC (rf. Sect. 8.2).

1.12.5 Silicon Drift Detector

Figure 1.78 shows the concept of a silicon drift detector including the drift path of the created electrons and holes.

Silicon drift sensors are today used in the heavy ion experiment STAR [327] at Brookhaven [331] and have been implemented in the ALICE [330] experiment at CERN. These sensor types were first proposed in 1984 [117]. The concept is identical to gaseous drift chambers.[79] There are p^+ strip implants on both planes of the n-type sensor plus n^+ diode segments on the end of the sensor. A set of voltages is applied to the p^+ strips with suitable gradient decreasing towards the n^+ anode. In the n^+ implant region the charge cloud is then focussed to the anode. The y-coordinate is measured from the anode segmentation while the x-coordinate is represented by the drift time (typically µs/cm); a traversing time stamp is necessary. The homogeneity is very important and puts strict limits on the quality of the sensitive device with respect to material uniformity and small number of defects. The drift velocity needs to be predictable, therefore temperature needs to be monitored continuously and a strict calibration of the field homogeneity is necessary. The device delivers 2D information over large areas. ALICE covers 1.3 $m^2 \cong 30 \cdot 10^6$ pixels.

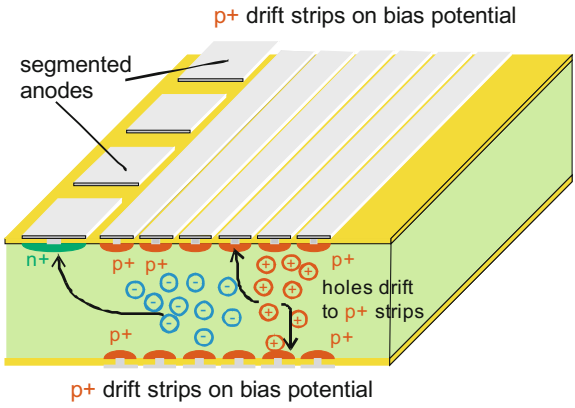


Fig. 1.78 The concept of a silicon drift sensor. Several p^+ strips on the same potential build a homogeneous field between sensor planes while the edge is structured with n^+ elements where the free charge carriers drift to; the Y-coordinate is defined by the n^+ elements while the X-position is defined by the drifting time. Depletion zone builds up horizontally

[79] Although, there is no amplification like in gas detectors.

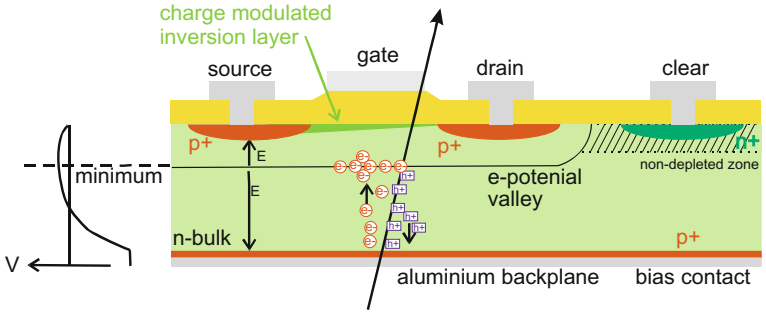


Fig. 1.79 The concept of a DEPFET sensor. The volume is depleted from the side n^+ strips down to the back p^+ implantation. The potential minimum of the sideways depletion is shifted towards the FET side by optimizing bias configuration. An ionising traversing particle creates electron–hole pairs in the depleted volume. Holes are lost in the back of the device, while electrons travel to and accumulate at the potential minimum below the external GATE at the so-called internal GATE, thus increasing charge density and thus modulating *source–drain* current of the FET. The electrons stay there until actively *cleared* [194]

1.12.6 Depleted Field Effect Transistors DEPFET Detectors

In contrast to all the silicon sensors previously described in this book the **D**epleted **F**ield **E**ffect **T**ransistors Detectors DEPFET sensor has an intrinsic amplification feature. The concept was proposed by J. Kemmer and G. Lutz in 1987 [166, 167]. It combines a fully depleted zone (detector bulk) and a field effect transistor FET (surface). A schematic of a DEPFET is displayed in Fig. 1.79. When applying a negative voltage on the back the depletion grows sideways, as for the silicon drift sensors described earlier. Electrons created by a traversing ionising particle now drift along the electric field versus the surface of the sensor. Different from all the other sensors, the charges are not collected directly and fed to an external charge amplifier. Electrons travel to and accumulate at the minimum potential zone formed below the gate – the so-called *internal gate*. This accumulation increases the charge density, thus increasing the FET (source–drain) current accordingly. Since the charge is not collected/'drained' another active mechanism is later needed to remove it from the internal gate. Actually quite high "CLEAR" voltages in the order of 12 – 20 V are needed. The CLEAR voltage increases the depletion zone towards the clear contact to drain the accumulated charges. Recent developments foresee a punch-through contact, which removes the charges from the internal gate – the constant clear design [194]. A correlated double sampling can be achieved by reading the current signal before and after a CLEAR. DEPFET sensors are used in pixel design. Low noise (just a few electrons per pixel) and intrinsic amplification allow them to be thin and make them ideal candidates for tracking detectors at the future **I**nternational **L**inear **C**ollider ILC. The missing radiation hardness prevents usage in a hadron collider, additional charges in the interfaces and bulk from radiation damage would deteriorate the FET behaviour; open channels and constant currents would be the

result. Figure 1.79 depicts the first proposed version of a DEPFET detector. In today's DEPFET sensors, since deep implantation is possible, additional deep N-wells are applied to localize the electron accumulation [195]. DEPFET applications will probably be used in the **X**-ray **E**volving **U**niverse **S**pectrometer XEUS [338] mission as well as in the vertex detectors of the future ILC. The most inner pixel detector of the Belle 2 experiment will consist of DEPFET sensors [32].

1.12.7 3D Silicon Detectors

Initially introduced in [229], 3D sensors give a good possibility to gather enough signal from the whole sensor volume for highly irradiated sensors, even after the onset of trapping. It will be shown in Sect. 2 that radiation levels of $\Phi_{eq} = 10^{15}\, n_{1\mathrm{MeV}}/\mathrm{cm}^2$ and beyond mainly affect the charge collection efficiency and are the limiting factor for detector operation. For $\Phi_{eq} = 10^{16}\, n_{1\mathrm{MeV}}/\mathrm{cm}^2$ a thickness above some tens of microns does not contribute anymore to the signal formation – the whole charge carriers are trapped before reaching the electrodes; or better said they are all trapped after a drift distance of some tens of micrometer and therefore the induced signal is significantly suppressed. The basic concept of 3D sensors is to etch pillars into the bulk volume and implant alternatingly boron and phosphorus doping via diffusion to the pillar walls or pillar filling with doped polysilicon. A scheme is displayed in Fig. 1.80. The sensors are then depleted horizontally instead of vertically thereby limiting the collection length to a few tens of microns but having a substantial volume at their disposal, for example the standard 300 μm thickness. In addition the small distance between p and n pillars allows for a low depletion voltage[80] even after irradiation. Lower depletion voltage also corresponds to lower power consumption. The technological challenges are the deeply etched pillars. Today **D**eep **R**eactive **I**on **E**tching DRIE (Bosch or cryogenic) is an industry standard procedure, but unfortunately it is much more expensive than the planar processing techniques. The 3D technology is becoming HEP standard. The capacity for a couple of square meter is available although the technology is not available among the big mass production sensor FABs of HEP. Figure 1.81 shows a pillar produced with the Bosch DRIE process. The wave-like walls clearly shows the alternating of etching deep down and depositing screeners (here polysilicon) to avoid a broad horizontal etch.

Diffusion processes finally implant the walls with p- or n-type dopants. This is also a two-step diffusion process, because alternating p- and n-type pillars are needed where one type of pillar is always masked, when the others are implanted. To operate the sensor, both bias voltage potentials need to be routed to the different pillars. Figure 1.82 shows a cut through a pillar with a zoom to the bottom part. The profile of polysilicon deposition and implantation is nicely visible.

The main challenge of 3D sensors is the additional etching process not compatible with the standard equipment used for planar technology. The technology nevertheless

[80] Reminder: $V_{FD} \propto distance^2$.

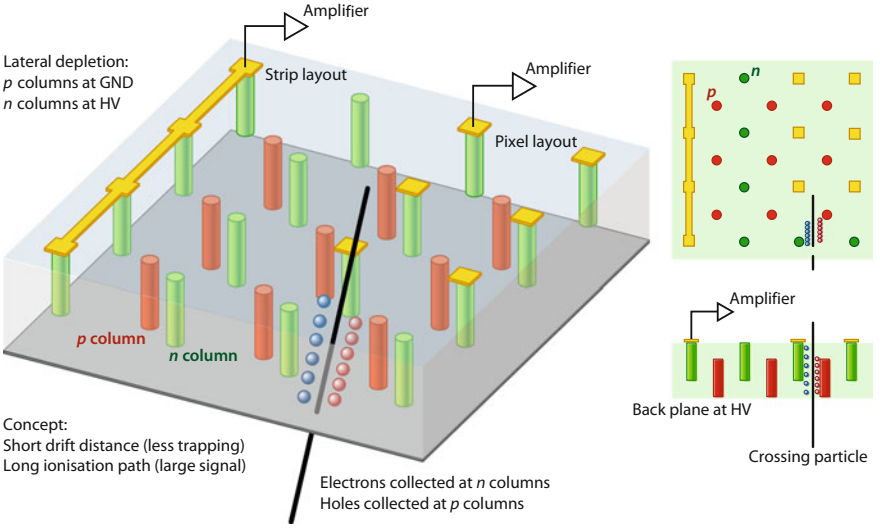


Fig. 1.80 Deviating from the standard planar sensor process, deep holes are etched into the silicon to achieve pillars finally serving as electrode junctions to span the depletion zone in a horizontal way instead of the standard vertical one. The electrons and holes travel a much shorter way and are therefore less sensitive to trapping. The pillars can be combined to a strip or pixel pattern. With the high radiation levels at the inner radii the normal configuration is the pixel one. Abbreviations: *GND* ground potential; *HV* bias voltage

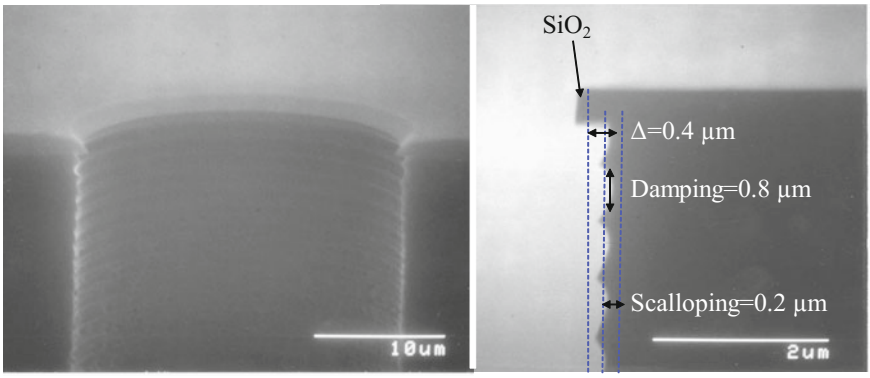


Fig. 1.81 The wave-like profile of the hole walls tells of the BOSCH etching method that was used. Frequent changes between etching and polysilicon deposit to mask from further etching ends in a deep hole with a uniform wall. The picture on the *right* shows the zoom of the one on the *left*. Courtesy of CNM-IMB (CSIC), Barcelona [62]

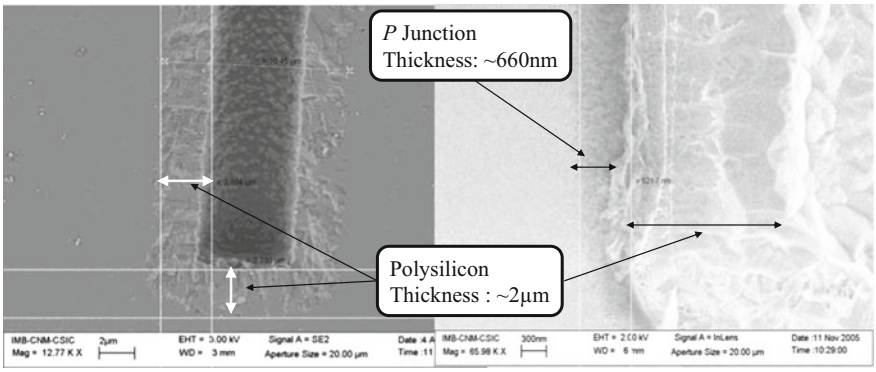


Fig. 1.82 The *p* and *n* implantation is realized via a diffusion process, the final polysilicon layers serve as protection. Courtesy of CNM-IMB (CSIC), Barcelona [62]

matured and today 96 sensors are installed in the **I**nsertable **B**-**L**ayer IBL detector at the innermost radius and high z location of the ATLAS pixel detector [287]. Also CMS/TOTEM have installed 24 3D sensor in the **C**MS-**T**OTEM **P**recision **P**roton **S**pectrometer CT-PPS detector, see more in [252].

Still mass production and yield is a challenge.

On the way to maturity as a sensor technology several different configurations have been explored and several proved effective.

The main species are

- single type versus double type columns
- single-sided versus double-sided processing
- full-3D versus not passing through columns

and several combinations are possible and in addition some are more suitable to also process an active edge, where the DRIE is utilized to 'cut' out the sensor without cutting. This is a very precise method to form the sensor edges with close to zero defects and no detrimental chips from diamond saw cutting.

The **S**canning **E**lectron **M**icroscopy SEM pictures in Fig. 1.83 display a 3D single-side **S**ingle **T**ype **C**olumn STC (*left*) suffering from a low field region between columns due to lateral depletion. In the STC scheme biasing is simpler than in a single-sided **D**ouble **T**ype **C**olumn DTC where GND and high voltage have to be routed on one side to the individual column electrodes (see Fig. 1.84). But the STC collection mechanism is not very efficient and the technology is not processed anymore. 3D **D**ouble-sided **D**ouble **T**ype **C**olumns DDTC (*right*) are more complicated but have a full electric field over the whole volume and GND and high voltage to bias the detector can be applied separately from both faces. Here, the holes are etched from both sides into the bulk to have n-type pillar from above and p-type from below, allowing for an easier masking and final voltage connection. This two-sided processing also compensates for mechanical stresses from the pillar etching. Figure 1.80

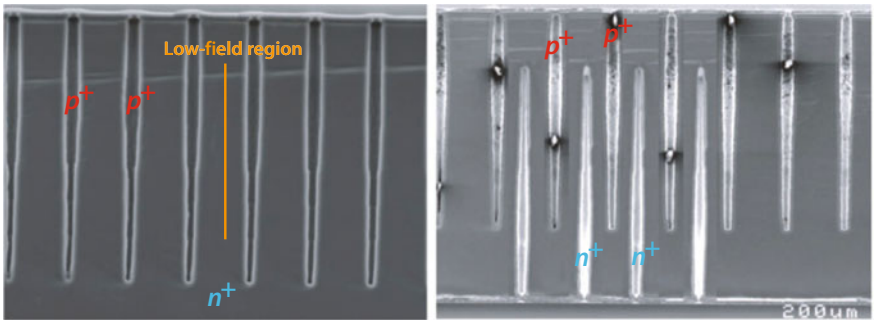


Fig. 1.83 The SEM micrograph pictures show a cut through the 3D sensors. Two different configurations are displayed, namely a single-type column and the newer, double-sided double-type column. The latter can overcome some low-field regions and simplifies the biasing scheme (top and bottom). Courtesy of CNM-IMB (CSIC), Barcelona [62]

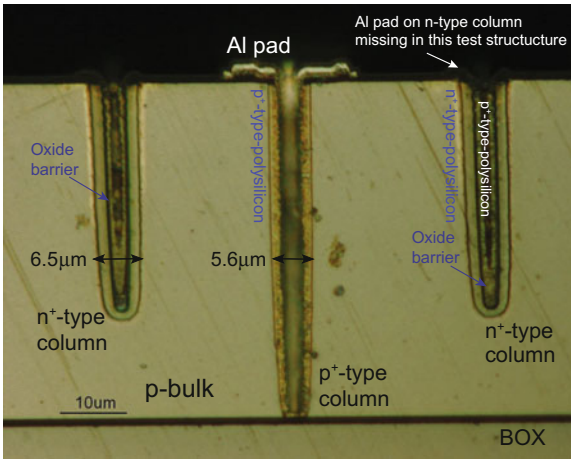


Fig. 1.84 The photo shows a cut through a test structure of a single-sided **D**ouble **T**ype **C**olumn DTC sensor. Columns with n- and p-doped polysilicon fillings are visible. First the n-type columns are etched, filled with polysilicon, which is then n-doped by a diffusion process. Then a global oxide layer is grown. Afterwards, the p-type columns are etched. To facilitate the following processing, all columns are filled with p-doped polysilicon, while the p-part inside the n-type columns, serving no function, is well separated by the oxide layer. Courtesy of CNM-IMB (CSIC), Barcelona

shows double-sided double type columns DDTC with not passing through electrodes (no full-3D). Another common nomenclature for a 3D-**D**ouble-sided **D**ouble **T**ype **C**olumn detector is 3D DDTC detectors (non through passing columns) and 3D-DDTC+ (through passing columns). The ATLAS IBL detector [287] uses double type column in full-3D and not pass through configuration.

An important difference between full-3D and double-sided 3D is the possibility to process active edges (see also Sect. 1.6.2 Slim Edge on page 49). Only in a

single-sided process where all columns are etched from one side the sensors can be oxide bonded to a handle wafer. In this case the etching process can etch trenches around the sensors at the same time the columns are processed. As the columns, the sensor edges can now also be doped or filled with doped polysilicon. After removing the handle wafer the sensors have smooth edges much better than diamond saw diced or laser cut ones. But without the handle wafer the individual sensors would simply fall away, thus trench processing is not possible in a double-sided process.

In addition to the different basic etching schemes described above, there are further combinations of active (connected to the readout) and passive (not connected) columns/electrodes. Most often more passive than active type columns are processed. Simulations are mandatory here to understand and optimise the field configuration. Often individual pixel cells are composed of more than one active column plus several passive ones.

There are additional "features" of these sensors which have to be taken care of in the full system design. The 3D pixels have much higher load capacitance compared to planar pixels of 'standard' thickness and therefore come along with a higher noise which has to be accounted for in the front-end chips. For very thin $D = 50 - 100\,\mu\text{m}$ planar sensors the capacities are comparable to 3D. While a planar sensor is sensitive in the full volume, the 3D columns themselves (empty or filled with polysilicon) do not contribute. Islands (column positions) of full inefficiency are clearly visible with incident angles of about 90°. This can be easily overcome by tilting the sensors or in a barrel configuration having them at higher z-position, away from the primary interaction where they are naturally hit at an angle (at high η). The second is the configuration for the IBL detector of ATLAS; where they are located at the end of the staves.

Small 3D Cell Size – the Challenge for the HL-LHC Detector Upgrades

3D sensors are candidates for the innermost layers due their high radiation tolerance and the high radiation environment at that location (see example in Fig. 7.1). Such high radiation goes hand in hand with a very dense particle environment (high occupancy) requiring very small pixel cells to be able to resolve individual tracks. Cell sizes of 50×50 and/or $25 \times 100\,\mu\text{m}^2$ are being considered for the HL-LHC detectors. This in turn requires a narrow 3D column diameter, especially in multi-electrode configurations for individual pixels cells. To reach a reasonable depth the aspect ratio depth:diameter becomes difficult to process but with smaller depths the concept loses its advantage. Currently aspect ratios of (depth:diameter) 25:1 are achievable, e.g. 8 μm diameter holes down to a depth of 200 μm. Cryogenic DRIE is being investigated to achieve aspect ratios of about 40:1 in the future. In addition such small cell sizes pose more difficulties in cell isolation and connectivity; a common challenge for planar and 3D sensors. Obviously also the cell load capacity will increase with smaller column distances.

1.12.8 Low Gain Avalanche Detectors – LGAD

Section 2.2.5 describes charge amplification encountered in highly irradiated diodes and strip detectors. This finding triggered the evaluation if one could utilize the amplification effect in a controlled way. A detector design to achieve a moderate gain has been studied and achieved within the RD50 collaboration – it is called **L**ow **G**ain **A**valanche **D**etector LGAD, and the reader is referred to [56, 101, 232, 233, 262, 263] for more details. Generally, electric field strengths above $2 \cdot 10^5$ V/cm activate the impact ionising multiplication. With the high signal (signal-to-noise), in-chip comparator thresholds are crossed fast (small "jitter"), thus ultra fast timing (several tens of picoseconds) is manageable.

Today, 2017, ultra fast timing detectors are in high demand for the LHC very forward detectors and the future HL-LHC detectors (ATLAS and CMS).

The concept is to have a localised controlled amplification stage at the readout cell, implementing a high n^{++}-strip[81] with an underlying doped p^+ deep implant, all embedded in a p-bulk with p^{++}-backplane. This creates a high field between n^{++} and deep p^+ implants – the multiplication layer. Figure 1.85 shows the scheme of an LGAD with two different cell varieties. The mechanism is based on the avalanche photodiode technology also used in **Si**licon-**P**hoto-**M**ultiplier Si-PM. The high field necessary to trigger the avalanche is realized due to the highly doped n^{++} and p^+ implants plus the bulk fields squeezed to the n^{++} implant edges. The doping concentration is therefore critical to reach a high enough field of 30 V/μm, a concentration of at least $5 \cdot 10^{15}$ cm^{-3} is necessary. A small increase in doping concentration translate to substantially higher gain – or an early breakdown, the doping concentrations must be precise. In addition, the electric field across the bulk forces the electron drift to the amplification stage.

Active thicknesses of LGADs between 50 and 300 μm have been realised, although thin sensors are clearly preferred for fast timing due to the increased bulk field strength (easier to over-deplete) and highly reduced drift length/time and thus reduced trapping effect – with the high field the drift velocity is also saturated. The thinner the sensor, the faster is the signal, the faster the rise time, the faster the slew rate dV/dt! Leakage current and noise are correlated with gain. Gain has to be optimized since noise rises faster than signal at higher gains.

Individual cells are mostly kept small to have a low current and capacitance (mind, with a thin detector cell, capacitances are on the high side). High currents could also evoke cell breakdown due to the internal gain. **J**unction Cells with **T**ermination **E**xtension JTE, distributing the high field more homogeneously, as shown in the right part of Fig. 1.85 are more reliable [101].

The JTE serves another purpose, namely making sure the charges generated "in-between" cells (inter-strip or inter-pad region) are not amplified. An amplification of particle hitting the inter-strip section would generate a "late" signal with respect to the initial drift thereby diluting the timing information. This has an unsatis-

[81] The n^{++} has a higher doping concentration than the deep p^+ deep implant.

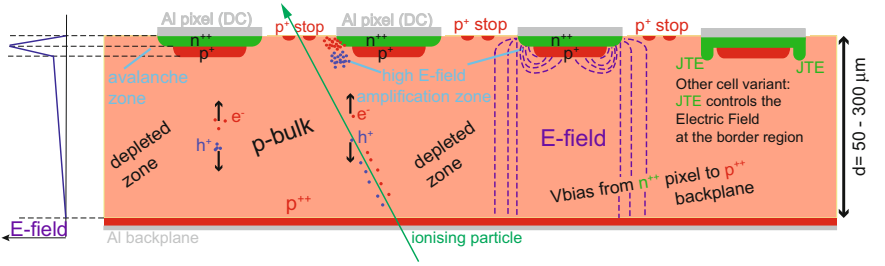


Fig. 1.85 A schematic of a segmented **L**ow **G**ain **A**valanche **D**etector LGAD is shown emphasising several different features – electric field, amplification, cell isolation, cell varieties. Cells can be strip, pads or pixels – mostly millimetre sized pads these days. The bulk is fully depleted. The amplification stage is localised between the deep p^+ implant and the n^{++}-electrode, see field configuration on the right. As for an n^{++}-in-p sensor the cells need to be isolated, here by p^+ stops. On the right another variant with a Junction Termination & Guard Ring is shown – **J**unction **T**ermination **E**xtension JTE. The JTE controls the electric field at the border region. A significant high bias voltage is applied between n^{++} cells (pixels or pads) and p^{++}-backplane

factory implication on the sensor fill factor since the inter-strip part is then insensitive. Fill factors up to 95% haven been achieved.

Time resolution, the relevant parameter, is mainly limited by Landau fluctuations and the "jitter" term (noise and steepness of the signal) and TDC[82] digitization. Signal lengths are of the order of a couple of nanoseconds. At relatively low gain, time resolution scales with gain (high gain – high signal – faster rise – smaller "jitter"), and for gain values of 20 – 30 a time resolution of $\sigma_t = 30\,\text{ps}$ has been realized [55].

The detectors are operated at low temperature (T∼-20°C); leakage current decreases thus does shot noise, gain increases thus slew rate dV/dt increases with an all in all net gain of signal-to-noise thus decreased σ_t.

Today, 2017, the situation is fluid and many optimizations and developments are to be expected; especially with respect to radiation tolerance.

Radiation Tolerance of LGADs

The standard radiation effects (leakage current, depletion voltage, trapping) and their mitigation strategies, as discussed in Chap. 2, are valid for LGADs. In addition the p^+ deep implant is significantly effected by acceptor removal, meaning the effective doping concentration N_{eff} is reduced below the value necessary for a high field resulting in amplification. Boron atoms from the lattice become interstitials thereby turning ineffective as acceptors. At fluences of $\Phi_{eq} = 10^{15}\, n_{1\text{MeV}}/\text{cm}^2$ the gain is significantly reduced, still a σ_t of 55 ps has been measured. At $\Phi_{eq} = 1.5 \cdot 10^{15}\, n_{1\text{MeV}}/\text{cm}^2$ the LGAD behaves more and more like a normal planar detector. At around $\Phi_{eq} = 2 \cdot 10^{15}\, n_{1\text{MeV}}/\text{cm}^2$, depending a bit on initial doping concentration, the acceptor removal is basically complete. The goal, though, is to be radiation tolerant up to $\Phi_{eq} = 10^{16}\, n_{1\text{MeV}}/\text{cm}^2$, e.g. for HL-LHC. For early studies, the reader is referred to [180] but the investigation is in flux.

[82] TDC **T**ime to **D**igital **C**onverter.

Table 1.7 The table summarizes the advantages and disadvantages of the main different sensor technologies. It does not claim to be exhaustive

Technology	Advantage	Disadvantage	Usage in HEP
Strip detector	Large area possible	Mostly 1D only	NA11, L3
	High precision (< 5 μm possible)	High strip capacitance	ALEPH, DELPHI, OPAL
	Standard processes/inexpensive		CDF, D0, Belle I&II[2]
	Many suppliers		CMS, ATLAS
	Fast		LHCb, ALICE
			ILC, HL-LHC
Hybrid pixel	2D info, small pixels	Bump bonding ↔ large pixel	DELPHI
	Standard for small r (vertex)	Electronics on top (material)	ATLAS, CMS
	Fast	High channel number	ALICE
	Electronics ∼commercial IC	Expensive Bump bonding	HL-LHC
CCD	2D info, small pixels	Slow	
	Industry standard	Not radiation hard	SLAC
	Thin devices possible	High power consumption	
	High spatial resolution	Needs many different voltages	
	Integrated amplification		
DEPFET	2D info	Only small devices	
	Integrated amplification	Not radiation hard	BELLE II
	Low power consumption	High power at CLEAR	ILC(?)
	Thinning possible	Slow	
	Low noise	No industry standard	
CMOS	2D info, small pixels	Only small devices	
	Standard IC process	Short technology lifetime	
	Integrated amplification	Small signal	ALICE upg.
	Low noise	Only NMOS (in sensitive area)	ILC; CLIC (?),
	Thin	High T dependence ↔ cooling	
	Low power	Only moderately rad tolerant	
	High resolution		

(continued)

Table 1.7 (continued)

Technology	Advantage	Disadvantage	Usage in HEP
Silicon drift	2D info	Slow	
	Large area	Good homogeneity needed	STAR
	Standard process	Exact calibration necessary	ALICE
		Double-sided process	
3D sensor[a]	V_{FD} independent on thickness	High depth/diameter ratio difficult	ATLAS IBL
	Very radiation hard	No standard planar process	AFP (ATLAS)
	Small collection distance and time	High capacitances	CT-PPS (CMS)
	Low power		HL-LHC(?)
SOI[b]	Reduction of total thickness	No HEP standard	
	No Bump bonding necessary	Back-gating effect	ILC; CLIC (?)
	Electronics and sensor separated		
HV/HR CMOS[c]	2D info	No global/full IC standard	ILC; CLIC
	Potential for full monolithic	Not yet mature	HL-LHC(?)
	Potential for low material budget	Speed!?!?	FCC(?)
	Potential for low cost, no BB but glue	Higher capacitances	
	Configuration for strips & pixels	Power	
	Radiation tolerant		
LGAD	Amplification	No HEP standard yet	CT-PPS (CMS)
	Fast timing	Fill factor < 1	AFP (ATLAS)
	Segmentation	Acceptor – removal in p-layer	HL-LHC
		High rad. tolerance to be shown	

[a]The 3D technology is becoming HEP standard
[b]SOI merges advantages of hybrid pixel and CMOS in a monolithic structure
[c]There is a SOI/HV-CMOS variant which looks very intriguing

One idea to counteract the acceptor removal effect is to use gallium instead of boron as dopant, which is more difficult to de-place. Another possibility would be to add carbon, filling the interstitial places, keeping N_{eff} high. This is somehow a paradigm change since in the past (see also Fig. 2.10 on p. 149 and the associated section), we wanted to avoid carbon – difference is, in the bulk we want to have low N_{eff} (avoid carbon), while in the p^+-deep implant, we want high N_{eff} (add carbon).

1.12.9 Technology Advantage – Disadvantage – Usage

The main advantages and disadvantages of the different technology types are summarized and examples of historic, current and future usage are given in Table 1.7.

1.13 Some Last Words About the Design of Detectors for High Energy Physics

The main purpose of silicon detectors in high energy physics is the precise and efficient measurements of particle tracks without disturbance of the track itself. What are the requirements and obstacles of a real detector in a nutshell? Low mass, low noise, high signal, fast response, low power and radiation tolerance are required. Unfortunately the requirements are partly orthogonal, e.g. reduction of mass and radiation tolerance leads to thin detectors resulting in a low signal resulting in a low noise requirement resulting in increased power resulting in additional mass in cabling and cooling resulting in more mass.

The DELPHI detector, the largest of the LEP silicon detectors, is described in Chap. 4. Chapter 5 describes the world's largest detector in the year 2000 at the TEVATRON at Fermilab, the CDF II detector. In Sect. 6.4 examples of measures to reach an acceptable level of radiation tolerance are presented. In Sect. 8.1 some innovative strategies are visioned to realize a silicon detector for the ILC, where radiation is low but multiple scattering is a major problem, with ultra-high precision physics planned. The next level of radiation tolerance challenge and very high occupancy at the high luminosity HL-LHC is described in Sect. 7.1.1.

1.14 Some Always Unexpected Problems Along the Way

Unlike space technology, all the HEP experiments always target for the newest possible technology. Silicon areas need to be large, sensors and electronics must be radiation resistant, space requirements are painfully tight. Only specialized companies produce sensors often in close collaboration with the HEP community. The

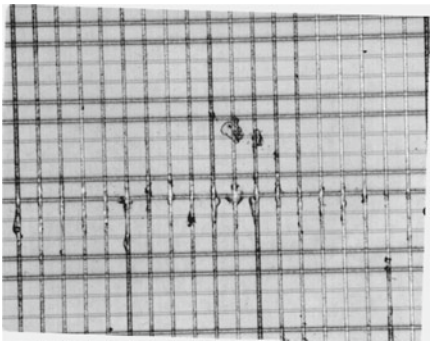


Fig. 1.86 The picture shows a "flake", a residual substance, remaining from a plastic packaging material. The real problem was the chemical reaction with the SiO_2 coupling capacitor isolation, resulting in a short circuit between p^+-implant and readout electrode, in short a pinhole

technology is similar but also sufficiently different to standard ASIC production and light structures are preferred to very robust ones. Also high electric fields need better material purity, well-defined doping profiles, voltage robust designs, etc. The HEP community encountered many problems along the way and this subsection gives an exemplary, but small, picture of some problems encountered in the last 20 years.

Pinholes Creation During Shipment

DELPHI, one of the LEP detectors pioneering the use of silicon microstrip and pixel sensors as tracking and vertexing detectors faced a mysterious sensor behaviour at its last upgrade. Perfect sensors, fully qualified at the company and subsequently delivered, showed a large number of pinholes, not at all compatible with the vendor measurement! It has to be acknowledged that measurements at these time were done manually strip by strip [17] and investigations were demanding. As a side effect, this incident led to the installation of fully automatic probe stations in the university laboratories.

It was finally identified that the plastic package of the sensor was reacting with the thermal-grown oxide thereby creating pinholes. The "flakes" caused by the "evil plastic" packaging are shown in Fig. 1.86. Because of this problem, the 1996 upgrade was partially delayed by 1 year to 1997.

Low Inter-strip Resistance

Low Inter-strip Resistance at CDF

Another incident with sensor degradation due to shipment was encountered during the CDF II sensor production. Sensors arrived at the quality control centres with regional low bias resistors and high leakage currents hinting to low inter-strip resistances. A dedicated inter-strip resistance measurement confirmed the suspicion. After some investigation the problem was attributed to be lint content in the packaging causing charge-up in the passivation oxide. Bathing in de-ionised water cured the charge-up and therefore the inter-strip resistance problem. Measurement results are shown in Fig. 1.87.

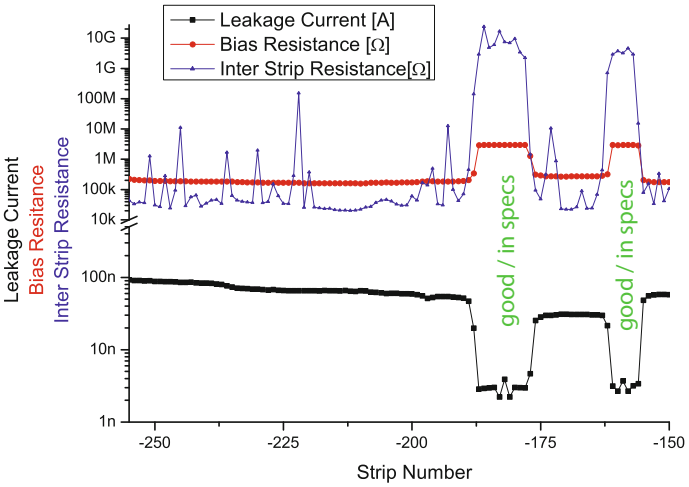


Fig. 1.87 Low inter-strip resistance after shipping. Lint content leads to charge-up. Low inter-strip resistances were indirectly detected by leakage current plus bias resistance measurements and were confirmed by dedicated additional measurement runs [35]

Low Inter-strip Resistance at CMS

During a certain period sensors were delivered to CMS with inter-strip resistances in the order of megaohms instead of gigaohms. The dedicated process control spotted the defect on dedicated test structures. It was immediately confirmed on the large sensors. Several hundred sensors were affected and accordingly rejected. A contamination in the passivation oxide created an accumulation layer in the Si – SiO_2 interface beneath "short-circuiting" the strips. The failure was traced back to a small parameter variation in one machine in the production line. The inter-strip resistance history is displayed in Fig. 1.88.

High Current on Isolated Strips

High leakage currents on isolated strips are a source for shot noise and in some cases with an irregular time structure due to micro-discharges, also a source of high common mode noise, affecting full chips (128 channels, see [60]). Such failures could also create fake hit signals. The cases in this chapter are just examples of leakage current origins and are not meant to be exhaustive. As a reminder HEP sensors are fully depleted devices with high electric field strengths on the strip side and especially on the edges of the implants or in the CMS case on the edges of the implants and the aluminium strips (Sect. 6.4.2). This region is kind of delicate and deviations as point-like structures disturb the field configuration and can cause local breakdown. Figure 1.89 shows a CDF II example where small protrusions on the p^+ implant strips caused a current increase from several nanoampere to several microampere. The exact defect location was identified by an IR camera finding a hot spot on the sensor.

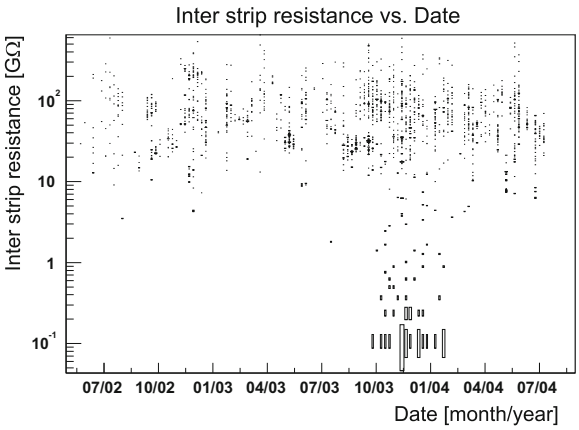


Fig. 1.88 History of inter-strip resistances of sensors delivered during 2 years for the CMS tracker. Late 2003 and early 2004 several sensor batches showed a low inter-strip resistance [35]

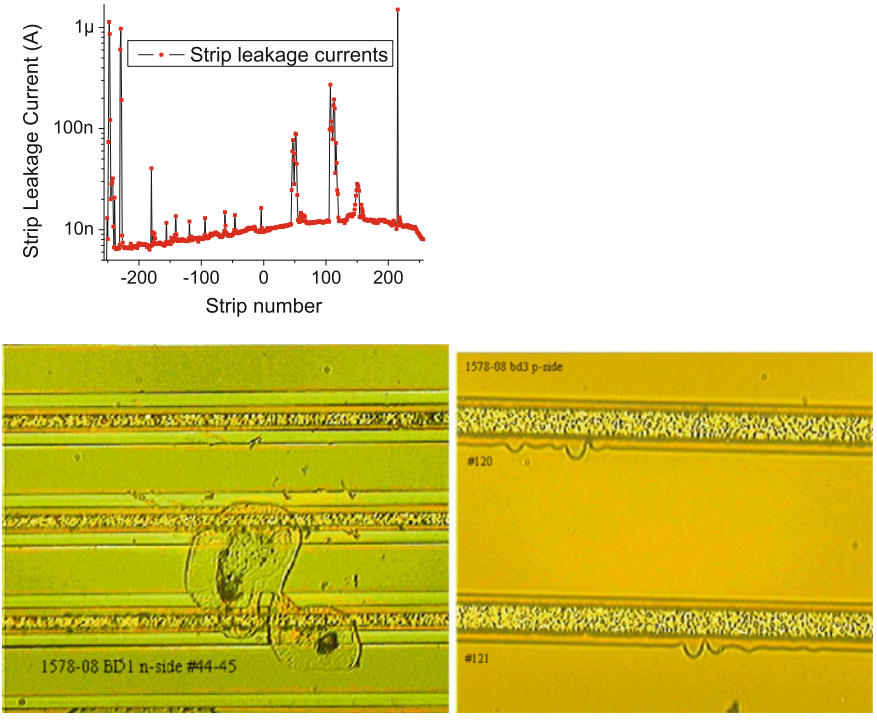


Fig. 1.89 A single-strip fault increasing strip leakage current by 2 orders of magnitude. The small protrusion or un-regularity between the implants on the p^+-side distorted the field strong enough to increase the leakage current locally. These defects often come in clusters

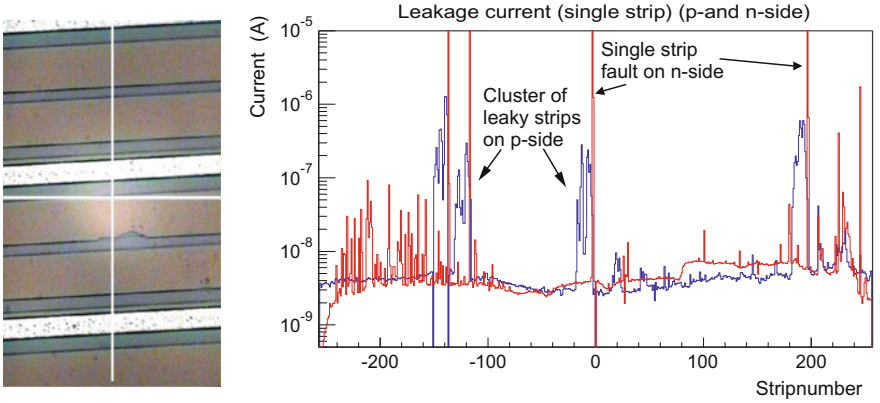


Fig. 1.90 A single-strip fault increasing strip leakage current by 2 orders of magnitude. The photo on the *left* shows a single-strip fault on the n-side where the p^+-stop is defective. The corresponding strip currents on the n- and p-side are displayed in the *right* plot. The single defect affects all geometrical corresponding strips on the p-side geometrically covered by the stereo layout

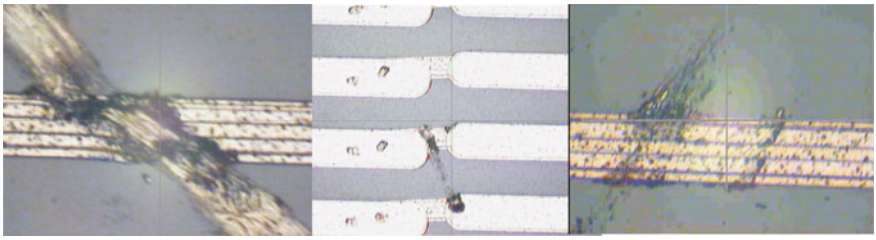


Fig. 1.91 Scratches resulting in high leakage currents. Most prominent is the scratch in the middle picture, it was present on about 50 sensors in a row, until the producer was notified. A pick-up tool was damaging the sensor with a placement precision of about 2 – 3 μm

Another prominent and severe leakage current problem was identified on the CDF II ISL double-sided sensors where one defective strip on the n-side affected a strip cluster on the p-side. A defect, a "mousebite", in the p^+-stops disturbed the field. The number of leaky strips on the p-side correlates to the geometric coverage of p strips to one n-strip, with the 1.2° stereo angle configuration. Figure 1.90 shows an example of a "mousebite" and the corresponding strip currents of the p- and n-side.

A prominent CMS case was superficial scratches only affecting the aluminium layer. Figure 1.91 shows examples of these scratches. Scratches were introduced during assembly but also prominently at the producer. On several sensors, the exact same scratch pattern on exactly the same strip number for many sensors was found where a pick-up tool impressed its mark [130].

Fig. 1.92 Some pinholes with optical counterparts

Pinhole Development

With the use of charge amplifiers, direct currents are prone to saturate the preamplifiers. Pinholes, low ohmic connections from p^+ implant to aluminium readout strip must be avoided (see Sect. 1.6.3).

During the prototype phase of the SVX II sensors, pinholes developed in the SiO_2. In the double-sided sensors there is a net electric field across the dielectric when bias voltage is applied. In addition wire-bonding weakened the isolation with respect to breakdown voltage. With this feedback the company was able to resolve the issue by applying a mix of "wet" and "dry" oxidation [43].

During the sensor prototyping of the CMS sensors, coupling capacitors were also weak and the number of pinholes per sensor was well above specification. Introducing an additional silicon nitride layer to the existing silicon oxide layer solved the situation.

Pinholes are sometimes optically identifiable, see Fig. 1.92. Pinholes are sometimes "created" in final steps during shipment and also during module assembly by careless handling, thus producing deep scratches.

Corrosion

During the CMS quality and process control, including long-term testing of the silicon sensors, strange stains on the guard and bias rings have been observed. The result of a systematic investigation proved that the detectors are affected by Al_2O_3 corrosion after some time under voltage in a humid environment. Depth profiling showed that the metal structures are compromised down to the level of the SiO_2. Finally, significant concentrations of potassium were found at precisely the location of the corrosion, serving as a catalyst for the electrochemical reaction.

Some examples of these dark-hued shades are displayed in Fig. 1.93. A full description of the incident can be found in [139]. Systematically the parameter space of time, voltage, humidity and temperature was investigated at passivated and non-passivated areas. Test duration went up to 2,000 h. Conclusive results on the necessity of voltage and humidity to drive the stain developing process were obtained. Testing at $T = -10°C$ still showed a slow stain development. An electrochemical process fuelled by water, heated by electricity and catalysed by some impurities corroded the aluminium structures.

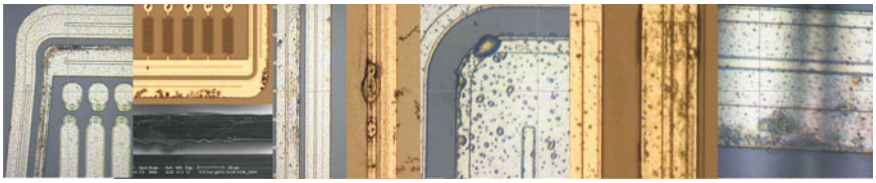


Fig. 1.93 The pictures give an optical impression of the severity of the corrosion [139]

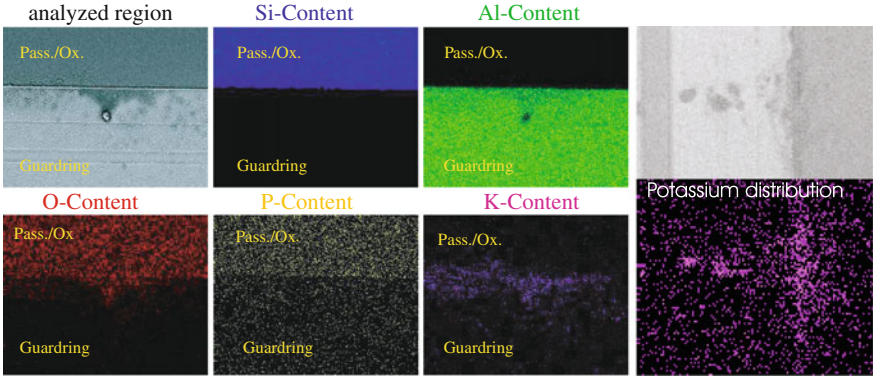


Fig. 1.94 Lateral element distribution. Comparing the optical picture with the material analysis results the missing aluminium and the abundance of oxygen catches the eye. Some spurious potassium conglomeration triggers a closer look (see *left figure*). Zooming in another region, a significant matching of potassium remnants and corroded regions is identified [139]

Investigating the origin of the stains, samples were submitted to an **E**nvironmental **S**canning **E**lectron **M**icroscope ESEM. Structural damage of the region under investigation was visible. Secondary electron emission reveals the element distribution. An atomic weight distribution of 24.31% O, 72.07% Al, 1.67% Si, 0.2% P, 0.45% S, 0.8% K was identified. By varying the primary electron energy in the ESEM analysis one varies the penetration depth. In some places the same element composition was found independent of the depth. Scanning a large lateral area and recording the element content, the missmatch of Al:Si:O in the affected regions (see Fig. 1.94) becomes obvious. Examining deeply the low content impurities, such as P, S and K, a significant matching of potassium concentration in the region was identified (see zoom on the right side of Fig. 1.94). Maybe leftover potassium from a SiO_2 etching step ended up causing some undesired side effects, namely aluminium corrosion. Since potassium hydroxide is a common agent in ASIC production [210], an improper removal of acids is assumed to be the cause of the problem.

Chapter 2
Radiation Damage in Silicon Detector Devices

The intent of this chapter is to introduce the radiation effects and give a general understanding of radiation damage – its mechanism, microscopic and macroscopic effects. The very basics of radiation damage are presented in [219] (Description of the so-called "Hamburg Model") and [356, 357], recent studies on fully segmented sensors on a large sample can be found in [79] and [80]. The three main effects (bulk and surface defects) introduced by radiation are

- displacement of atoms from their positions in the lattice (bulk)
- transient and long-term ionisation in insulator layers (surface)
- formation of interface defects (surface)

2.1 Bulk Damage

Detectors at the LHC and also already at the TEVATRON operate at high particle fluxes, necessary to achieve a large statistical sample on particle collisions, to understand physics of the events in a hadron collider environment. To understand the depletion voltage, leakage current or trapping of an irradiated sensor the following mechanisms have to be taken into account:

1. the damage to the lattice created by traversing particles
2. the following diffusion[1] processes – annealing

The following sections describe the microscopic and macroscopic changes in the material with radiation and subsequent changes with time. The introduced models

[1] The term "diffusion" used here is more a descriptive one combining effects like diffusion, migration, break-up, re-configuration of defects or better reactions between defects propagating through lattice – also often summarized by the term "annealing".

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_2

describe the evolution of sensor parameters like leakage current, depletion voltage and **C**harge **C**ollection **E**fficiency CCE. A significant dependency on the type of radiation particle exists. In addition, for charged particles, a strong dependence is seen for different irradiated silicon materials – especially on their oxygen or carbon content. The models, with the initial constants, describe the data very well. But for a large production, it is strongly advised to launch a dedicated radiation campaign allowing subsequently the re-fit of Hamburg model parameters for the specific sensors.

2.1.1 Damage by Particles

Traversing particles are not only ionising the lattice but they also interact with the atomic bodies via the electromagnetic and strong forces. Atoms are displaced and create interstitials I, vacancies V and more complex constructs, e.g. di-vacancies V_2 or even triple-vacancies V_3, also di-interstitials I_2 are common. All these defects deform the lattice. Some examples are depicted in Fig. 2.1. In addition diffusing Si atoms (interstitials I) or vacancies often form combinations with impurity atoms, like oxygen, phosphorus or carbon, again with different properties. All these lattice displacements populate new levels in the band gap, changing the initial silicon properties.

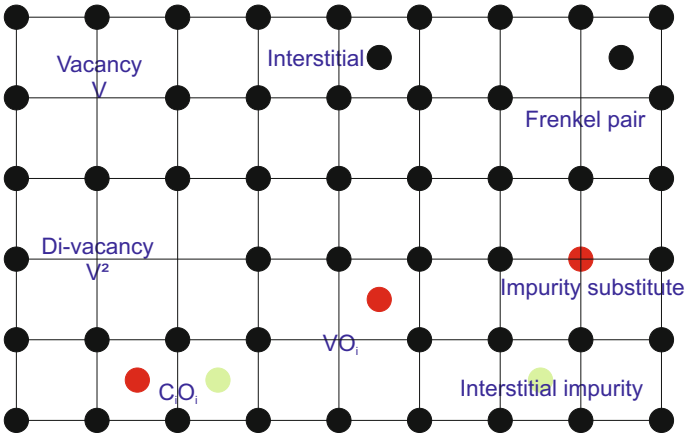


Fig. 2.1 The figure shows an exemplary selection of atomic displacements in the lattice after collision with traversing particles. These vacancies, interstitials and complex clusters are creating new levels in the energy scheme of the semiconductor and therefore change the elementary properties. As abbreviation, vacancies are labeled V, interstitials I, di-vacancies V_2. Impurities are labeled with their atomic sign, their index defines their position as substitute or interstitial, e.g. C_s or C_i

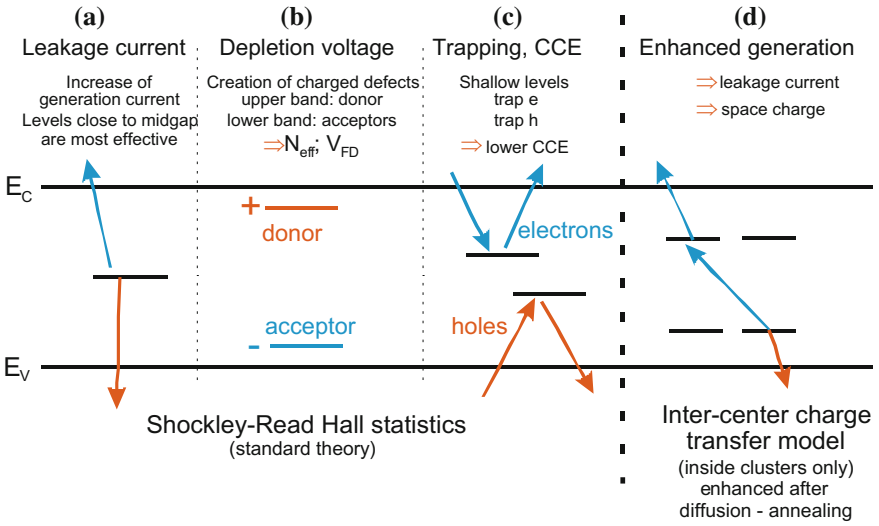


Fig. 2.2 The different defect level locations and their effects. All relevant defect levels due to radiation are located in the forbidden energy gap. **(a)** Mid-gap levels are mainly responsible for dark current generation, according to the Shockley–Read–Hall statistics and decreasing the charge carrier lifetime of the material. **(b)** Donors in the upper half of the band gap and acceptors in the lower half can contribute to the effective space charge. **(c)** Deep levels, with de-trapping times larger than the detector electronics peaking time, are detrimental. Charge is "lost", the signal decreases and the charge collection efficiency is degraded. Defects can trap electrons or holes. **(d)** The theory of inter-centre charge transfer model says that combinations of the different defects in so-called defect clusters additionally enhance the effects

The resulting macroscopic property changes are

- increase of dark current
- change of depletion voltage level (N_{eff}) due to creation of mainly additional acceptor levels
- decrease of charge collection efficiency due to defect creation, acting as traps

The basic important levels and their roles and macroscopic effects are shown in Fig. 2.2. Defect analysis and even defect engineering was started in the R&D collaboration Rose/RD48[2] and is nowadays continued within the RD50 collaboration at CERN. Within this chapter the basic mechanisms are described, special cases and defect engineering will be described in Sects. 6.4.2 and 7.1.1. The topic of defect engineering is fully covered and up to date in [336]. For example vacancy plus phosphorus *VP* removes the donor property of single phosphorus. A special case is the di-vacancy plus oxygen combination V_2O, which introduces additional negative space charge. A detailed table of defect combinations and respective quantitative energy levels can be found in [194, 219]. The defect distribution and clustering in

[2]ROSE: **R**&**D** **O**n **S**ilicon for future **E**xperiments.

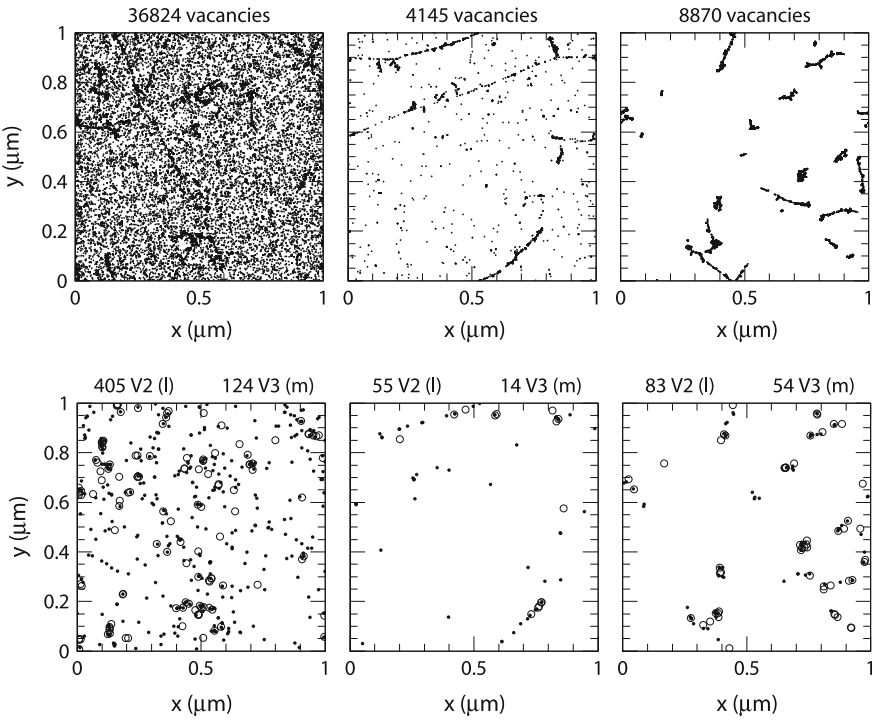


Fig. 2.3 Simulation of defect formation with radiation and diffusion. The upper three simulations show the microscopic picture of defect distribution. About 10 MeV protons (*left*) produce a quite homogeneous vacancy distribution, while more energetic protons with 24 GeV (*middle*) form more clustered and discrete defects. Neutrons with 1 MeV (*right*), interacting only due to strong interaction, do produce more isolated clustered defects. The plots are projections over 1 µm of depth (z) and correspond to a fluence of 10^{14} $n_{1\mathrm{MeV}}/\mathrm{cm}^2$. The *lower three* figures are displaying final constellations after a certain annealing time and therefore diffusion effects occurred. Many initial defects decay, e.g. Frenkel pairs, where interstitials recombine with vacancies. Others form more local clusters, like formations of di- and triple vacancies, with again different levels and therefore different properties. The full study is described in [150]

Fig. 2.3 clearly show the difference between irradiation with charged and neutral particles and different energy levels that emerge from them. Especially the additional Coulomb force of a charged particle enhances small energy transfer and therefore local short distance defects, known as Frenkel pairs. Neutral particles, like neutrons, acting via the strong force result mainly in long-range cluster defects. With enough energy the initial **P**rimary **K**nock on **A**tom PKA acts further on several additional lattice atoms. In the non-relativistic approach, the maximum transferred energy $E_{R,max}$ can be calculated in the case of recoil for a particle with mass m_p and kinetic energy E_p with

$$E_{R,max} = 4E_p \frac{m_p m_{Si}}{(m_p + m_{Si})^2} \tag{2.1}$$

For electrons with energy E_e and mass m_e, higher kinetic energies are needed for lattice damage. $E_{R,max}$ can be calculated by the approximate relativistic relation

$$E_{R,max} = 2E_e \frac{E_e + 2m_e c^2}{m_{Si} c^2} \tag{2.2}$$

Table 2.1 shows the average and maximum energy transfer of different particle types with the incident energy of 1 MeV.

Also, the energies needed to knock an atom from its original lattice place are clearly dependent on the binding forces and therefore on the material. In the case of silicon, the minimum energy needed to displace a single lattice atom (Frenkel pair) is $E_{\text{single defect}} \approx 25$ eV, while $E_{\text{cluster}} \approx 5$ keV [343] is needed to produce a defect cluster. Table 2.2 lists the minimum kinetic particle energies needed to transfer $E_{\text{single defect}}$ or E_{cluster}, subsequently creating a single point defect or a cluster.

The **N**on **I**onising **E**nergy **L**oss NIEL hypothesis allows a first-order normalization of radiation damage with respect to different particles with different energies, see also Table 2.2. According to NIEL, the damage manifestation depends only on energy transferred in collisions regardless of particle energy and type. We will see later that this is not always true.

The displacement damage $D(E)$ can be calculated by

$$D(E) = \sum_i \sigma_i(E_{kin}) \int_0^{E_{R,max}} f_i(E_{kin},\ E_R) P(E_R) dE_R \tag{2.3}$$

where all possible interactions are summed up. σ_i is the cross-section of the process and $f_i(E, T)$ is the probability of having a collision of a particle with E_{kin}, transferring a recoil energy of E_R. $P(E_R)$ is the Lindhard partition function [190], describing

Table 2.1 Maximum energy transfer ($E_{Recoil,max}$) versus average ($E_{Recoil,av}$) of different particle types with incident energy of 1 MeV. Clearly the long-ranging Coulomb force favours a small energy transfer and therefore point defects, while the strong force produces mostly clusters

	Electron	Proton	Neutron	Si^+
Force	Coulomb	Coulomb and nuclear	Elastic nuclear	Coulomb
$E_{R,max}$[keV]/$E_{R,av}$[keV]	0.155/0.046	133.7/0.21	133.9/50	1000/0.265

Table 2.2 The minimum kinetic particle energies to create single point or cluster defects can be derived with formula 1.45 (p and n) and 1.46 (e^-). The ^{60}Co-photons could only create a cluster defect via a secondary electron (dominantly via Compton effect), but it cannot energize e^- up to 8 MeV

Particle	$E_{\text{single defect}}$	E_{cluster}
neutron; protons	185 eV	35 keV
electrons	225 keV	8 MeV
^{60}Co-gammas	1 MeV	No cluster

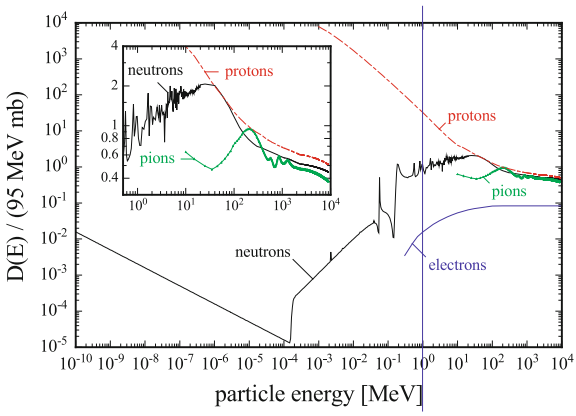


Fig. 2.4 NIEL – non ionising energy loss. Displacement damage function D(E) normalized to 95 MeVmb for neutrons, protons, pions and electrons. The inserted graph shows a detail of the figure around relevant energies for high energy physics [150, 219]

the fraction of energy going into silicon atom displacement, e.g. $P(E_R) \approx 50\%$ for 10 MeV protons or $P(E_R) \approx 42\%$ for 24 GeV protons and $P(E_R) \approx 43\%$ for 1 MeV neutrons [150]. The resulting displacement functions are plotted in Fig. 2.4. As standard the 1 MeV *neutron equivalent fluence* $[n_{1\mathrm{MeV}}/\mathrm{cm}^2]$ is used and abbreviated with Φ_{eq}, corresponding to $D_{neutron}$ $(1\,\mathrm{MeV})/\mathrm{cm}^2 = 95\,\mathrm{MeV\ mb/cm}^2$, with millibarn: $\mathrm{mb} = 10^{-27}\mathrm{cm}^2$. It is therefore possible to scale radiation damage from different particles and different energies by a simple numerical factor κ to make comparison possible. κ is defined as the ratio of the individual damage particle factor for a given energy and 1 MeV neutrons.

$$\kappa = \frac{\int D(E)\phi(E)dE}{95MeV\,mb \cdot \Phi} = \frac{\Phi_{eq}}{\Phi} \tag{2.4}$$

where $\phi = \int \phi(E)dE$ is the irradiation fluence. The 1 MeV neutron equivalent fluence Φ_{eq} is then calculated by

$$\Phi_{eq} = \kappa\Phi = \kappa \int \phi(E)dE; \quad [\Phi_{eq}] = n_{1\ \mathrm{MeV}}/\mathrm{cm}^2 \tag{2.5}$$

In plots and in literature the $n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ unit is often omitted and without specific reference to a particle type all fluences are given as 1 MeV neutron equivalent.

Leakage Currents

Mid-gap defects are constantly produced during radiation. They degrade mainly the lifetime τ, being efficient electron–hole pair generators they therefore increase the dark currents. It was found in many experiments that there is a linear behaviour of dark current versus fluence. The shot noise increases with $ENC_{IL}\sqrt{I}$ respectively; see also Sect. 1.5.

$$\frac{\Delta I}{V} = \alpha\Phi_{eq} \tag{2.6}$$

where V normalizes for a given volume. α is called the current-related damage rate. The correspondence is shown in the left part of Fig. 2.5. The good linearity over several orders of magnitude allows the technical use of diodes to determine the particle fluence by the increase of current.

Depletion Voltage

The situation for the effective space charge concentration is a bit more difficult. It is displayed in Fig. 2.6. Starting with an n-type-doped silicon bulk, a constant removal of donors ($P + V \rightarrow VP$ centre) together with an increase of acceptor-like levels (one example is $V + V + O \rightarrow V_2O$) shifts the space charge first down to an intrinsic level and then up to a more p-like substance. The material "type inverts". When the material changes from n to p, the space charge changes sign this is often referred to as **S**pace **C**harge **S**ign **I**nversion – SCSI. The depletion voltage therefore drops first and starts rising later. Figure 6.28 in Sect. 6.4.2 shows an example of the CMS sensor irradiation campaign and the evolution of depletion voltage.

$$N_{eff} = N_{D,0}e^{-c_D\Phi_{eq}} - N_{A,0}e^{-c_A\Phi_{eq}} - b\Phi_{eq} \tag{2.7}$$

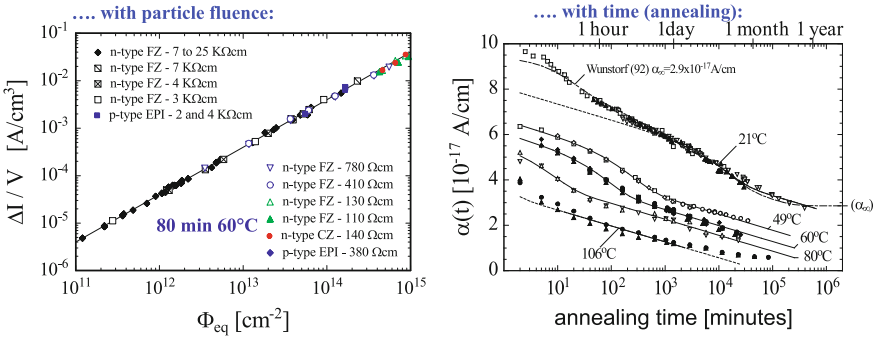


Fig. 2.5 Leakage current versus fluence and annealing time [219, 356]

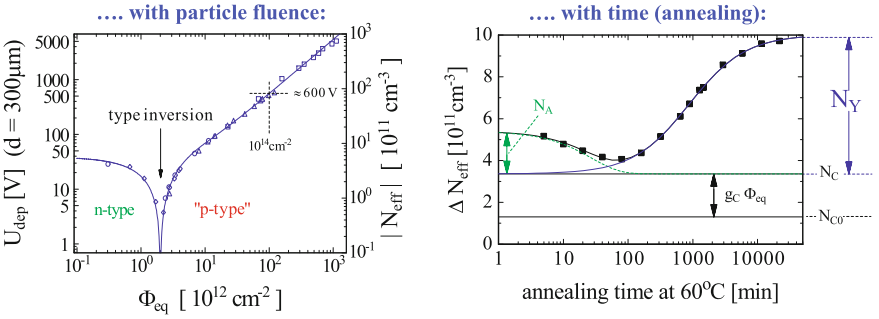


Fig. 2.6 Depletion voltage versus fluence and annealing time [219]

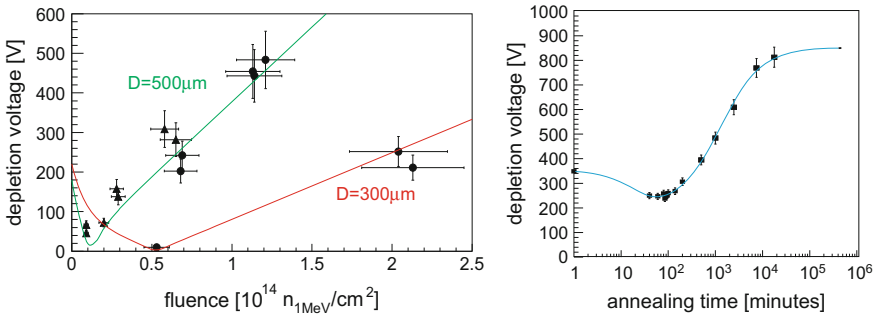


Fig. 2.7 Evolution of V_{FD} for different fluences and annealing durations. To have a basis for radiation evaluation, CMS irradiated several sensors and modules to get actual adapted fit parameters to the Hamburg model for the specific procured sensors. In this case, the beneficial constants g_a were found to be $(1.11 \pm 0.16) \cdot 10^{-2}\ \mathrm{cm}^{-1}$ and $t_a(60°\mathrm{C}) = 21 \pm 8$ min; the reverse constants are $\mathrm{g}_Y = 4.91 \pm 0.27 \cdot 10^{-2}\ \mathrm{cm}^{-1}$ and $\mathrm{t_y}(60°\mathrm{C}) = 1290 \pm 262$ min. The different $\mathrm{V_{FD}}$ curve behaviours in the *left plot* can be explained by the different sensor thicknesses of D =500 µm (*upper curve*) and 320 µm (*lower curve*) – mind $\mathrm{V_{FD}} \sim \mathrm{N_{eff}} \cdot \mathrm{D}^2$. At fluences of $\Phi_{eq} = 10^{14}\ n_{1\mathrm{MeV}}/\mathrm{cm}^2$, $\mathrm{V_{FD}}$ of the thick sensor would have increased above 1000 V. The initial compatible $\mathrm{V_{FD}}$ values are due to the different sensor resistivities. Data are compared to calculations for an annealing time of 80 min and an annealing temperature of 60 °C at each fluence step. More about this study is described in Sect. 6.4.2 and [351]

With the evolution of N_{eff} can be parameterized in first approximation with the donor and acceptor removal rates c_D and c_A plus the most important acceptor creation term $b\Phi_{eq}$. Since there is a significant temperature-dependent diffusion, Formula (2.13) parameterizes the evolution in a more common and general description.

Charge Trapping

The trapping rate is proportional to the concentration of trapping centres N_i, resulting from defects. Therefore the trapping probability can be formulated by

$$\frac{1}{\tau_{eff}} = \sum_i N_i (1 - P_i) \sigma_i v_{th} \tag{2.8}$$

with P_i the occupation probability and σ_i the charge carrier cross-section. In first order the fluence dependence is linear and can be written as

$$N_i = g_i \Phi_{eq} f_i(t) \Rightarrow \frac{1}{\tau_{eff}} = \gamma \Phi_{eq} \tag{2.9}$$

with the introduction rate g_i; $f_i(t)$ describes the annealing with time. An example is plotted in Fig. 2.9 (left side). The slope is different for electron and hole trapping, they are differently affected due to their different mobilities. Some initial numbers for proton and neutron irradiations are given in Table 2.3. The degradation of **C**harge **C**ollection **E**fficiency CCE can then be described by

Table 2.3 Introduction of trapping centres, significant for electrons and holes of neutron and proton radiations [184]

	$\gamma_{e,0}$ [10^{-16} cm^2/ns] for Electron	$\gamma_{h,0}$ [10^{-16} cm^2/ns] for Holes
Fast charged hadrons	(4.97 ± 0.14)	(5.25 ± 0.17)
Neutron	(3.53 ± 0.24)	(5.10 ± 0.39)

$$Q_{e,h}(t) = Q_{0_{e,h}} e^{-\frac{t}{\tau_{eff_{e,h}}}}, \text{ where } \frac{1}{\tau_{eff_{e,h}}} \propto N_{\text{defects}} \tag{2.10}$$

The above assumption is valid as long as the drift velocity v_D is much smaller than thermal velocity vth and trapping distance $\lambda = v_{th} \cdot tau_{eff}$. At effective fluences of $\Phi_{eq} = 10^{15}\ n_{1\text{ MeV}}/\text{cm}^2$ and above, trapping becomes the most limiting factor of silicon usage as a particle detector. The charges no longer arrive at the collecting electrodes in 300 µm thick sensors. Examples of charge travelling distances x for $\Phi_{eq} = 10^{15}\ n_{1\text{MeV}}/\text{cm}^2$ and $\Phi_{eq} = 10^{16}\ n_{1\text{MeV}}/\text{cm}^2$ are

- $\tau_{eff}(10^{15}\ n_{1\,\text{MeV}}/\text{cm}^2) = 2\,\text{ns} : x = v_D \cdot \tau_{eff} = (10^7\text{cm/s}) \cdot 2\,\text{ns} = 200\,\mu\text{m}$
- $\tau_{eff}(10^{16}\ n_{1\,\text{MeV}}/\text{cm}^2) = 0.2\,\text{ns} : x = (10^7\text{cm/s}) \cdot 0.2\,\text{ns} = 20\,\mu\text{m}$

The following list summarizes the main operation limiting effects of radiation bulk damages for the different Φ_{eq} levels:

- at $10^{14}\ n_{1\,\text{MeV}}/\text{cm}^2$ the main problem is the increase of leakage current
- at $10^{15}\ n_{1\,\text{MeV}}/\text{cm}^2$ the high resulting depletion voltage is problematic
- at $10^{16}\ n_{1\,\text{MeV}}/\text{cm}^2$ the fundamental problem is the CCE degradation.

2.1.2 Annealing – Diffusion of Defects

Interstitials and vacancies are very mobile at temperatures $T > 150\,\text{K}$. The lower part of Fig. 2.3 displays the result of defect diffusion. Basically, there is the possibility of

- Frenkel pair recombination ($I + V \rightarrow Si$)
- multi-vacancy and multi-interstitial combination (e.g. $V + V \rightarrow V_2$)
- combination of more complex defects (e.g. $C_i + O_i \rightarrow C_i O_i$ or $V + P \rightarrow VP$)

where the former types are short-range and very mobile processes and therefore happen with a shorter time constant, while the latter happens with a longer time constant. The whole process is called *annealing* with a beneficial part reducing the damage and a reverse annealing part degrading macroscopic sensor properties. Some parts are stable and do not evolve with time. The diffusion processes are naturally temperature dependent and some effects, e.g. depletion voltage evolution, can even be effectively frozen out at temperatures below 0 °C. In addition, different levels in the energy band behave differently with respect to time constants and temperatures.

Table 2.4 α and τ_I for different temperatures

Annealing temperature in °C	21	40	60	80
α_0 in 10^{-17} A/cm	7	6	5	4
τ_I in min	140000	260	94	9

The table shows the α parameter and the time constants τ_I for the current annealing for different temperatures. Below room temperature the time constants are longer than 100 days and annealing is almost frozen out

Annealing – Leakage Current

The annealing of the dark currents is displayed in the right part of Fig. 2.5.

The α parameter, respectively the dark currents, can be parameterized. In first order α can be fitted by a sum of exponentials, pointing to the existence of several contributing defects with different decay time constants. The radiation afflicted current continuously decays exponentially until it follows more or less a logarithmic behaviour or even saturates for higher temperatures after several months. According to [219] the annealing behaviour can be described by

$$\alpha = \alpha_0 + \alpha_I e^{-\frac{t}{\tau_I}} - \beta \cdot ln\frac{t}{t_0} \tag{2.11}$$

with $\alpha_I \sim 1.25 \cdot 10^{-17}$ A/cm, $\beta \sim 3 \cdot 10^{-18}$ A/cm and $t_0 = 1$ min. τ_I takes the annealing temperature T_α dependence into account, where

$$\frac{1}{\tau_I} = k_{0I} \cdot e^{\frac{E_g}{k_B T_\alpha}} \tag{2.12}$$

with $k_{0I} = 1.2^{+5.3}_{-1.0} \cdot 10^{13}$ s^{-1}, e.g. $\tau_I \approx 10$ days at room temperature. $\alpha_0 = -(8.9 \pm 1.3) \cdot 10^{-17}$ A/cm $+ (4.6 \pm 0.4) \cdot 10^{-14}$ AK/cm $\cdot \frac{1}{T_\alpha}$ is a fitted parameter dependent on the annealing temperature. Table 2.4 gives α and T_I examples for different temperatures.

The average α after a standard annealing scenario of 80 min at 60°C is $4 \cdot 10^{-17}$ A/cm, measured at T = 20°C. The α value changes about 15% every 1 degree. The effect is temperature dependent and also effective but strongly reduced at sub-zero temperatures. Different from the time dependence of the depletion voltage, which starts to rise at later times, the current annealing always decreases the dark current and is therefore only "beneficial".

Annealing – Depletion Voltage

The annealing of N_{eff} and therefore the depletion voltage evolution can be described by

$$\Delta N_{eff}(\Phi_{eq}, t, T) = N_{C,0}(\Phi_{eq}) + N_A(\Phi_{eq}, t, T) + N_Y(\Phi_{eq}, t, T) \tag{2.13}$$

where Φ_{eq} stands for 1 MeV neutron equivalent fluence, with the stable term $N_{C,0}$, the short-term annealing term N_A and the second-order long term N_Y. This description is called the Hamburg model and it is depicted in the right part of Fig. 2.6. In its basic nature it parameterizes the space charge change due to donor removal plus acceptor creation with fluence and latter annealing. It was proposed in [219] and with some adaptation of the initial time constants to the different sensors it has been very successful.

Before discussing the three annealing terms in some detail, it must be mentioned that the stable term is the most relevant one in a high-radiation environment. For the LHC experiments, the beneficial annealing needs to be exploited during the maintenance periods to heal the short-range defects and therefore reduce the depletion voltage, while the reverse annealing term needs to be suppressed by freezing out at sub-zero temperatures, both are possible. It also has to be mentioned that all constants in this chapter are taken from [219] and need to be re-fitted for the user case, see e.g. [350], some constants reflect, for example, the oxygen content. As an example, Fig. 2.7 illustrates the CMS radiation campaign, where sensors were subjected to several fluences and annealing parameters were fitted.

The Stable Term $N_{C,0}$

In the current understanding, the stable damage term consists of two components, the donor removal and the acceptor creation rate.

$$N_C(\Phi_{eq}) = N_{C,O}(1 - e^{-c\Phi_{eq}}) + g_c\Phi_{eq} \tag{2.14}$$

where c is the initial dopant (donor in n- and acceptor in p-bulk) removal constant and $N_{C,O}/N_{eff,non\text{-}irradiated}$ is the fraction of initial dopant removal, depending strongly on the oxygen concentration, where oxygen can bind vacancies, which would otherwise combine with phosphorus ($V + O \rightarrow VO$ instead of $V + P \rightarrow VP$). This is of course grandly simplified.

Michael Moll – spokesperson of the RD50 collaboration responds to this subject:

> For p-in-n sensors, the donor removal component of the Hamburg model can not be described by a simple process $V + P \rightarrow VP$ only. There is something more behind that and we still do not exactly understand what it is.

Ranges from 10 to 80% donor removal are observed after neutron irradiation in p-in-n sensor. On average the initial donor removal rate here is $N_{C,0} \cdot c = (7.5 \pm 0.6) \cdot 10^{-2}\ \mathrm{cm}^{-1}$. The second term describes a creation rate of stable acceptors with an average measured $g_c = (1.49 \pm 0.04) \cdot 10^{-2}\ \mathrm{cm}^{-1}$. Neither terms are time dependent as there are no evolutionary diffusion processes.

The Short-Term Annealing N_A, the Beneficial One

Monitoring the sensors directly after irradiation, a fast change of the depletion voltage can be observed. For type-inverted material V_{FD} decreases with time, while it increases before type inversion. This can easily be interpreted as an increase of the effective doping concentration N_{eff}. The introduced acceptors decay and inverted

sensors with negative space charges become less negative while sensors that are not type inverted with positive space charge become more positive.[3] The decay of defects can be factorized in a series of first-order exponential decays. In extremely long duration HEP experiments, the short time constants of minutes and hours are not relevant and the series can be reduced to

$$N_A(\Phi_{eq}, t) = \Phi_{eq} g_a e^{-\frac{t}{\tau_a}}; \quad N_A = g_a \Phi_{eq} \tag{2.15}$$

The average value of $g_a = (1.81 \pm 0.14) \cdot 10^{-2}$ cm^{-1} was determined. The time constants τ_a, describing diffusion processes, are naturally temperature dependent and can be parameterized by

$$\frac{1}{\tau_a} = k_a = k_{0a} \cdot e^{-\frac{E_{aa}}{k_B T_a}} \text{ with } E_{aa} = (1.09 \pm 0.03) \text{ eV and } k_{0a} = 2.4^{+1.2}_{-0.8} \cdot 10^{13} \text{ s}^{-1} \tag{2.16}$$

They are summarized in Table 2.5 together with the time constants τ_y describing the reverse annealing term.

The Long-Term Annealing N_Y, the Reverse One

After a long time another first-order effect,[4] with another time constant τ_y, becomes effective – reverse annealing. Space charge becomes even more negative, more acceptor levels form. Reverse annealing can be parameterized by

$$N_Y = N_{Y,0} \cdot (1 - e^{-t/\tau_Y}) \tag{2.17}$$

with

$$N_{Y,0} = g_Y \cdot \Phi_{eq} \tag{2.18}$$

being directly proportional to the particle fluence, with the introduction rate g_Y depending on the radiation type and radiated sensor material, e.g. neutron radiation on standard n-type FZ has $g_{Y,neutron} = 4.8 \cdot 10^{-2}$ cm^{-1}. Time constants are given in Table 2.5. The diffusion is temperature dependent and can be described after [219] by

$$\frac{1}{\tau_Y} = k_Y = k_{0,Y} \cdot e^{-\frac{E_{YY}}{k_b T_a}} \text{ with } k_{0,Y} = 7.4 \cdot 10^{14} \text{ s}^{-1} \text{ and } E_{YY} = 1.31 \text{ eV} \tag{2.19}$$

[3] Reminder: acceptors are introducing negative, donors respectively positive, space charge.

[4] Reverse annealing is also often described in literature (e.g. [107, 266, 267]) as a second-order effect with a parameterization of $\left[1 - \frac{1}{1+t/\tau_y}\right]$, describing accurately N_{eff} versus time for long annealing times at higher temperatures. Nevertheless, the physical mechanism is ruled out due to the missing dependency of the effect on fluence. The rate, depending on the probability of **two** defects combining, does not increase with the number of defects.

Table 2.5 Annealing time constants; beneficial and reverse annealing [219]

Annealing temperature (°C)	−10	0	10	20	40	60	80
Short-term annealing τ_a	306 d	53 d	10 d	55 h	4 h	19 min	2 min
Reverse annealing τ_Y	516 y	61 y	8 y	475 d	17 d	21 h	92 min

The numbers define immediately the running and maintenance conditions of experiments in a high-radiation environment. Operation temperatures below 0 °C freeze out the reverse term completely and largely the beneficial one. During the maintenance periods the detectors temperatures should be elevated in a controlled way to benefit from the short-term annealing and to absolutely avoid the reverse one. An example of the foreseen maintenance scenarios of CMS are described in Sect. 6.4.2

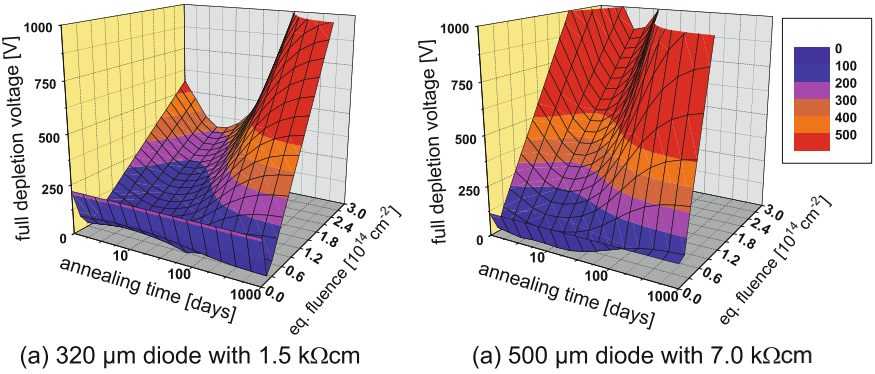


Fig. 2.8 Depletion voltage versus fluence and annealing time at room temperature. The plot shows the simulated depletion voltages according to the Hamburg model for 300 and 500 µm thick silicon, the two CMS sensor configurations. The ordinate shows the depletion voltage versus fluence and temperature. For detector operation in high-radiation environment, it is clearly necessary to make use of the annealing up to the point where the reverse annealing becomes too strong. Of course there is one plot to be drawn per temperature, which defines the annealing time constants [80]

The full depletion voltage evolution with respect to fluence and time for a fixed temperature is presented in a condensed form in Fig. 2.8.

Annealing of Effective Trapping Probability

Formula (2.9) [175] already introduced the term $f_i(t)$ and the proportional factor γ. Experimental data show a decreasing probability of electron trapping and an increase of hole trapping in time. The right part of Fig. 2.9 shows the annealing of the effective inverse trapping times. The interesting operational parameter **C**harge **C**ollection **E**fficiency CCE is then basically proportional to the electric field and trapping.

$$\gamma(t) = \frac{1}{\tau_{eff}} = \frac{1}{\tau_0} \cdot e^{-\frac{t}{\tau_a}} + \frac{t}{\tau_\infty}(1 - e^{\frac{t}{\tau_a}}) \tag{2.20}$$

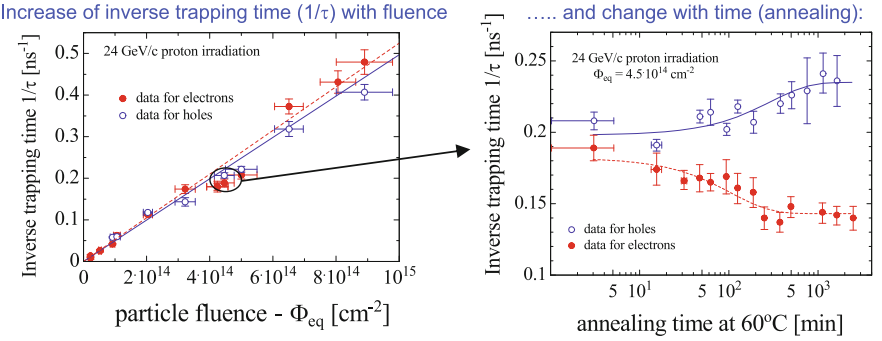


Fig. 2.9 Trapping time versus fluence and annealing time [184]

Still today, the statistical sample is not large, especially not on low-temperature samples. Saturation already starts after several hours at $T = 60\,^{\circ}\mathrm{C}$, when the annealing time constants τ_a for electrons and holes are on the order of hours at elevated temperatures. The effect counts for roughly 20% for electrons and 20 – 50% for holes. It can be neglected at sub-zero temperatures. Pixel detectors[5] at ATLAS and CMS even benefit from the fact that inverse trapping times decrease for electrons and therefore CCE increases with time.

To summarize, the macroscopic changes deriving from diffusion, called annealing, are highly temperature dependent, while frozen out at sub-zero temperatures they are dominant at room temperature and above. Leakage current, **C**harge **C**ollection **E**fficiency CCE and depletion voltages evolve with time in the following way:

- leakage current always decreases
- trapping probability decreases for holes and increases for electrons
- acceptor levels first decay in the beneficial phase and increase later in the reverse annealing phase. This leads to an increase/decrease of depletion voltage before and decrease/increase after type inversion.

The recipe is to benefit from the beneficial annealing for voltage and current and avoid the reverse annealing phase to stay in applicable bias voltage levels.

2.2 Defect Analysis, New Materials and Detector Engineering

It was mentioned, in the last sections, that impurities can influence the radiation hardness. Due to meticulous studies, mostly in the framework of RD48 [191] and RD50 [336], several influences are understood at the microscopic level or at least

[5] ATLAS and CMS pixels use n-in-n technology where most of the charge is induced by electrons, while for standard strip p-in-n sensors most of the charge is induced by holes.

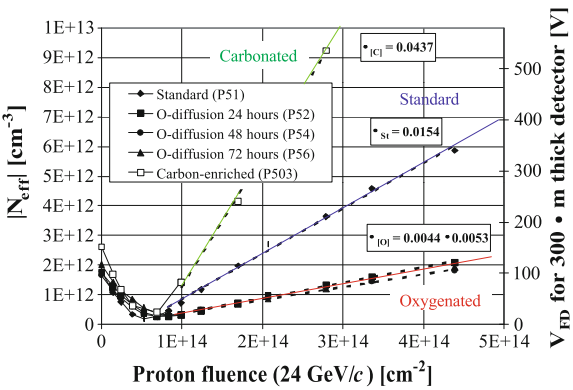


Fig. 2.10 Evolution of V_{FD} versus time of differently engineered silicon diodes. The beneficial influence of oxygen and malevolent effect of carbon are clearly visible. Today the ATLAS and CMS pixel sensors are composed of oxygenated silicon sensors [Courtesy of RD48 [191] and RD50]

empirically on the macroscopic level. A lot of different silicon sensor substrates with several types of diffused atoms were investigated, e.g. oxygenated, carbonated, *Li*-covered, etc. Substrates created with different growth techniques were investigated and irradiated to several fluence levels, e.g. FZ, CZ, magnetic[6] CZ, epitaxial material. To present all studies is beyond the scope of this book but a fair number of examples from RD50 will be described in the next paragraphs. The discussion of final sensor choices for the CMS Tracker Upgrade for the HL-LHC will be presented in Sect. 7.1 in detail.

Today, the most important beneficial effect identified was, that for oxygen-enriched material, the stable damage parameter g_c decreased and the reverse annealing time constant τ_Y increased. However, this effect is only valid for irradiation by charged particles.[7] The effect is already being technologically exploited. Today, the pixel sensors of the ATLAS and CMS experiment are oxygenated. The beneficial effect of high oxygen concentration and the degrading effect of carbon content are shown in Fig. 2.10.

The next paragraphs *introduce* the study of microscopic effects and their impact on macroscopic parameters and the fact that NIEL is no longer valid for N_{eff}, nor for effective trapping times, for different materials; depletion voltage becomes an academic concept, how sensors behave at very high fluences, and briefly discuss the effect of amplification in HEP sensor prototypes. In general, investigations are ongoing and the understanding of radiation damage mechanism and device engineering is growing every day.

[6] mCz: CZ crystal growth in a magnetic field to achieve a homogeneous oxygen distribution.

[7] Note that this violates the NIEL hypothesis.

2.2.1 Study of Microscopic Defects and Their Impact on Macroscopic Parameters

Meticulous studies about defects induced by radiation have been conducted in the RD50 and WODEAN [354] framework. The goal is to answer the following questions for microscopic defects:

- correlation between chemical constellation and energy level; e.g. VO-complex at $E_C - 0.176eV$, or E5 seems to be a tri-vacancy-complex [160]. Not much is known for more complex structures
- correlation between defects and macroscopic effects (operations/measurement) – (a) Leakage current? (b) Space charge (Donor? Acceptor? Neutral?)?, (c) Trapping?
- which are point or cluster defects?
- how do they evolve with time (annealing)? Differently said, which defects are responsible for reverse annealing?
- are they charged or neutral at operation temperature?
- which particles (p, π, n, γs and e radiation) at which energy induce which defects? NIEL?
- Defect Engineering: which materials (FZ, DOFZ, Cz, mCz, Epi) are affected by which radiation type and which defects develop?
  - What can we do to avoid certain malicious defects? E.g. add oxygen since the VO complex has no negative effect and is stable.

A combination of radiation and measurement, standard mainly DLTS and TSC (both briefly introduced in Sect. 1.8), is shedding light on the topic. The phase space has not yet been exhaustively explored but more and more pieces of the puzzle have been identified in the past 2 decades. Figure 2.11 tries to summarise the different energy levels of certain defects and their role as we understand them.

A comprehensive set of corresponding publications can be found at [153]. Reference [250] tabulates the defects with their energies, cross-sections and the most recent understanding of their effects. Selected, relevant examples will be presented in the following paragraphs. Some examples of defects and their potential effects:

The I_P(0/-) point-defect is being generated with γ radiation and contributes to the leakage current. H116K(0/-), H140K(0/-), H152K(0/-) are the main culprits for the reverse annealing of the depletion voltage (more later). The current indications mark E205a(-/0) and H152K(0/-) as important trapping centres. It seems that oxygen binds vacancies and also the bi-stable BD is oxygen dependent with a higher introduction rate with proton than neutron irradiation. The E4/E5 are believed to triple-vacancy-complexes.

Two prominent and pedagogic examples of the microscopic to macroscopic correlations will be presented for illustration; a tiny fraction of the many studies. The understanding is continuously growing.

One example is presented, where levels H116K, H140K, H152K, as acceptors, can be identified being responsible for reverse annealing. These levels do not form

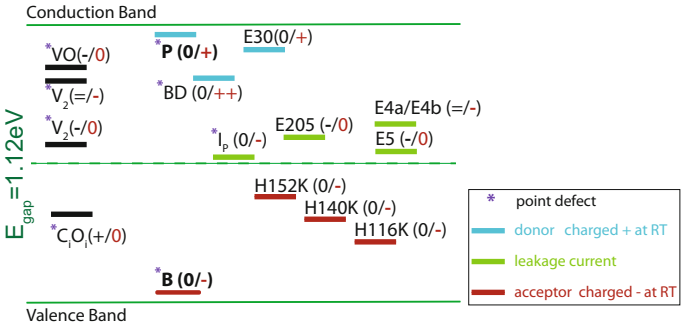


Fig. 2.11 A selection of radiation induced defects and their energy level in the band gap. To put everything in perspective, the levels of the common dopant atoms, phosphorus and boron, are given. The asterisk (*) indicates point defects, the others are more complex! The *upper right* indices gives the potential charge-states of the defects. The right value defines the space charge state of the defect in the SCR at room or operation temperature. Two examples: The free electron from the P atom (0/**+**) acts as majority charge carrier and the P is positively charged (*n*-doped). The VO-complex (-/**0**) is neutral – 0. "ZERO" means they are neutral; not active as dopants. Therefore the "*red*" defects (0/**-**) in the lower half act as acceptors and the upper "*blue*" defects (0/**+**) as donors. The left index value represents the defect state when a charge carrier has been trapped. Mind, for acceptors the free charge (carrier) occupying the trap is a hole and for donors it is an electron. The "*green*" defects, located near mid-band, contribute to the leakage current. The charge state transition from left to right is what we measure with DLTS (space charge change – capacitive change) or TSC (released charge – current). For example, the H152K(0/-) complex releases an electron measured as a current by TSC and space charge changes from neutral to positive, where DLTS measures the corresponding capacitive change

with γ radiation and are therefore cluster defects. The concentration of these levels increases with long-time annealing and are quantitatively compatible with negative space charge build-up (N_{eff} change). Figure 2.12(left) shows the Thermally Stimulated Currents Method TSC to determine the defect level concentrations while Fig. 2.12(right) shows the corresponding N_{eff} change. The goal of material defect engineering is now to either avoid the creation of these defects or to create counteracting donor effects with similar annealing behaviour.

E4a and E4b, bi-stable defects, have been discovered by [105]. A study, described in detail at [159], uses the bi-stability of the E4 effects to prove the strong correlation with the leakage current. This is visualised in Fig. 2.13.

Several diodes of different materials have been irradiated to moderate levels (regime where DLTS still works). Due to the bi-stability of the defect, the E4 concentration can be changed with charge injection. The DLTS spectra were recorded after three different steps: at first after annealing at 200 °C for 30 min, where both E4 levels annealed out completely (open squares), secondly after injection of 1 A/cm^2 forward current (full circles) which leads to a full recovery of both levels and finally after a subsequent annealing at 80 °C for 60 min (solid line), restoring the initial state of the spectrum. The leakage current clearly follows the E4 concentration. The very similar behaviour of E4a with respect to E4b suggests that both defect states are

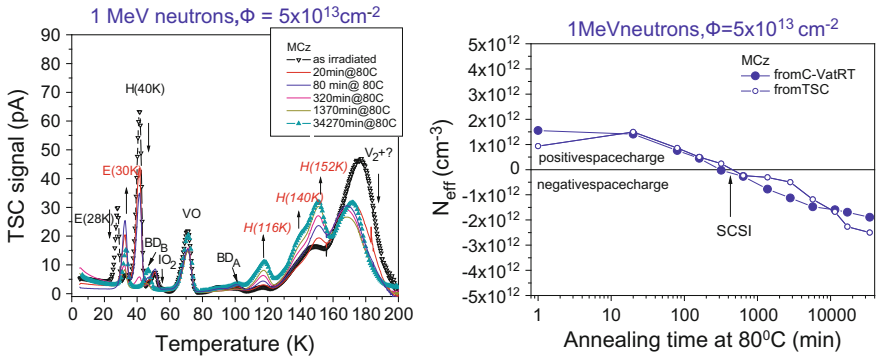


Fig. 2.12 Thermally Stimulated Currents Method scans were done after each annealing step. The rise of the microscopic levels H116K, H140K, H152K can be observed after each annealing step in the *left* figure while the *right* figure shows the corresponding change in depletion voltage (N_{eff}) for each annealing step, determined by the CV characteristic and a TSC scan [241]

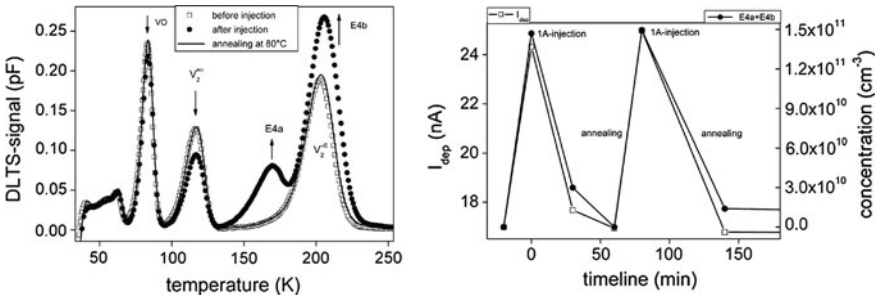


Fig. 2.13 The DLTS measurement on the *left* show a clear rise of the E4 concentration after current injection (1 A forward current) and a decrease to the original level after isothermal annealing of 80 °C. The leakage current strongly follows the E4 concentration, shown in the plot on the *right* [159]

the same defect complexes in a different "charge" state. Their annealing behaviour is similar to that of double-vacancies, suggesting that the complex is an n-vacancy complex.

The knowledge, which defects corresponds to which macroscopic parameter and which irradiation type introduces them, has grown considerably in the past decade. In addition the knowledge which defect is a point defect and which a cluster improves our understanding of the annealing process. All this can be fed, in a more simplistic way, into dedicated simulation and we start to achieve predicting power of parameter evolution with radiation and annealing.

2.2.2 *Different Materials and Different Radiation Types – NIEL Violation*

Already RD48 [191] proved the beneficial effect of high oxygen concentration in silicon material (DOFZ) with respect to depletion voltage evolution, violating the NIEL hypothesis – see also Fig. 2.10. This led to the exploitation of Czochralski material (Cz) and later to magnetic Czochralski (mCz[8]) where oxygen enrichment comes naturally during the melt process.

Radiation damage studies produced surprising results and in Fig. 2.14 no distinct SCSI point is present for these materials. After a long campaign of CV and TCT[9] studies, it became clear that with the new materials and with high fluences applied, one can no longer assume a linear electric field with one single junction at one side. A double peak or double junction can be qualitatively explained by two opposite linear fields at both ends defined by different space charge regions at both ends and possibly a zero or constant field region in the middle. More quantitatively, fits suggest a parabolic field throughout the sensor volume (more in the next Sect. 2.2.3). Often, with charge trapping, charges (TCT signals) drifting from the injection side are trapped before they reach the other side and double peaks are smeared out; thus a trapping corrected TCT analysis is mandatory.

As a result, the depletion voltage parameter becomes a more abstract concept and for high radiation levels, CCE or better signal-to-noise becomes the more realistic and important parameter[10] to measure. Consequently these are the main parameters evaluated for the HL-LHC upgrade program; see Sect. 7.1.1.

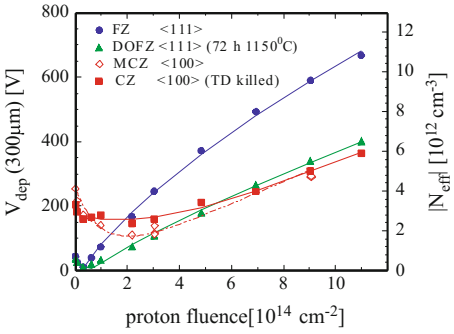


Fig. 2.14 Cz and mCz (*red points*) do not exhibit the distinct point of space charge sign inversion SCSI [220]

[8] An applied magnetic field during the melt creates an electric current distribution and an induced magnetic field. The active Lorentz force then dampens the oscillations in the melt, resulting in a more homogeneous oxygen distribution.

[9] In a **T**ransient **C**urrent **T**echnique TCT measurement the current slope represents the field and a sign change in slope indicates SCSI. Today we see a double peak thus a double junction (see also Fig. 1.48 and Sect. 2.2.3).

[10] With higher and higher "depletion voltages" even above a possible operation voltage, the only important parameter is the collected charge at the amplifier.

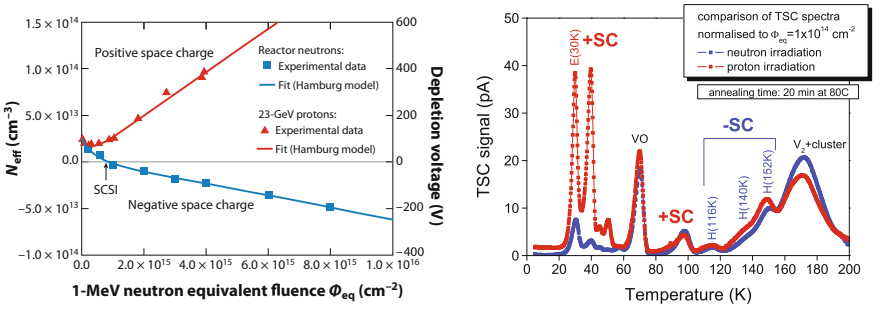


Fig. 2.15 Change of N_{eff} in EPI-DO material versus irradiation with different particles. Acceptor introduction is enhanced for neutron irradiation, similar to n-FZ material, while protons generate mainly donors [242]. In the corresponding study, with the Thermal Stimulated Current TSC method, the deep level states E30K have been identified to act as donors [158]

Furthermore, it has been realized that for some materials charged particles introduce distinctly different defects than neutrons. Figure 2.15 shows for EPI-DO (DO – diffused oxygen) the introduction of negative space charge after neutron irradiation with the corresponding SCSI. Instead, for protons, donor generation is enhanced (positive space charge) and therefore no SCSI is observed.

In the case of n-FZ sensors, both neutron and proton radiations introduce predominantly p-type defects. In the case of n-mCz, the neutrons introduce mainly acceptor (p-type) defects while charged particles produce mainly donors (n-type) defects – a clear violation of the NIEL hypothesis. This particular feature of the n-mCz silicon can have a favourable consequence on the degradation rate of the electrical properties of the detectors when the damage is due to a comparable mix of neutron and charged hadrons because the radiation induced defects can partially compensate each other [177]. To test this effect, n-in-n FZ and n-in-n mCz detectors have been irradiated with neutrons only, 25 MeV protons only and with an equal mix of neutrons and 26 MeV protons to a total dose of $1 \cdot 10^{15}$ $n_{1\text{MeV}}/\text{cm}^2$. Figure 2.16 shows the CCE(V) measurements of these devices and confirms the compensation effect. The two n-FZ detectors exhibit almost identical CCE(V) characteristics after the neutron, proton and mixed irradiations, while the n-mCz shows a faster rise of the CCE(V) in the case of mixed irradiation relative to the neutron and proton irradiations. The compensation effect of n-mCz is very interesting for locations/radii with similar radiation levels from neutrons and charged particles but, in reality, the location with the highest levels (inner radii) are largely dominated by charged particle radiation.

Another interesting plot showing the difference in annealing of FZ and mCz can be found in Fig. 7.17 on page 313.

The Role of Oxygen as we understand it

In the inner pixel detector, ATLAS and CMS are using sensors processed out of **D**iffusion **O**xygenated **F**loat-**Z**one DOFZ wafers. Materials with high oxygen concentration are high on the ingredients list for the future upgrades, due to the findings of RD48 & RD50. As shown in the previous and later sections higher oxygen

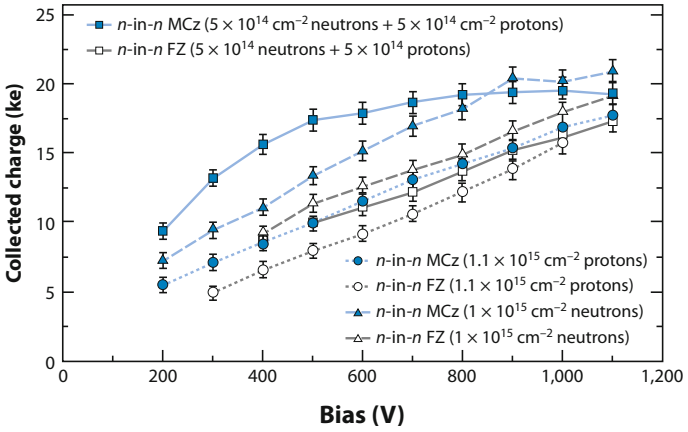


Fig. 2.16 Charge Collection Efficiency of mCz and FZ detectors after a total dose of $1 \cdot 10^{15}$ $n_{1\mathrm{MeV}}/\mathrm{cm}^2$ obtained with neutrons only, 26 MeV protons only or mixed (equal dose of neutrons and 26 MeV protons) irradiation. For FZ sensors (open bullets) neutron, proton or mixed irradiation give similar results at the same equivalent NIEL fluence – the mixed irradiation is just the average of protons and neutrons. For mCz (solid bullets), the picture is quite different – the mixed radiation shows a much higher CCE than either proton or neutron only, specifically at lower voltages. This is a clear indication of donor – acceptor *compensation* with the different radiation types (decrease of $|N_{eff}|$). At higher voltages, above depletion, the CCE difference becomes marginal

concentration seems to improve the radiation tolerance with respect to the change in depletion voltage (see e.g. Fig. 2.10 on page 149 and Fig. 7.17 on page 313). In general, in oxygen enriched silicon, the built-up of net negative space charge (acceptors) after charged hadron radiation is suppressed. Oxygen catches/binds vacancies. Therefore some cases (EPI, mCZ) do not exhibit "type inversion" after charged hadron irradiation and/or exhibit a donor/acceptor compensating effect (see former sections), clearly violating NIEL. Figure 2.17 shows a clear microscopic to macroscopic correlation: The radiation induced E(30K) concentration in oxygen enriched material is much higher while the H116K, H140K, H152K defect concentration seems unaffected by radiation. E(30K) is an electron trap and a donor (positive space charge) in the upper half of the Si band-gap while the unaffected H complexes are hole traps with acceptor levels (negative space charge) in the lower band-gap; the H complexes are also relevantly responsible for the reverse annealing. Thus the donor E(30 K) is an oxygen-related defect.

Another beneficial side effect is that interstitial oxygen strengthens the lattice and reduces the brittleness of the sensor.

New Materials and NIEL

Obviously, the "old" NIEL mantra is not really adequate any more for the new materials! Charged particles damage differently, protons may even compensate for neutron damage. NIEL is still useful for scaling between different proton energies especially to evaluate the leakage current after hadron irradiation. But, while new

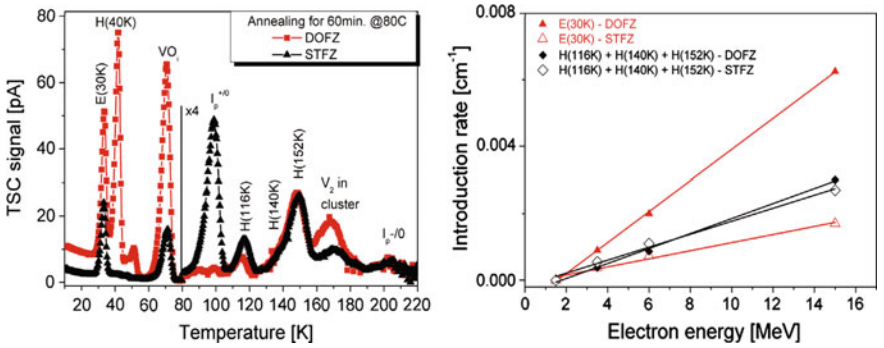


Fig. 2.17 The *left* figure shows the the microscopic defects for standard FZ and diffusion oxygenated Float-Zone DOFZ material determined with the TSC technique; both samples have been irradiated with 6 MeV electrons (creating point and cluster defects). The corresponding introduction rate for E(30K) (donor) and H (acceptor) defects versus electron energy is shown in the *right* plot. Clearly E(30K) introduction (red triangles) is enhanced at high oxygen concentrations while H (black diamonds) is not. More information in [249]

materials seem to be more radiation tolerant, a complete evaluation of each material must be conducted separately for neutron, proton and mixed irradiation. Even the correct radiation mixtures at different radii in the experiment should be checked. Much more complicated and extensive campaigns are necessary to evaluate new materials, processing schemes or companies.

2.2.3 Double Junction

One of the important questions in the last years was: "Does material x,y type-invert or not?". Another question: "Why do we collect charges from considered un-depleted zones in a given device after high radiation?" Is the simple description of a linear field through the full sensor bulk, as depicted in Fig. 1.8 on page 14 still applicable after high irradiation?

Figure 2.18 and earlier Fig. 1.48(lower right) on page 76 show TCT spectra not compatible with a standard assumed linear field across the entire sensor bulk. No clear single slope is visible but a double peak representing fields on both bulk ends with opposite sign. Since the fields are results from space charge, we also expect opposite space charge at the different ends. More examples of the phenomenon will be given, ending with an explanation.

Figure 2.19 depicts the expected linear field configuration after inversion (left) plus a simple two linear field configuration approximating the double junction (middle) and finally the reality parabolic situation determined by simulation and further evaluation of the double peak (right). The lower part of Fig. 2.19 shows the TCT spectra of a 320 μm thick *p*-in-*n* diode after radiation with front side (electron) injection with red laser, plus its derived electric field configuration.

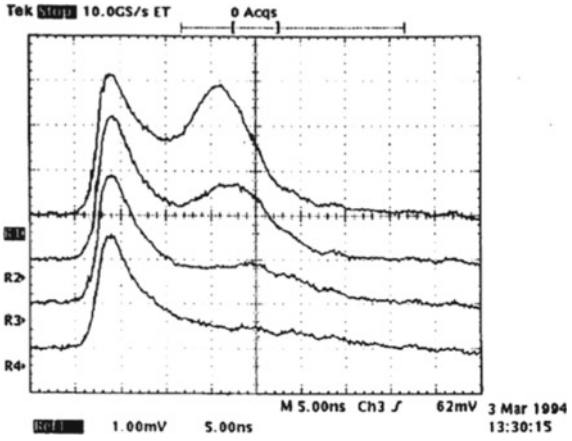


Fig. 2.18 Oscilloscope photo of one of the early reported double peaks in a TCT scan [94]

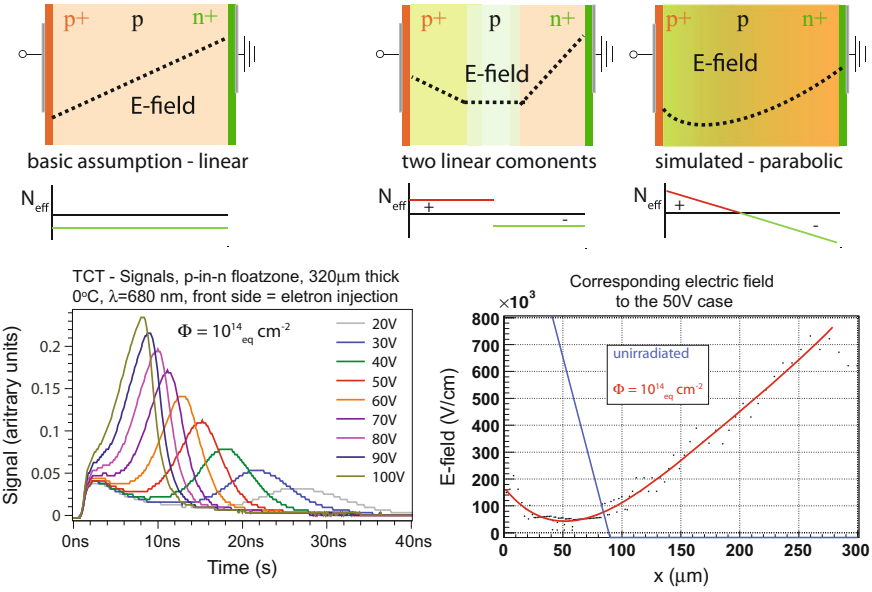


Fig. 2.19 The *left upper* picture depicts the standard linear field expected after radiation – before trapping becomes relevant. The *middle* one shows a double junction described simply by two linear fields and, on the *right*, the parabolic truth derived by simulations. Below the corresponding effective doping concentrations N_{eff} are depicted. The *lower* figure shows the TCT current scans for different voltages; charges are injected by a red laser from the front side (electron injection). The *black dots* on the right show the 'necessary' field to reach the next velocity value of the TCT plot; the *red* represents a simple fit of the points while the *blue* depicts the situation of the corresponding field configuration before radiation [90]

Why do we have a parabolic field after high radiation levels and not the simple expected reversed linear field after type inversion as we have for lower radiation levels? It's another manifestation of bulk current generation centres and, more important, trapping centres in the band-gap. Where is the asymmetry coming from? Defect level concentrations are constant across the entire bulk. But thermally generated charge carriers drift in the electric field to opposite sides (holes travel towards p- and electrons towards n-electrodes) creating an asymmetric but dynamic situation. With the presence of deep level defects, these charges are "trapped", thus "static" and thereby alter the space charge asymmetrically. N_{eff} is no longer uniform leading to a parabolic field across the bulk also described as double junction – distinct junctions at each sensor diode face. N_{eff} corresponds to n (p) doping at the p^+-side (n^+-side).

For values below depletion voltage, the un-depleted zone is located in the middle of the sensor instead of at one end. Still even in the un-depleted zone, fields are present visible in the non-zero current of the TCT and edge-TCT signal (see Figs. 2.19 and 2.21). For $V \gg V_{FD}$ when the field fully reaches through the sensor, one junction, the "main junction", dominates and the field becomes 'more' linear again, the double peak smooths out. After the onset of trapping, the field configuration changes even more with radiation and annealing because it is now defined by three components – intrinsic doping concentration (including radiation defect levels), leakage currents and trapping, all changing with radiation but also with annealing. In addition, charges from ionisation are also trapped thereby altering N_{eff} and thus field configuration. More on the topic in [94, 213]. As for the question about type inversion or no type inversion, often one peak is hidden by trapping thus injecting from one side or the other show a different peak thus a different slope (the one from the visible peak) thus hints to inversion or no inversion depending on the injection side.

The following paragraph shows how the double junction reflects in operation with particles. A CMS n-in-n pixel sensor (DOFZ material) after irradiation beyond the point of type inversion ($\Phi_{eq} = 8 \cdot 10^{14}\ n_{1\mathrm{MeV}}/\mathrm{cm}^2$) is being investigated in a dedicated test beam campaign [65, 88]. The particle beam hits ("grazes") the sensor under a very small angle and thus the traversing particle path in the sensor is very long,

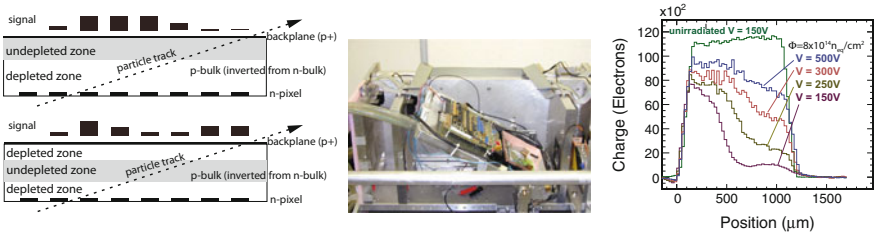


Fig. 2.20 On the *left* we see a cartoon of the incident angle, the position where the ionisation happens and the expected signal distribution. The under-depleted case is depicted for a fully type-inverted case and (*below*) for the double junction case where the un-depleted zone stays in the middle and not at the p^+-face. The expected signal distribution is significantly different. The *middle* photo shows the setup. On the *right*, the data shows a simple flat line for a fully depleted un-irradiated sensor plus the data for under- and fully depleted irradiated sensor given by the different voltages. The data is incompatible with a simple case of type inversion [65, 88]

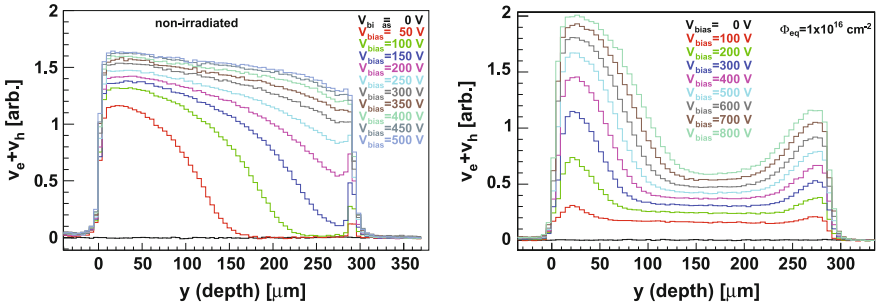


Fig. 2.21 Velocity profiles for different charge injection depths with edge TCT for an un-irradiated and in comparison for a highly irradiated sensor with a double junction. The scans have been done for several voltages mainly all below depletion voltage (which is about $V_{FD} \gg 800\,\mathrm{V}$). The velocity in the middle expectedly un-depleted region is substantially above zero proving an existing electric field. More details and definitely worth to read in [179, 201]

spanning multiple pixels. The z-position corresponds basically to the depth where the ionisation happens; it is called grazing angle method. The cartoon in Fig. 2.20 illustrates simplistically where the charge is created, and the resulted signal distribution expected for a full "type inverted" and one for a double junction configuration (no signal from the un-depleted zones). The measured distribution is incompatible with full "type inversion": (a) charges are collected in the entire volume and (b) the double peak at low voltages indicates high field on both sensor faces and (c) the integrated charge/signal does not scale with the expected depletion depth (with depletion depth $d \sim \sqrt{V}$, $V_B = 300\,\mathrm{V}$ should give $\sqrt{2}$ more signal than $V_B = 150\,\mathrm{V}$). Setup, field configuration and results are complex and only a full simulation sheds light and indicates a parabolic field configuration.

Again, the edge-TCT method (see Sect. 1.8.3 - Edge TCT) demonstrates its strength in Fig. 2.21 investigating the double peak structure injecting charges at certain depths of the sensor and measuring the velocity profile for different voltages – see [178]. The results confirm the above described picture and show a) high fields at both sensor ends but also substantial field strengths in the middle region, even for voltage below V_{FD}. The concept of depletion voltage becomes academic: electric fields are present throughout the sensor and charges are being collected via drift, not random walk. Looking closely, the velocity peak at the back for $V = 500\,\mathrm{V}$, is even larger than the one expected for saturated drift velocities. This hints to charge amplification as described in the next section.

Double junction in a nutshell:

- at high radiation levels with traps present, we have a parabolic instead of a linear field with high but opposite sign maxima at both sensor faces
- the concept of "type inversion" becomes academic
- the concept of V_{FD} becomes academic, since we have field everywhere and therefore directed charge drift everywhere, also below V_{FD}

  - charge collection also from the un-depleted zone

- the different space charge concentrations N_{eff} are due to polarisation, a result from different filling of traps at the different sensor ends (holes/electrons in traps at n/p electrode respectively), due to different drift direction
- the double junction has been proven by standard TCT, grazing angle method in a test beam, and further investigated by edge-TCT. The complex situation is been well described by simulation and simulations are necessary to fully describe, compare and understand the situation
- in the end field strength matters!

2.2.4 *Sensors After Very High Radiation Levels*

With higher fluences, around 10^{16} $n_{1\text{MeV}}/\text{cm}^2$, trapping (trapping time) $\tau_{eff} \sim \Phi_{eq}$ becomes the dominant damage factor, where the electrons and holes, from the ionising traversing particle, are trapped before they reach the readout electrodes. The concept of depletion voltage becomes more and more academic at these fluences. Radiation induced introduction of trap levels differs for the different materials (n, p, FZ, mCz, EPI, oxygenated) but, as reference point, NIEL is applicable for the different radiation types and energies. Substantially large differences exist for electron versus holes collection (n- vs. p-electrode configuration).

Figure 2.22 teaches us that at very high fluences trapping becomes the dominant damage factor (reducing signal) and different particle radiation result in the same effective CCE.

Sensors deployed at the LHC are radiation tolerant up to about 10^{15} $n_{1\text{MeV}}/\text{cm}^2$. The situation gets difficult at fluences of $\sim 5 \cdot 10^{15}$ $n_{1\text{MeV}}/\text{cm}^2$ and above, as foreseen

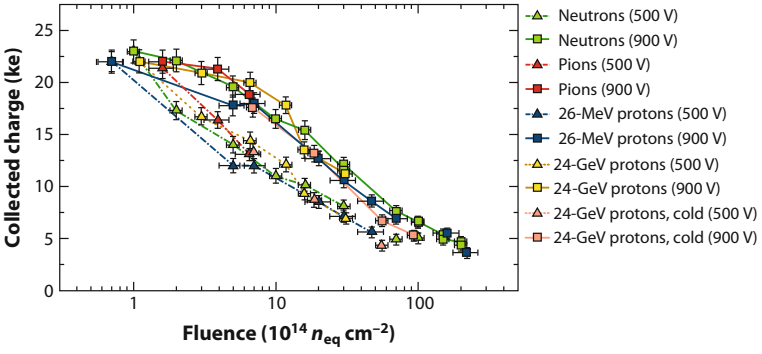


Fig. 2.22 The plot shows CCE for n-in-p FZ strip detectors versus fluence of different particles. At high fluences trapping becomes the dominant factor and damage becomes almost particle independent. The knee in the most right tail looks even a bit too high and could be a hint to charge amplification [11]

for the HL-LHC. New materials and detector schemes had to be developed mainly within RD50 and LHC collaboration efforts (more in Sect. 7.1).

At these fluences, the main relevant question is how much charge is being collected, or better how much charge participates via induction to the signal before being trapped? Adequate signals can be achieved by a combination of the strategies below:

- maximize electric field at the collecting electrode; where the weighting field $\mathbf{E}_W$ is also large
  - if the electric field cannot be established in the full volume; have it at the readout electrode
- establish a geometry with a favourable weighting field
- minimize drift length
- maximize $\mu \cdot \tau_{eff}$ – read electrons

All the above is true for strip/pixel sensors with n-electrode readout either in an n-in-n or n-in-p configuration. After radiation (after SCSI), the depletion zone grows from the n-side and therefore also under-depleted operation is possible differently to p-in-n sensors. n-in-n and n-in-p pixel/strip sensors have a favourable combination of weighting and electric field in heavily irradiated sensors. In reality $\mathbf{E} \cdot \mathbf{E}_W$ after irradiation is much larger for n-in-p compared to p-in-n. Also the collection of electrons seem favourable due to their higher mobility μ and small τ_{eff}; they seem less affected by trapping. For a more detailed discussion the reader is also referred to [173].

The superiority of electron readout is less obvious for pad sensors where electron and holes participate equally to the signal due to the induction process. In general thinner sensors have a higher field but also less volume for the initial ionisation process of the traversing particles and therefore less charges to begin with. This has to be decently balanced. We will see in the next paragraph that field strengths after irradiation can be as large as to amplify the signal. 3D sensors described in Sect. 1.12.7 are reducing drift length by etching narrow electrode columns through the entire sensor volume with spacings much smaller than the sensor thickness.

The baseline choice for the ATLAS and CMS upgrade are n-in-p sensors substantially cheaper than n-in-n ones, which need double-sided processing. Thin n-in-p sensors are even candidates for the innermost radius of the future HL-LHC experiments together with the more special 3D sensors. A drawback of n-electrode readout (n-in-p or n-in-n) sensors is the fact that the high voltage reaches the sensors sides and upper face where additional measures of insulation have to be applied to allow for the readout at GND potential. This is especially true for pixel sensors where the chips are bump bonded directly to the sensors.

More about HL-LHC strategies, plans and R&D in Sect. 7.1.1 where also the annealing behaviour of n-in-p sensors is being discussed.

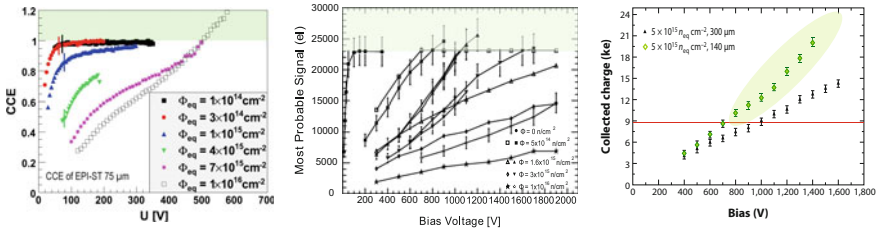


Fig. 2.23 Several groups claim to collect more charge after irradiation than before and even more charge than a MIP deposits in the given material volume. The first plot [188] shows a higher signal in n-EPI material after irradiation, the second [203] a higher signal in p-FZ sensors after neutron irradiation (reactor Ljubljana). The last [59] shows a higher signal in a p-FZ 140 µm thin sensor with respect to the 300 µm thick sensor and also with respect to charge deposited by a MIP in the corresponding volume (after $5 \cdot 10^{15}$ $n_{1\text{MeV}}/\text{cm}^2$ with 25 MeV-p). Clear signatures of charge amplifications have been identified

2.2.5 *Charge Amplification*

Several devices of different material exhibit, after high irradiation, a higher Charge Collection Efficiency than before.

In several cases, more charge per volume has been recorded than the charge deposited by a MIP due to ionisation. Figure 2.23 shows three examples hinting at a charge *amplification* mechanism. It is now of utmost importance to evaluate if the charge amplification is really the wished modus operandi for silicon sensors in the HEP environment. How is the leakage current and the noise affected, what is the resulting effective signal-to-noise? One study shows a correlation of charge collection and leakage current; see Fig. 2.24. The amplification mechanism works for electrons coming from signal as well as from dark current. The impact ionisation comes from the local changes in N_{eff} and the corresponding change/increase in electric field. The situation becomes even more complicated since N_{eff} also changes with annealing and thus the level of amplification can change with time and radiation fluence.

Dedicated investigations with the edge-TCT technique (described in Sect. 1.8.3) are shown in Fig. 2.25. The measurement exhibits a direct indication for charge amplification, the second time-delayed peak in the current pulse can be explained by electron–hole creation at the very high electric field at the strip face. The effect is very similar to gas wire detectors, where ions from the primary ionisation are registered early while a larger ion signal from the secondary avalanche ionisation within the high field region near the wires is recorded much later. The corresponding holes from the amplification process have been excited later than the original ones from light injection and then drift from the strip region to the backside. A second observation by this method, not detailed here, is that the velocity and electric field profiles do not give a consistent picture without charge amplification (more in [178]).

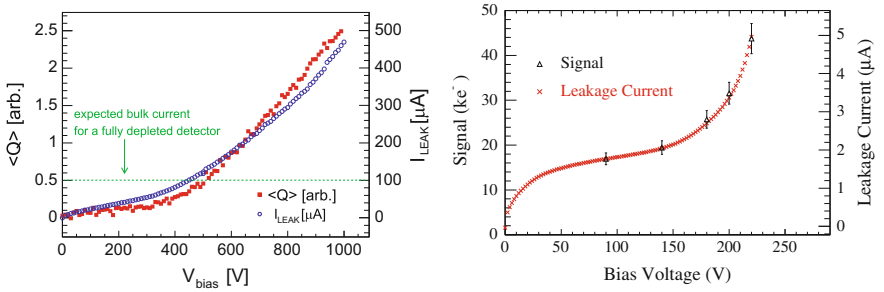


Fig. 2.24 Collected charge versus leakage current in the regime of amplification. Two examples indicating that charges generated by traversing particles as well as the thermally generated charges that constitute the leakage current are multiplied by the same factor once the field is high enough for avalanche multiplication. The *left* [178] is a measurement on a planar sensor, the *right* [230] on a 3D sensor

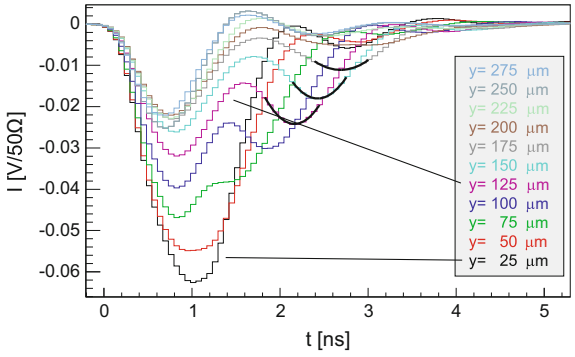


Fig. 2.25 The second peak in the hole tail is evidence of charge amplification in n-in-p sensors in the high E-field strip region. The first peak corresponds to charge carriers from the initial laser light, the additional peak in the tail represents additional holes created at a later time together with electrons in the amplification process when the electrons reach the high field strip region; the holes then need to drift to the back [178]

In the end S/N, efficiency, resolution and power consumptions are the important parameters. Dedicated designs might be able to make use of the amplification feature. Inspired by the amplification, in the framework of RD50 dedicated structures are being designed and produced, namely **L**ow **G**ain **A**valanche **D**etectors LGAD [56, 232, 233, 262] – more in Sect. 1.12.8. LGAD represents a new concept of silicon radiation detector with intrinsic multiplication of the charge. These new devices are based on the standard Avalanche Photo Diodes APD, normally used for optical and X-ray detection applications, but they give a low gain, suitable for detecting high energy charged particles, and allow fine segmentation pitches.

Another aspect of amplification sensors is the increase in speed/slew rate and they open the possibility for fast timing [263]. CMS is exploring to use LGADs as a timing layer in the forward direction in front of the future high granularity calorimeter for the HL-LHC and they are candidates for ATLAS/CMS very forward detectors.

2.3 Surface Damage

The term surface damage describes all radiation-induced damages in the SiO_2 layer and in the SiO_2–Si interface. This effect is therefore prominent in AC-coupled sensors and in the NMOS and PMOS transistors of all kinds of electronics.

The damage is introduced by ionisation and not atomic displacement, unlike in the silicon bulk damage described earlier. In contrast to the situation in the silicon bulk, creation of electron–hole pairs is not fully reversible in an insulator. Depending on the oxide quality, recombination varies between several percent and almost 100%. In addition to recombination, generated charge carriers can also be captured by existing defects, where the emission is highly suppressed, since the band gap is much larger in oxide and nitride layers ($E_g = 8.8$ eV in SiO_2 and $E_g = 5$ eV in Si_3N_4). In the oxide, the mobility of electrons ($\mu_{e,SiO_2} \approx 20\ \frac{\text{cm}^2}{\text{Vs}}$) is several orders of magnitude higher in the oxide than that of holes ($\mu_{h,SiO_2} \approx 2 \cdot 10^{-5}\ \frac{\text{cm}^2}{\text{Vs}}$). This leads to a fast separation of electron–hole pairs. Electrons drift to the metal electrode, while the holes drift by a hopping mechanism via shallow levels to the Si – SiO_2 interface, especially when a voltage is applied (Fig. 2.26). The effect is enhanced for a positive voltage applied on the metal side during radiation; electron movement is accelerated to the metal side and holes drift to the interface, a standard mode for a transistor. In addition the defect concentration is especially high at the interface due to lattice mismatch and dangling bonds. The trap density in the interface region is around 10^9 – 10^{10} traps/cm^2. As a result, positive static charges accumulate at the interface, saturating at about $N_{ox} \approx 3 \cdot 10^{12}$ cm^{-2} [358]. The introduced additional oxide volume and interface charge concentration is now increasing the flat-band voltage by

$$\Delta V_{flat} = -\frac{1}{\varepsilon_{ox}\varepsilon_0}\left(\sigma_{int} d_{ox} + \int_0^{d_{ox}} \varrho(x) x dx\right) \tag{2.21}$$

where d_{ox} is the oxide thickness, σ_{int} is the surface charge at the Si – SiO_2 interface and $\varrho(x)$ reflects the introduced additional charge density in the oxide volume.

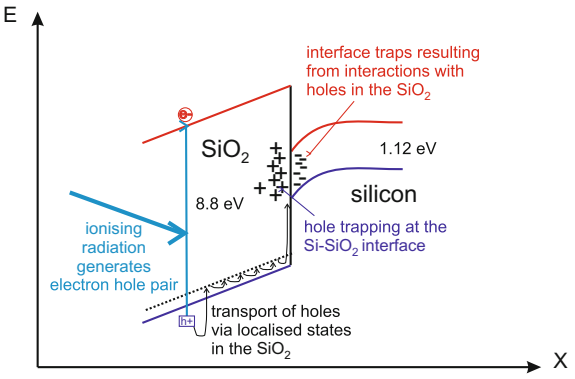


Fig. 2.26 Radiation damage in the Si – SiO_2 Interface

Therefore, measuring the flat-band voltage changes is a diagnostic tool to investigate surface damage. Negative charges are now attracted from the bulk side towards the interface and accumulate there, decreasing the inter-strip resistance and increasing polarisability, thus capacitance. For thick oxides, with a large enough positive oxide charge the 'surface depletion' can switch to 'electron accumulation', decreasing inter-strip resistance. As a second step also negative traps are attracted towards the interface from the bulk side. Figure 2.27 displays the field configuration without and with high oxide charge concentration. Section 6.4.2 describes how the use of <100> silicon minimizes the effect by reducing the number of dangling bonds.

To summarize, due to ionisation, insufficient recombination and subsequent trapping of holes at the Si – SiO_2 interface, the following macroscopic results deteriorate the sensor functionality.

- increase of inter-strip capacitance, thus increasing noise
- decrease of inter-strip resistance, thus increasing cross-talk
- increase of flat-band voltage, as an indicator for oxide charge.

A small annealing effect at high temperatures ($T \gg RT$; around 100 – 400 °C) is also observed here, explained by the drift of bulk electrons into the near-interface region and recombination with holes trapped there. Unfortunately the reverse annealing effects at these temperatures do not allow to benefit from surface damage annealing.

While the damage mechanism in electronic circuits is the same as for the MOS structure in AC-coupled sensors, the static charge centres have several additional effects for NMOS and PMOS. Figure 2.28 shows the basic schema. The main effects are

- threshold voltage shift of transistor V_{thr}
- increased noise
- increased leakage current.

The main drawback of radiation damage in electronics is the increase of transistor threshold voltage V_{thr} and leakage current. The increased threshold voltage derives

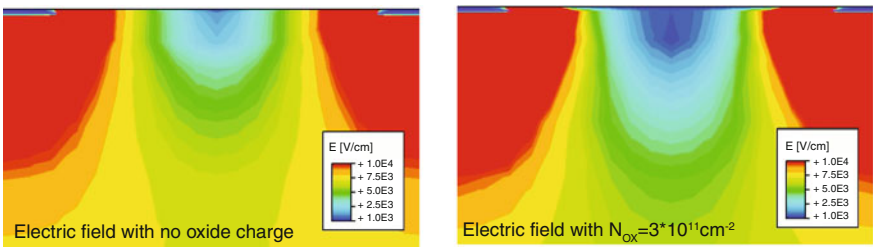


Fig. 2.27 Inter-strip region fields with/without oxide charges. The additional charges do disturb the field distribution. The low-field region in the intermediate region of the strips (*right picture*) allows for negative charge carriers. These are increasing polarisability and thus capacitance and are decreasing the strip-by-strip insulation, which is completely defined by the field distribution [79]

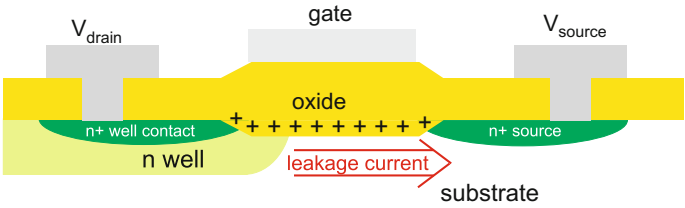


Fig. 2.28 Scheme of an NMOS transistor with deteriorating oxide charge from radiation. The oxide charge screens the gate voltage and therefore a higher threshold voltage V_{thr} is needed to operate the transistor. The resulting attracted charge carriers in the substrate region increase leakage current. The accumulating negative traps in the substrate finally affect mobility. Resulting energy levels in the mid-band region also reduce lifetime and therefore increase leakage current

from the screening effect of the oxide charge concentration. The shift of V_{thr} is proportional to a power n of the oxide thickness d_{ox} : $V_{thr} \propto d_{ox}^{n}$. The power factor n depends on the processes used to grow the oxide and its thickness. It has been measured to be in the range of 1 – 3. The increase of leakage current is induced by the traps acting as a mid-bandgap level introducing recombination centres, thus decreasing lifetime and therefore increasing leakage current. In addition the static traps in the interface affect the mobility in the conducting case. Finally, the accumulation of positive charges in the oxide, especially in thick oxides, can form parasitic charge transfer paths in the substrate by creating inversion layers between N-wells in the substrate. This can lead to a decrease in the signal-to-noise ratio or even to short circuits in the transistor. In the case of NMOS transistors, the charge-up of the oxide may reach a concentration that makes it impossible to completely switch off the transistor.

In Sect. 6.4.1 on page 239 the enclosed transistor design together with the deep sub-micron processing is described which is the key technology for radiation-tolerant electronics.

Chapter 3
First Steps with Silicon Sensors: NA11 (Proof of Principle)

3.1 From Semiconductor Detectors in the 1950s as Spectroscopes to First Tracking Devices in the 1980s

Around 1943 in the very early dawn of time, at least with respect to ionising semiconductor nuclear detectors, P.J. Van Heerden in Utrecht[1] was concluding that a *crystal* being able to react to photons via the photoelectric effect could also be sensitive to β and α particles [142, 341]. The first tests were done with AgCl, 4 mm thick, 4 cm in diameter and γs as reference. To establish an insulator or in the language of today, void the material of free majority charge carriers, the crystal was cooled down to low temperature in a dewar with liquid air and a high voltage was applied.

> Principle of the new method. If it is possible to find substances, in which a β-particle is stopped over a short distance, while the ionisation is still sufficiently strong to be measured, this would provide a means to analyze a β-radiation of small intensity. This leads to the question: what will be the ionisation produced by β-rays in liquids, compressed gases and solids? "…"
>
> In *solids* in general, the ionisation current is very small, owing to the small mobility of the ions. Now the idea which underlies the present investigation is that the ionisation of a single β-ray may be well-observable in a special kind of solids: namely, *crystals* that show photoelectric conductivity. Inside these crystals, light is able to free electrons which, in some of these, can move over a distance of several millimetres, if we apply an electric tension to the crystal; why should β-rays not do the same? [142, 341]

A few years later, around 1950, a germanium (Ge) detector was developed in the Bell Telephone Laboratories in Murray Hill, New Jersey, by Kenneth G. McKay [211]. In this case already a pn-junction was set to reverse bias to deplete the volume. Electrons and holes were collected and the resulting charge was amplified.

> The number of electron–hole pairs produced in germanium by alpha-particle bombardment has been determined by collecting internally produced carriers across a reverse-biased pn-junction. "…" the carriers are swept across the barrier in a time of less than $2 \cdot 10^{-8}$ s. "…" The energy lost by an alpha-particle per internally produced electron–hole pair is 3 ± 0.4 eV. [211]

[1] Mededeling uit het Physisch Laboratorium der Rijksuniversiteit te Utrecht, Nederland.

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_3

Later, from 1955 to 1965, when monocrystalline silicon became available, Si detectors became a fashion, for example at the Oak Ridge and the Chalk River nuclear facilities, but also at Harwell in the UK and the CEA in France. At first simple gold surface barrier devices were constructed but within months pn-junctions were established by gaseous diffusion of phosphorus. The excellent energy resolution of a few electron volts predestined these semiconductor radiation detectors as miniature particle spectrometers. Still today all tracking detectors give additional energy information, where the velocity β to dE/dx dependence is described by the Bethe formula.

The first pattern, multiple rectangular diodes on one substrate, was established in 1961 – a pixel device [76]!

The first silicon strip sensors were developed in the early 1970s in the Kernforschungszentrum Karlsruhe, the Argonne Labs, Fermilab and the Department of Nuclear Medicine in Southampton, though they were not used in high energy physics [26, 104, 127, 186]. The ones in Karlsruhe [121] (1971) for example consisted of 5 or 12 gold 3 mm wide strips, spaced by 0.2 mm, while an aluminium contact connects to analogue readout. The purpose of the strips was to measure on the "large area sensor" the angle of the incident α particle from e.g. a $^{12}C(\alpha, \alpha')^{12}C$ reaction while the other electrode measured energy. The gold contacts were produced via a thin wire mask. The sensors were operated at $-20\,°C$ with 30% over-depletion.

The first use of silicon sensors at CERN [33] (I found) was in 1973 where unstructured silicon was used as "live" targets where the ionisation is correlated to the recoil energy. A telescope of five totally depleted silicon detectors was used as a "live" target in a coherent production experiment $\pi^- + Si \rightarrow \pi^+\pi^-\pi^- + Si$ with a 16 GeV/c π^- beam at CERN. The active area of the detectors was 300 mm^2, thickness was 200 μm or 1 mm.

This section tries to assemble all relevant achievements and milestones but an overview of a lively field like early semiconductor developments cannot be complete or exhaustive.

3.2 Development of the First Silicon Strip Detector for High Energy Physics NA11 and NA32

The theoretical expectations in the 1970s for lifetimes of the lowest mass charm states were of the order of 0.1 ps corresponding to $c\tau \sim 30\,\mu m$ with a production cross-section of a few microbarns and the charged multiplicity of hadronic interactions [112].

The requirements for a suitable detector to resolve this are

- spatial resolution: better than 10 μm and good particle separation
- rate capability about 10^6 Hz
- low multiple scattering and photon conversion – thin sensors

With their high density and small energy band gap, silicon sensors collect ~108 (e–h pairs)/μm within some tens of nanoseconds collection time. Silicon is

therefore perfectly suited to be used as thin and high rate detector material. The missing ingredient, very high spatial accuracy and two particle resolution, is achieved by a very fine structuring of the silicon wafer surface into a strip pattern. In the 1980s this was initially achieved by the so-called surface barrier diode method.

The surface barrier diode structure is a straightforward technique but rather delicate. After final cleaning and etching a metal (Au, Cu) strip pattern is applied to form a surface Schottky barrier junction. The pattern can be achieved by evaporation through a mask [141] or a bit more like the planar process with metal etching after applying photo-impression technique [15] as with printed circuit boards.

A new development – planar technology [151, 165] – as used in integrated circuit fabrication allowed a more robust device with higher structure precision. This technique was finally used to produce the silicon sensors for the NA11 and NA32 experiments at the SPS at CERN to identify lifetime and mass of the charm mesons $D^0(c\bar{u})$, $D^-(\bar{c}d)$, $D^+(c\bar{d})$, $D_s^+(c\bar{s})$, $D_{\bar{d}}^-(\bar{c}s)$.

The detectors are made of high-ohmic (3 kΩcm) n-doped silicon single crystal wafers of 2 in. diameter and 280 μm thickness. Using the planar process, p-doped strip diodes, covered by aluminium contacts, are implanted into one side of the wafer. The sensor is DC coupled. On the other side a common aluminium contact is evaporated. The sensor is reverse biased to deplete the diodes of free charge carriers. The bias voltage also generates an electric field in the n-bulk. In order to reduce the number of electronic channels the principle of capacitive charge division has

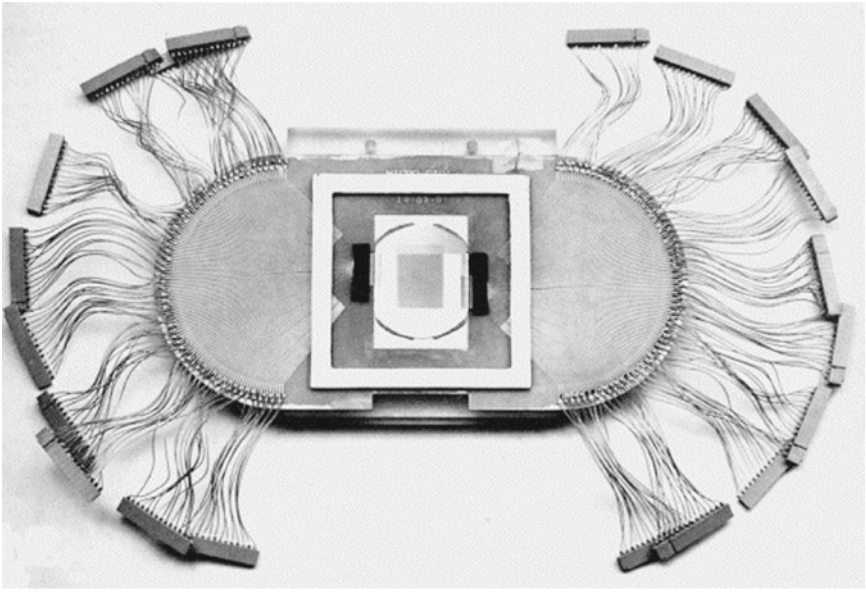


Fig. 3.1 Photograph of a mounted NA11 detector. [151, 165] The sensor seen in the centre is 2436 mm^2 in size with 1200 diode strips and readout of every third (sixth in the outer region) strip, resulting in 4.5 μm (7.9 μm) resolution. Relaxing the readout electronic connection density to 120 μm with a physical pitch of 60 μm is realized by connecting the even strips on one sensor side and the odd ones on the other

been used. Only every third (sixth) strip is equipped with an amplifier in the central (outer) region of the detector. Charges arriving at intermediate strips induce signals in the neighbouring readout strips proportional to their mutual capacitance. The strip pitch is 20 μm arriving at an effective readout pitch of 60 (120) μm central (outer). The resolution was measured to be 4.5 (7.9) μm central (outer); the two particle separation was 60 (120) μm. A photo is displayed in Fig. 3.1.

Comparing the fabrication of sensors in 1980 described in [151, 165] with processing of today described in Sect. 1.9.2, it becomes clear that the planar process is still the baseline for sensor processing

1. use a 500 Ωcm – 5 kΩcm wafer 250 – 500 μm thick as base material (2 in.)
2. apply 0.2 μm $Si\,O_2$ passivation at 1030 °C with dry oxygen or steam leading to 1600 – 6000 Å of SiO_2
3. open the diode windows by photolithography (SiO_2 etching)
4. use an ion implanter to get an asymmetric nearly abrupt *pn*-junction (ion sweeping over full area, hitting only at openings)
   - for *p*-type strip implantation boron was used with an energy of 15 keV with a dose of $5 \cdot 10^{14}$ ions/cm^2
   - the backside was implanted with arsenic with an energy of 30 keV and a dose of $5 \cdot 10^{15}$ ions/cm^2
5. clean and thermal cure (anneal) at 600 °C under dry nitrogen
6. evaporate $\sim$ 1 μm of aluminium onto both surfaces
7. photolithographic etching of strip pattern
8. cut, mount, micro bond

In retrospect the planar technology used by J. Kemmer et al. for the NA11 experiment established some wonderful devices, the core to all current designs. Up to now there is no *real* change in the basic functionality. Especially the conclusion of [151, 165] showed a remarkable foresight:

> We would like to end with the remark that high-precision silicon detectors of the type discussed in this paper have severe limitations for the use in high-energy experiments – in particular as central high-precision detectors for colliding beam experiments. For our counters the ratio of detector surface to the nearby electronics surface is approximately 1:300, which makes it impossible to cover larger areas with detectors. We think that integrating electronics on the detector chip can solve this problem.

After the intermediate step of high integration of electronics in adjacent chips (see Sect. 1.10, Chaps. 4, 5 and 6), today the direct integration of electronics on the sensors itself is a standard (see Sects. 1.12.2 and 1.12.3). Other remarks also influenced further developments, especially for the LEP era:

- detectors with readout strips on both sides to reduce multiple scattering
- large area 3 inch diameter detectors are already available
- large-scale electronics integration for amplification and data processing is needed

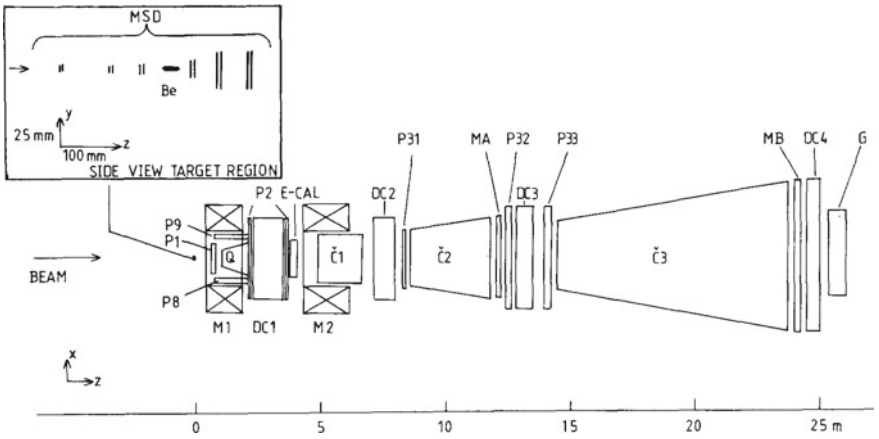


Fig. 3.2 The NA11 spectrometer. [21]. The NA spectrometer can be observed from the *top*. It is showing the Be target, silicon **m**icro**s**trip **d**etectors *MSD*, magnets *M*, drift chambers *DC*, Čerenkov counters (Q, Č), multiwire proportional chambers (*P*), scintillator arrays (*MA, MB*) and electron (photon) calorimeters (E-Cal, G)

3.3 Distinguish *c* Quarks from Others

In this section the detector arrangement of NA11 and NA32 is described and the flavour tagging mechanism by the identification of a second vertex is briefly introduced. The complete NA11 spectrometer[2] is depicted in Fig. 3.2 [21]. This chapter concentrates on the silicon **m**icro**s**trip **d**etector MSD. Eight silicon strip detectors (two in front and six behind the target) of the above-described sensors were used in the NA11 and NA32[3] spectrometer at CERN SPS. The detectors come in plane pairs with $\pm 14°$ inclination with respect to the horizontal. The configuration can be seen in the upper left part of Fig. 3.2. The system is fully suited for a fixed target experiment. All layers are arranged perpendicularly with respect to the beam, the number of layers allows precise reconstruction, especially with the planes in front of the target defining the initial particle flight path, thus giving a good estimation of the primary vertex. The spill rate of the SPS was in the order of 10^4 – 10^6 π^- per second at 175 – 200 GeV/c.

The full beauty and capability of the device are revealed in Fig. 3.3. The measurement is so precise that flight paths can be reconstructed, that secondary vertices can not only be identified but the flight path of the decaying particle from its production to its decay can be quantified. Finally a handful of events showed a charm decay out of several million recorded events. The measured distance between primary and secondary vertex gives the boosted flight path $ct\gamma$ which is directly correlated to the

[2]The other detector types are listed in the figure caption of Fig. 3.2.

[3]NA11 used a Be target, while NA32 used an active silicon target and some trigger decisions were different.

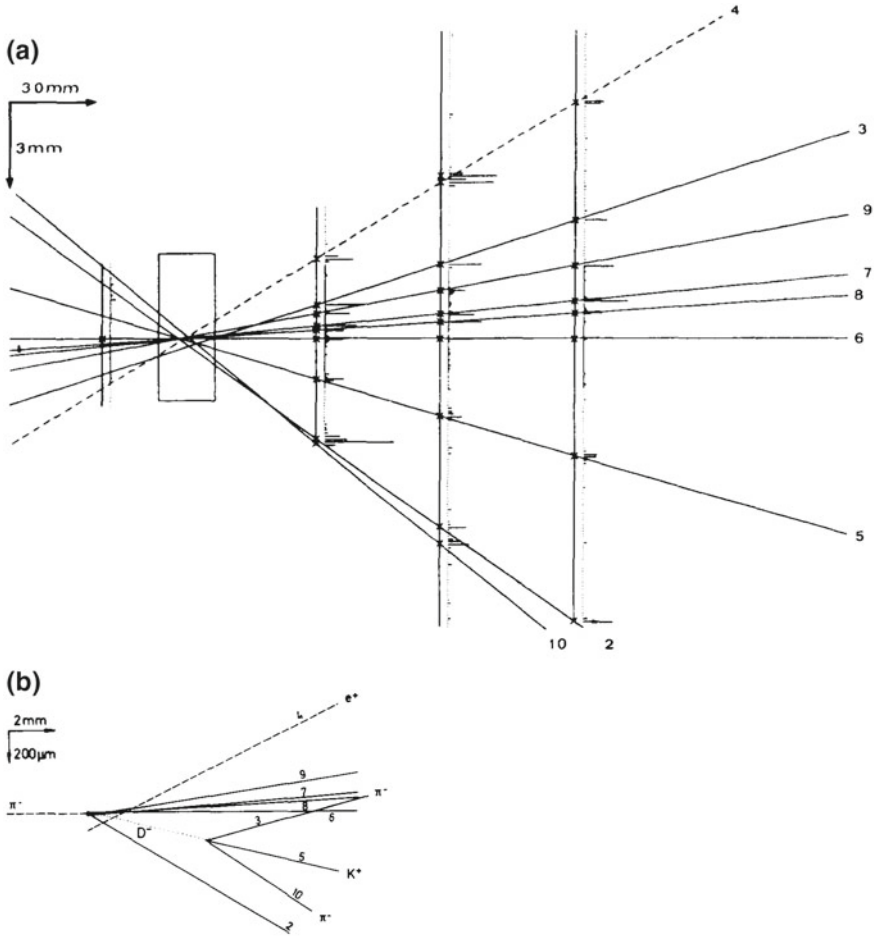


Fig. 3.3 Reconstruction of the production and decay of a $D^- \rightarrow K^+\pi^-\pi^-$ as measured in the NA11 experiment in 200 GeV/c π^- Be interactions [31]. The lengths of the horizontal lines on the planes indicate the measured pulse height in the silicon sensors. The connecting lines represent the reconstructed particle paths. With a precise reconstruction it becomes clear that trajectories 3, 5, 10 are not originating from the same decay point as the others. They are not starting from the primary vertex, but from a secondary vertex. Distance reconstruction gives the boosted flight path $ct\gamma$ of the associated charm particle which is directly correlated to its lifetime

lifetime of the particle. The accuracy of vertex identification is 130 μm. While in the case of NA11 the flight path was used to measure the lifetime, in later experiments the mere existence of a secondary vertex indicates a heavy quark. With the knowledge of the specific quark lifetime, thus flight path, the specific quark can be identified, which gives a lot of information about the full decay.

Chapter 4
The DELPHI Microvertex Detector at LEP

In the **L**arge **E**lectron and **P**ositron LEP collider era, quark tagging and spectroscopy of short-lived particles demanded high-precision tracking with flavour tagging possibilities via the second vertex identification method. Taking the success of silicon sensors in NA11 (see Sect. 3.2) and MARKII [193] into account, silicon was the obvious candidate for the innermost tracking detector complemented by drift chambers and/or time projection chambers at larger radii. Therefore the first silicon vertex detectors were installed in *DE*tector with **L**epton, **P**hoton and **H**adron **I**dentification DELPHI [41] and ALEPH [70], with OPAL [16] and L3 [9] joining soon after.

Figure 4.1 shows the full DELPHI detector configuration with all tracking components, the solenoid, the calorimeters and the particle ID detectors. In this book the DELPHI **M**icro**V**ertex **D**etector MVD is described, the largest silicon detector of its time. Figure 4.2 presents a quick historical walk-through of the initial stage and first upgrades. The main part of the chapter concentrates on the last upgrade finished in 1997 for LEP2 [137, 239, 319]. Looking back on the NA11 conclusion strong efforts were made to compact the structures, especially the readout electronics and support structures, to be able to fit into the small compartment between beam pipe and the inner drift chamber. The first MVD was constrained by the beam pipe at $R = 7.8\,\text{cm}$ and the inner drift chamber at $R = 12$ cm; therefore, only two layers were placed at $R = 9$ and $R = 11\,\text{cm}$.

A hint of the excitement of 1994 can be felt by looking into the internals of the DELPHI Microvertex Detector in Fig. 4.3.

4.1 Design and Strategies

What are the ingredients of the design and what are the requirements of high energy physics?

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_4

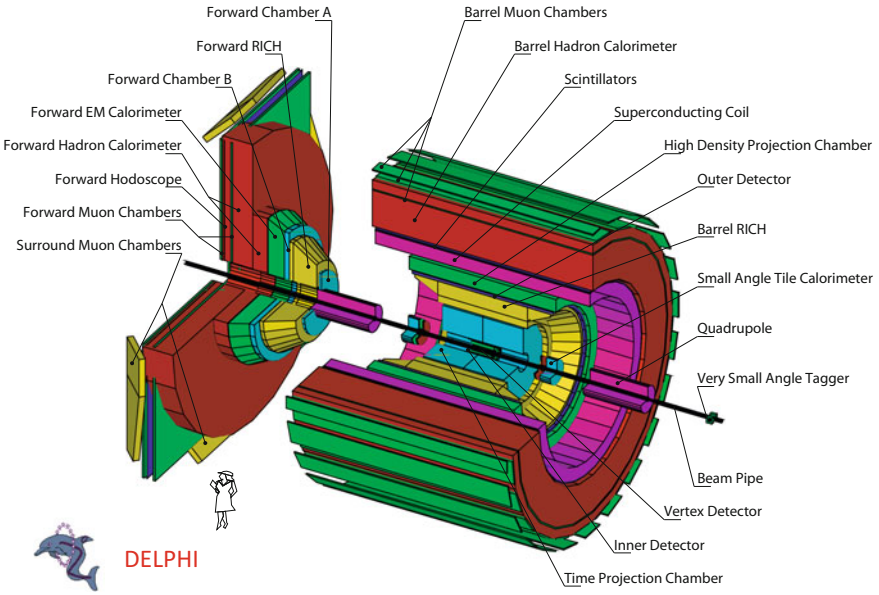


Fig. 4.1 The DELPHI detector, a schematic. The DELPHI detector is a textbook example of a multipurpose detector. Arranged from inside out, the following types of detectors are illustrated: Close to the beam pipe, (*1*) **the Microvertex Detector** *MVD*, a silicon detector, (*2*) the Inner Detector *ID*, a drift chamber, (*3*) the Time Projection Chamber, *TPC*, (*4*) the Ring Imagine Čerenkov Counter, *RICH*, (*5*) the Outer Detector *OD*, drift tubes, (*6*) the high density projection chambers, electromagnetic calorimeter, (*7*) the solenoid with 1.2 T, (*8*) Time of Flight, *TOF*, scintillators, (*9*) hadronic calorimeter, (*10*) muon chambers, drift chamber. [Courtesy of DELPHI, CERN [318]]

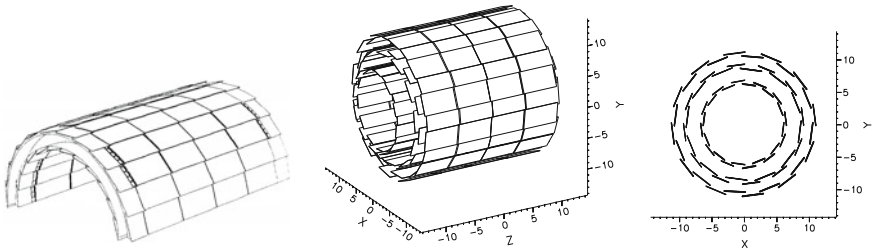


Fig. 4.2 History of the DELPHI vertex detector. DELPHI started in 1989 with two layers with single-sided readout, a R and ϕ sensitive detector as the first silicon vertex detector at LEP – a so-called $R\phi$ detector. In 1990 the beam pipe radius was reduced and a third layer, still single-sided, $R\phi$ only, was introduced. Its 1994 upgrade finally contained double-sided[1] silicon sensors on three layers to establish a 3D readout of R, ϕ and z. The middle scheme shows the 3D view, while the right gives a view from the perspective as the beam particle would see it. The final upgrade in 1996/1997 is displayed in Fig. 4.4. In every upgrade sensors were reused from the former versions, due to cost effectiveness [18, 63, 198, 352]

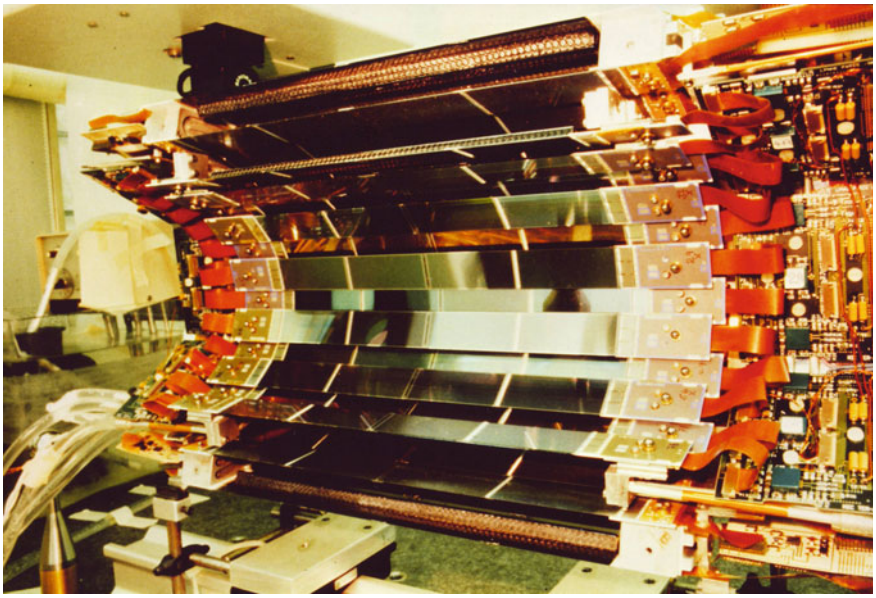


Fig. 4.3 The DELPHI silicon detector, view from within. After its first upgrade in spring 1994 the DELPHI **M**icro**V**ertex **D**etector MVD was State of the Art. The photo shows a rare view during construction, with all the modules in hand and three layers of the barrel detector assembled. The "low" number of modules allowed them to be handmarked. Compared to LHC detectors, it was quite small but the principle of geometrical module placement is still educational. The modules were placed in a so-called "staggered" arrangement that the adjacent modules overlapped by a few strips. This allowed for tracks through neighbouring sensors to be used for alignment and to make sure that the complete cylinder is covered. One of the modules is detailed in Fig. 4.6 while the exploded view of the hybrid sitting at both ends of the modules is presented in Fig. 4.7. [Courtesy DELPHI, CERN]

Starting from the physics point of view, secondary vertices need to be tagged mainly to identify heavy quarks[2] and the τ lepton to e.g. measure lifetime but equally important is to reduce background. The corresponding lifetimes are around 0.2 - 1.5 ps translating to flight paths $c\tau\gamma$ before decay in the millimetre regime at LEP energies. Track reconstruction must be precise enough to extrapolate back into the interaction region, resolving primary, secondary and tertiary vertices. An example of track and finally vertex reconstruction can be seen in Fig. 4.16 on p. 190. The ability to reconstruct or detect secondary vertices can also be expressed by the *impact parameter resolution* σ_{d_0}. The impact parameter d_0 is defined by the shortest distance between a reconstructed track and the primary vertex, it is a crucial quality parameter of the full detector performance. If the measured *i*mpact *p*arameter IP is

[1] DELPHI followed the example of ALEPH, which started right away with double-sided silicon sensors.

[2] Heavy quarks from e.g. J/Ψ containing a c quark or the Υ meson containing $b\bar{b}$.

significantly larger[3] than the experimental resolution in this quantity, a secondary decay vertex is probably present. How the impact parameter is related to physics will be discussed later, this section concentrates on the detector design aspects. d_0 is dependent on the detector geometry and multiple scattering, hence the material budget obstructing the flight path. For a simplified two-layer system the variance of d_0 can be expressed by

$$\sigma_{d_0}^2 = \sigma_{MS}^2 + \sigma_{geom}^2 \tag{4.1}$$

with

$$\text{with } \sigma_{geom}^2 = \left(\frac{\sigma_1 r_2}{r_2 - r_1}\right)^2 + \left(\frac{\sigma_2 r_1}{r_2 - r_1}\right)^2 \text{ and } \sigma_{MS}^2 = \sum_{j=1}^{n_{scatt}} (R_j \Delta\Theta_j)^2 \tag{4.2}$$

with σ_1 and σ_2 the intrinsic resolution in the measurement layers and $\Delta\Theta_j \simeq \frac{0.0136}{p_T^{beam}[\text{GeV/c}]}\sqrt{\frac{\Delta x}{X_0}}[1 + 0.039 \cdot ln(\Delta X_j / X_0)]$ the average multiple scattering angle of a particle with momentum p_T^{beam} traversing through the material of thickness ΔX_j (expressed in fractions of a radiation length X_0) located at radius R_j and n_{scatt} the number of layers in front of the last detection element. The IP resolution is often parameterized by $\sigma_{d_0}^2 = \sigma_{asympt}^2 + \left(\frac{\sigma_{MS}}{p_T}\right)^2$ with the transverse moment p_T in GeV/c (a fitted value derived from detector operation) – p_T is often also expressed as $p_\perp$.

All the above considerations lead to the following design goals:

- low mass for beam pipe and vertex detector, including cables and support structures to minimize Coulomb scattering (e.g. all electronic components were placed outside the detector volume). This is especially true in front of the very first measurement layer: keeping $\Delta X_j / X_0$ and Θ_j small results in a small σ_{MS}
- placement of the first detection layer as close as possible to the primary interaction point to minimize extrapolation error, thus maximizing impact parameter resolution: r_1 small!
- largest possible radius for the outer measurement layer: r_2 large
- high intrinsic detector resolution, thus silicon sensors with small pitch and analogue readout for hit interpolation in-between strips: σ_1 and σ_2 small
- take alignment into account from the very beginning, thus overlap sensors to allow extrapolation of exact position with tracks crossing overlapped modules
- establish good algorithms for alignment, pattern recognition and vertex identification in the early stage

The initial detector performed superbly in lifetime measurements of the τ lepton and b-flavoured hadrons, even with only two layers and single-sided sensors allowing only $R\phi$ information. The initial aluminium beam pipe outer radius located at $R = 7.8\,\text{cm}$ was consequently exchanged with a beryllium one at smaller radius

[3]The track does *not* pass through the primary vertex.

Table 4.1 Impact parameter resolution between 1989 and 2000. Initially the IP resolution was improved, then in 1994 the z-coordinate[4] joins in and in 1997 the polar angle coverage increases from 45° to 25°. Clearly all the design criteria were rigorously followed

Year	Coord.	$\sigma^2_{IP}(\mu m^2)$	Angle (°)	Layer
1989	$R\Phi$	$80^2 + (120/p_T)^2$	45	2
1990	$R\Phi$	$24^2 + (69/p_T)^2$	45	3
1994	$R\Phi$	$20^2 + (65/p_T)^2$	45	3
	Rz	$39^2 + (71/p_T)^2$	45	3
1997	$R\Phi$	$28^2 + (71/p_T)^2$	25	3
	Rz	$34^2 + (69/p_T)^2$	25	3
1997	VFT[a]		10.5	1

[a] VFT: The **V**ery **F**orward **T**racker includes pixel and Ministrips

of $R \sim 5.5$ cm allowing for a third silicon layer at $R = 6.3$ cm in 1990. The IP resolution improved from $\sigma_{IP_\phi}(1989) = \sqrt{80^2 + (120/p_T)^2}\,\mu$m to $\sigma_{IP_\phi}(1991) = \sqrt{24^2 + (69/p_T)^2}\,\mu$m where the latter corresponds to $d_0 = 21\,\mu$m for a 45 GeV/c track from a $Z \to \mu^+\mu^-$ decay corresponding to a point precision measurement of 8 μm. The full history of measured σ_{d_0} from 1989 to 2000 is listed in Table 4.1.

4.2 The DELPHI Microvertex Detector 1996/1997

The basic necessary design features are presented in the previous section. Now what are the ingredients to cope with energies of 195 GeV/c and above for LEP2? The main design optimizations for the 1996 upgrade were driven by physics expectations like four fermion processes which ask for a large angular coverage. Also Higgs boson and supersymmetric particle searches at LEP2 were relying strongly on precise tagging of b quarks down to low polar angles. This is important to reduce W^+W^- process[5] background. Some physics examples are described in Sect. 4.5. The MVD schematic is displayed in Fig. 4.4 while the full blossomed beauty of the microvertex detector is shown in Fig. 4.5. The main goal is to increase hermeticity and expand the b-tagging capabilities. These goals are met by extending the barrel region to achieve b-tagging capability down to 25° and active sensor area down to 10.5° with the additional pixel and Ministrip sensors. The three layers serve also the purpose to solve possible track ambiguities. The entire structure is 85 cm long and has 1.5 m^2 of active silicon sensors. An outer barrel module from 1994 can be seen in Fig. 4.6. The corresponding hybrids are displayed in Fig. 4.7.

[4] Rz IP resolution is given for perpendicular tracks, since the point resolution (inclined tracks) as well as the amount of material strongly depends on Θ for the z-coordinate

[5] The W^+W^- production cross-section close to threshold energies is a few picobarns only, thus the expected event rate is relatively small and a high detection efficiency is necessary.

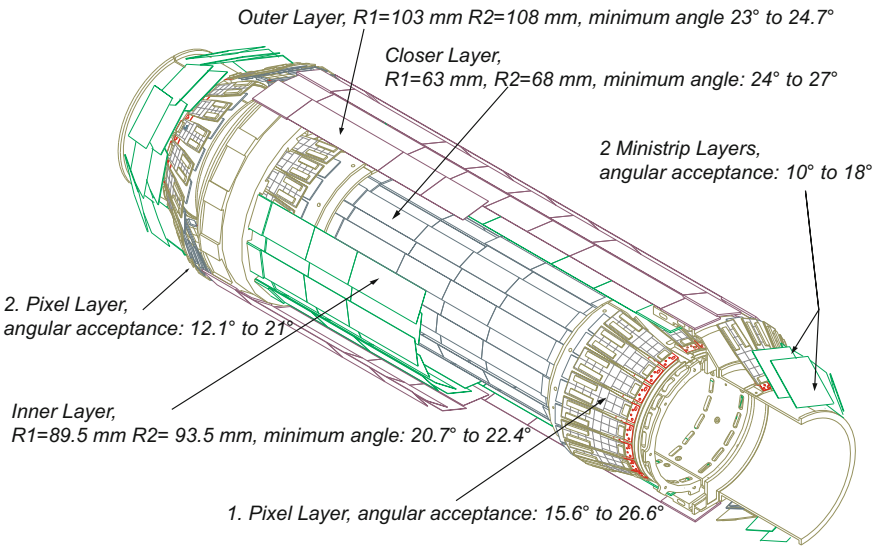


Fig. 4.4 The DELPHI microvertex detector [319]

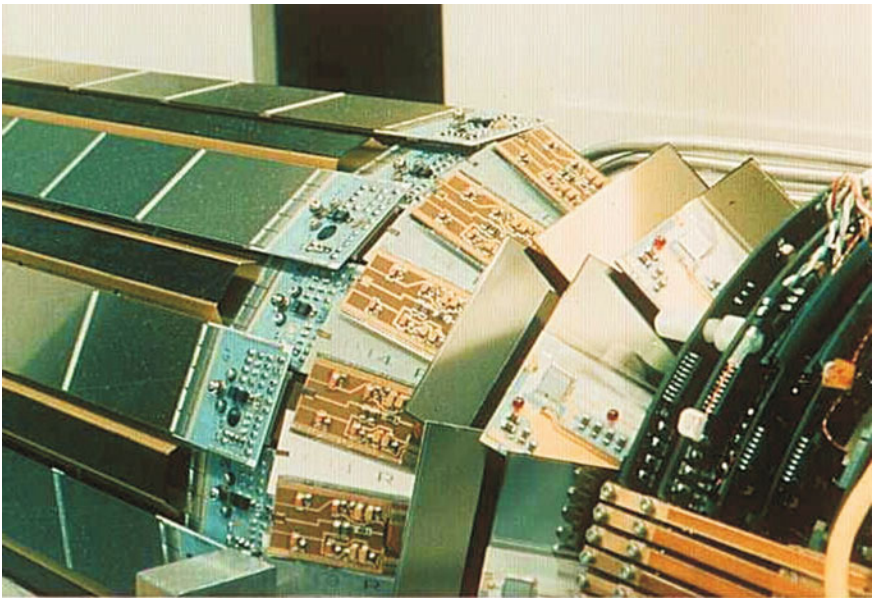


Fig. 4.5 The very forward part of the DELPHI tracker. The photo shows the forward pixel and Ministrip detector. At the time of design and construction these innovative parts enabled tracking down to $\Theta = 10.5°$. [Courtesy of DELPHI, CERN]

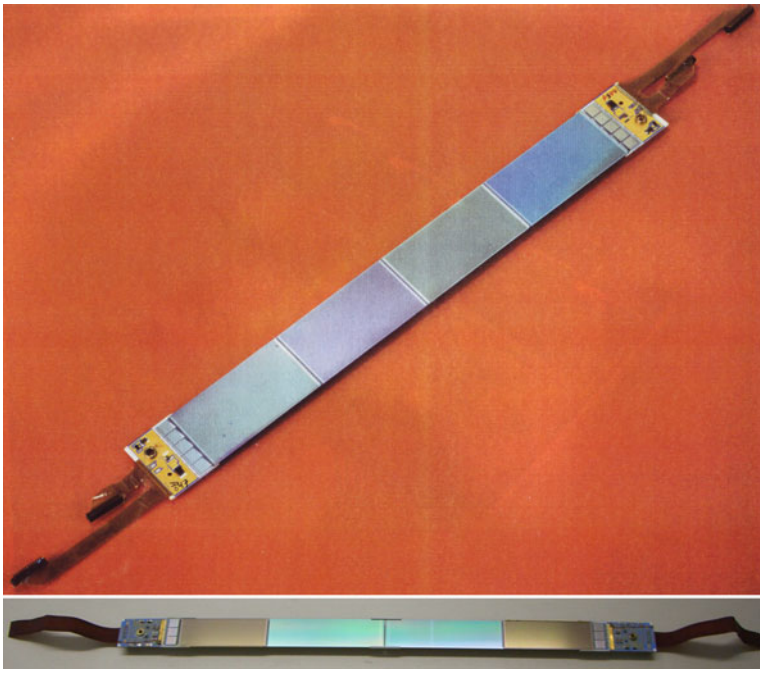


Fig. 4.6 A DELPHI inner and outer module. Each hybrid reads out two detectors with the daisy-chained strips connected to each other and to the amplifiers by wire-bonding. This assembly is chosen to carefully situate the electronics outside the active volume, thereby minimizing the material budget and also minimizing multiple scattering. The outer detector module contains five chips with a total of 640 strips on each hybrid side, while the inner detector module being narrower contains only three chips with 384 strips per side

The DELPHI MVD is divided into closer at $R = 6.6\,\text{cm}$, inner at $R = 9.2\,\text{cm}$ and outer layer at $R = 10.6\,\text{cm}$ tightly fitting in-between the beam pipe and inner drift chamber. The *smallness* of the detector allows for many fancy solutions and very distinct optimizations. The basic concept is the usage of double-sided sensors in the closer and inner layers and back-to-back modules in the outer layers where the Coulomb scattering is less critical. The hybrids reside at the end, equipped with MX6 chips serving two sensors each in the closer layer and two or four sensors in the inner layer. In the outer layer one hybrid serves four sensors with the newly developed Triplex chip, optimized for larger capacitances (a short discussion on these chips can be found in Sect. 1.5 on p. 40f and Sect. 1.10 on p. 96f). All hybrids are double-sided, having chips on both sides. In the $R\Phi$ plane the point resolution is around 8 μm and between 10 and 25 μm in the Rz plane, varying for different track inclinations. Throughout the tracker, great emphasis was placed on the overlap of sensitive silicon within each layer to allow for self-alignment procedure – a staggered design, see Fig. 4.3. The closer layer is 36 cm long, while the inner and outer barrel counts 55.5 and 55.9 cm in length. The bulkhead is constructed out of aluminium but all internal

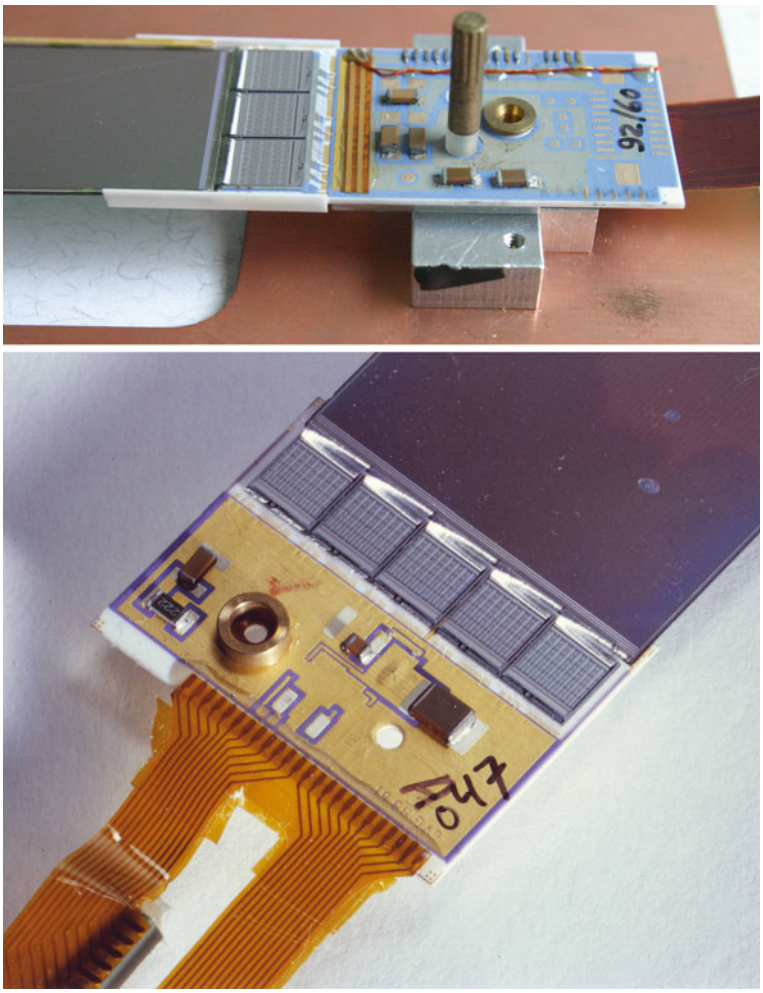


Fig. 4.7 DELPHI hybrids. Voltages and electrical signals, like clock and readout data, travel via the brown thin flex Kapton cable. Some passive components share the hybrids' space with the MX chips serving 128 channels each. The silverish connection between hybrid and sensor shows 384 (640) micro-wires. The "massive" round structure is the high-precision hole for precision mounting on the bulkhead. Notice the hand-made labels "92/60" and "A047" – the total number of modules was manageable. Without the electronics **V**ery **L**arge **S**cale **I**ntegration VLSI of the Microplex (MX and Triplex) chips any realization of a silicon vertex detector would have been impossible

structures are lightweight Kevlar plus carbon fibre only. A minimum of water cooling runs only in the bulkhead volume.

One has to value the fact that the DELPHI vertex detector is the first one using forward structures. These are necessary to improve track extrapolation towards the forward **R**ing **I**magine ***Č***erenkov **C**ounter RICH, to improve **P**article **Id**entification

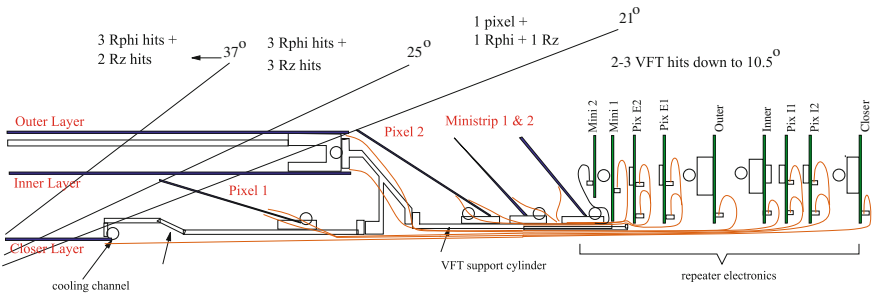


Fig. 4.8 Cross-section of the DELPHI tracker – a novelty for a collider vertex detector, a forward part. DELPHI pioneered the field in two ways: (*1*) forward tracking and (*2*) usage of hybrid pixel sensors. To cope with the tight space constraints, quadratic strip sensors were mounted back-to-back with the electronics hybrid glued on top of the sensors on both sides. A single type of sensor geometry was used – the Ministrip sensor, a $53 \times 53\,\mathrm{cm}^2$ quadratic single-sided sensor. Towards the centre DELPHI uses the new pixel technology to establish a good pattern recognition with an intrinsic 2D readout. Thin, long pixel modules just fit in this crowded region. The pixel modules reach fully into the barrel strip region [319]

PID. Schematic of Fig. 4.8 expresses in detail which detector component is active in the different polar angle regions.

The detector also features the first usage of *silicon hybrid pixels* (see Sect. 1.12.1 on p. 107) and rectangular strip sensors, where the hybrid is attached onto the sensor – Ministrips. These two new sensor types are elaborated upon here.

DELPHI Ministrip

The Ministrip detector consists of 48 modules arranged in two layers on the MVD bulkheads, the so-called "crowns". Every detector is quadratic $5.3 \times 5.3\,\mathrm{cm}^2$; therefore, two single-sided sensors can be mounted back-to-back with a 90° angle to establish an intrinsic 2D readout. A Ministrip module is displayed in Fig. 4.10. The hybrids with the two MX6 readout chips are glued directly onto the sensor serving 256 strips. The strips are arranged with a stereo angle of 2°. Due to Coulomb scattering and due to unavoidable material (e.g. hybrids from the barrel modules) the resolution is anyhow limited to about 100 μm and the pitch can be relaxed to 200 μm with one intermediate strip. The stereo angle helps to resolve track ambiguities in the crowded forward direction. Modules of Ministrip layer 1 are reversely[6] placed with respect to layer 2 with a resulting effective angle of 4° between strips. Two particle tracks hitting the same coordinates in one detector plane now hit, by construction, different coordinates in the second "tilted" plane.

[6]The "horizontal" planes face each other, therefore strips are oriented ±2°.

DELPHI PIXEL

The pixel detector is formed out of 152 modules, the component layout is shown in Fig. 4.9 and a photo of a pixel module can be seen in Fig. 4.10. A pixel intrinsically has a 2D readout and is therefore predestined to have a marvellous pattern recognition capability and reduces any ambiguity. The modules are inclined as seen in Fig. 4.8. Each pixel module area contains 8064 pixels with a size of $320 \times 320\,\mu m^2$. Unlike the strip sensors (described in the next section) the pixel sensors are DC coupled and one module is read out by 16 separate chips (see Sect. 1.11), which have the same size as the pixels silicon sensors. The chips are aligned directly on top of the sensors and are bump bonded. A multi-layer Kapton is then glued on top of the chips to distribute control signals, power supply lines and signal lines.

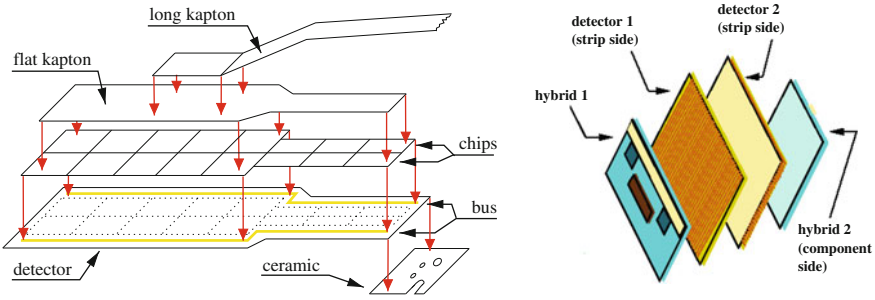


Fig. 4.9 A DELPHI Ministrip and pixel module – layout. 16 chips of two different varieties with 24×42 or 16×24 channels are flip-chipped in a single module to cover 8074 pixels in total. Two buses connect 8 chips each. The *right* picture shows the back-to-back arrangement of the Ministrip sensors and their electronics placement. [319]

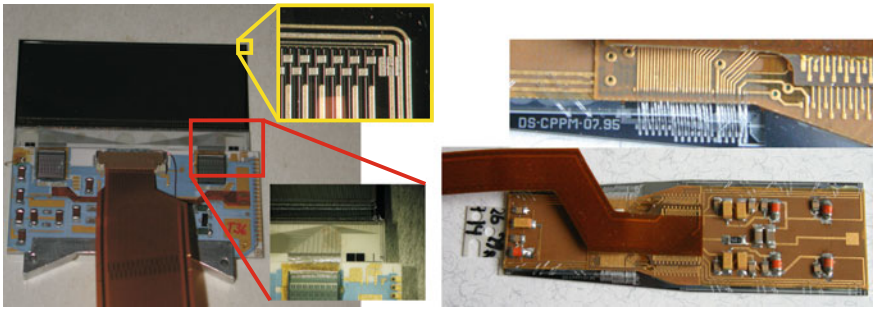


Fig. 4.10 A DELPHI Ministrip and pixel module. The *left* picture shows one side of a Ministrip sensor module. The hybrid to sensor wire-bonding happened in the middle of the sensor. The sensor itself is biased via FOXFET. The FOXFET ring layout is shown in the exploded view of the sensor corner, the guard ring was connected to ground. A second zoom shows the hybrid edge including the chip edge. Two chips are sufficient for the 256 strips. The *right* picture shows a complete pixel module. The sensor-chip sandwich (bump bonded) is covered by a Kapton flex structure carrying the necessary passive components and the bus lines. The exploded view shows the wire-bonding from pixel chip bus lines to Kapton bus lines [182]

4.3 The Silicon Sensors of the DELPHI Microvertex Detector MVD

In the last stage of the DELPHI tracker a variety of sensors were used to first optimize in all areas and second to reuse as many sensors as possible from the older detector versions. The different sensor types, sizes and technological choices are listed in Table 4.2.

The commonalities for all the DELPHI strip sensors are the use of the *planar* technology where the implants for the *pn*-junctions were either formed by diffusion or by ion implantation. As a novelty, integrated coupling capacitors were developed for the DELPHI Microvertex detector – AC-coupled readout [53]. Different from the NA11 sensors, a dielectric layer, around 200 nm thick, was introduced between the implant strip and the aluminium readout electrode. This reduced direct currents into the readout amplifier to zero. In earlier years the different strip currents caused baseline variations of the readout and therefore required an ADC of large dynamic range. With the integrated capacitors, charges arriving at the implants are capacitively coupled to the aluminium strip. The concept is described in Sect. 1.6.3. Unfortunately a short of an integrated coupling capacitor introduces a new vulnerability to the system – pinholes. A short between implant and aluminium first allows current to flow freely to the charge amplifier and secondly disturbs the potential between chip channels, resting at a different potential than the implants. On the *n*-side of some sensors, one of these shorts resulted in a loss of about 50 neighbouring channels. The standard solution is the removal of the wire-bond but in the case of a pinhole[7] occurring after[8] installation, this is not a viable option. Typical values of the coupling capacitances for e.g. 7 μm wide strips are 12–15 pF/cm, sufficiently high in comparison with the strip to backplane and the inter-strip capacitance that are on the order of 0.2 pF/cm and 1 – 1.5 pF/cm, respectively. As mentioned earlier, $\frac{C_{Coupling}}{C_{int}+C_{Back}}$ has to be large, coupling to the readout has to dominate parasitic charge coupling. The later sensors in 1996 already used a combination of SiO_2 and Si_3N_4 layers to increase capacitance at the same dielectric thickness. The use of additional nitride also enhanced the robustness vs. pinhole creation. As a direct consequence of the additional insulator layer the necessary potential on the implants could no longer be applied via the chip channel inputs. There were two novel bias concepts applied namely, FOXFET and the use of polysilicon resistors (see also Sect. 1.3). Running at room temperature, the polysilicon resistor values must be large, not to influence noise[9] $\left(\text{ENC} \sim \sqrt{\frac{k_B T}{R}}\right)$. The specification requested resistances of 5 MΩ or higher. Finally 4 – 20 MΩ resistors were accepted while most resistors reside well above 20 MΩ.

[7] Noise distributions around pinholes looked like a high mountain range and the occurrence was called “Mt. Fuji”.

[8] Fortunately, there were not too many such cases.

[9] Thermal noise from parallel resistances is proportional to the shaping time, thus crucial for LEP operations. In addition the modules daisy-chain up to four sensors. The parallel arrangement of R_{poly} decreases effective resistance (see Sect. 1.5).

Table 4.2 Different silicon sensor types for different layers

	Closer	Inner	Outer	Ministrip	Pixel
Radius (cm)	6.6	9.2	10.6		
Number of modules	24	20	24	48	152
Detectors/module	4 ds	4 ds + 4 ss	Rϕ : 8, z : 8	2ss	1ss
Sensitive area (cm^2)	292	208	103	324	378
Channels/module	1536	2560	2560	512	8064
Coverage $\pm\Theta$	25	21	23	10	12.2
Support	Kevlar + carbon	Kevlar + carbon	Kevlar + carbon	Al	Al
Angle to z-axis	0	0	0	49	12 and 32
Length	6.07; 7.91	5.75;6	6	5.3	6.9
Width	2.08	3.35	3.35	5.3	1.7–2.2
Readout pitch (μm)	Rϕ 50	Rϕ 50	Rϕ 50	200	330
	z 49.5, 99, 150	z 42, 84	z 44, 88, 176		
Intermediate strips	Rϕ 1 z0	Rϕ 1 z0	Rϕ 1 z 0,1	1	0
Biasing	R_{poly}	FOXFET/R_{poly}	FOXFET/R_{poly}	FOXFET	
n-side isolation	Field plate	p^+	–	–	–
Operating voltage	60	60	60–95	50–60	60
Readout chip	MX6	MX6	Triplex	MX6	SP8
RO channels/module	2 × 384	2 × 640	2 × 640	8064	300
Power/chip (W)	0.2	0.2	0.2	0.2	0.017

ds double-sided; *ss* single-sided. More detailed tables exist in [319]

The guard rings on all DELPHI sensors were connected in the *current drain configuration* and were bonded to ground, defining the electric potential, thus electric field, at the sensor borders. For this configuration the guard rings must surround the strip area closely, but are still located outside the bias ring.

The different choices of pitches, implant widths and aluminium widths are driven by point resolution needs ranging from 42 to 200 μm with intermediate strips for charge sharing to achieve a good resolution while keeping the number of readout channels low. Considerations, like effects of magnetic field, statistical fluctuation in the energy loss, noise performance of the electronic readout, also influenced the initial choice of thickness and pitches. The full system is small enough to allow the luxury of different parameter choices for different detector regions. The numbers are listed in Table 4.2. Along the z-axis the track incident angle increases, therefore the particle traversing paths are longer ($\cos\Theta$). For the $R\phi$ coordinates, this simply increases signal height, but for Rz coordinates with strips orthogonal to the particle path this would lead to many strips being hit, a many-strip cluster event. This is not particularly beneficial for the resolution. The signal charges are distributed over several strips thus individual strip signals are possibly too low to be significantly higher than the noise level. This problem can be circumvented by increasing the pitch along the z-axis. As an example in the outer layer, the four-sensor-long half-modules start with 50 μm pitch in the centre increasing to 100 μm pitch in the second sensor and ending with 200 μm pitch for the last two sensors.

With hybrids situated at the end of the modules (see Fig. 4.6) the strips running orthogonal to the beam axis need a routing to the readout, which is realized via a second metal layer. The readout strips orthogonal to the beam axis are connected by vias at defined locations to perpendicular strips in a second metal layer. The two metal layers are separated by either a 4 – 5 μm thick layer of SiO_2 or of polyimide. The photo in Fig. 4.11 and the schematic in Fig. 4.12 allows us to understand the system quite well. This is true for the double-sided sensors as well as for the single-sided Rz sensors in the outer layer. An aluminium thickness of 1 – 2 μm guarantees a high conductivity, thus low serial noise. Unfortunately and unexpectedly in the case of the outer layer, the metal-to-metal capacitance was around 1 pF/cm and a charge loss up to 50% had to be endured [67, 137]. Operation was still possible without compromise because the charge loss was within Rz sensors where the additional charge from the longer traversing particle path ($\cos\Theta$) compensated for the effect.

In the 1994 and 1996 upgrade, double-sided sensors were introduced. In Sect. 1.6.3 on p. 50 it is shown that on the n-side special precautions are necessary to guarantee strip isolation. The standard mechanism of p^+-stops[10] was used – visible in the photo on the left of Fig. 4.11. In addition DELPHI used the existing aluminium readout strips above the n^+ implants in a novel configuration. The aluminium width is wider than the width of the underlying implants. Furthermore, an appropriate potential was applied to the electrodes (negative with respect to the bulk) driving electrons away and resulting in surface depletion – the readout strips work as a gate-like electrode.

[10] p^+-stops was the common configuration; the p^+-spray technique was developed later.

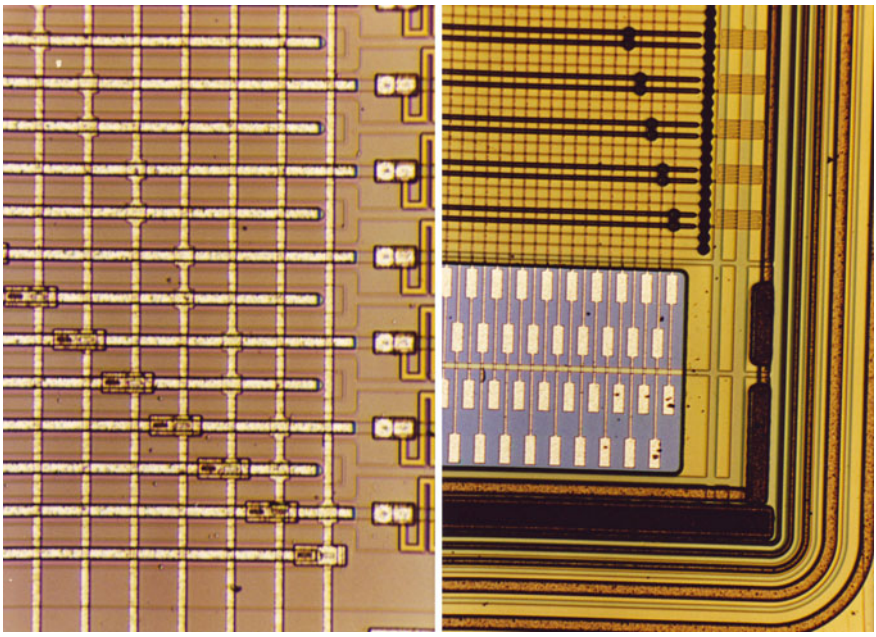


Fig. 4.11 The *left* photo shows the microscopic view of the *n*-side of a double-sided sensor. The *right* picture shows a single-sided *Rz* sensor. The perpendicular strip arrangement with the contact vias can be seen on both pictures. Both sensors are using the double metal routing to connect the strips to the readout channels. The meander structures, the polysilicon resistors, connect the bias voltage to the implants. The polysilicon length and narrow shape defines the high resistance. On the *left* picture the *n*-side structure, the p^+-stop structures surrounding every second implant are clearly visible. They are responsible for the strip-to-strip isolation. The sensor on the *right* is also interesting; always two metal strips are connected together with two non-metallized intermediate strips (implants only) in-between. With a 50 μm implant pitch, this gives 200 μm readout pitch with good charge sharing. The intermediate strips were connected to the bias ring via bias resistors on the other end of the sensor to guarantee a uniform potential on all implants. For the sensor on the *left* every single strip is connected to the readout. This gives a bit of an impression of the variety of sensors in the detector

With a voltage difference of 10 V and an overhang of 4 – 5 μm, inter-strip resistances exceeding 100 MΩ were achieved - the field plate configuration.

On the other hand, a metal overhang causes possibly strong fringe fields on the edges and makes the detector potentially vulnerable to micro-discharges or the so-called "popcorn" noise. Extensive measurements were done to avoid any surprises, the onset of micro-discharge was found to be well above 90 V, thus well above the operation voltage.

Figure 4.12 presents the ultimate technology mix, a double-sided sensor with integrated coupling capacitors, serving also as field plate on the *n*-side, bias voltage supplied via polysilicon resistors and finally a double metal layer to allow perpendicular strip routing. Finally, DELPHI used the "*flip chip*" sensor module configuration. With two sensors daisy-chained, the second one is "flipped", or in other words, the *p*

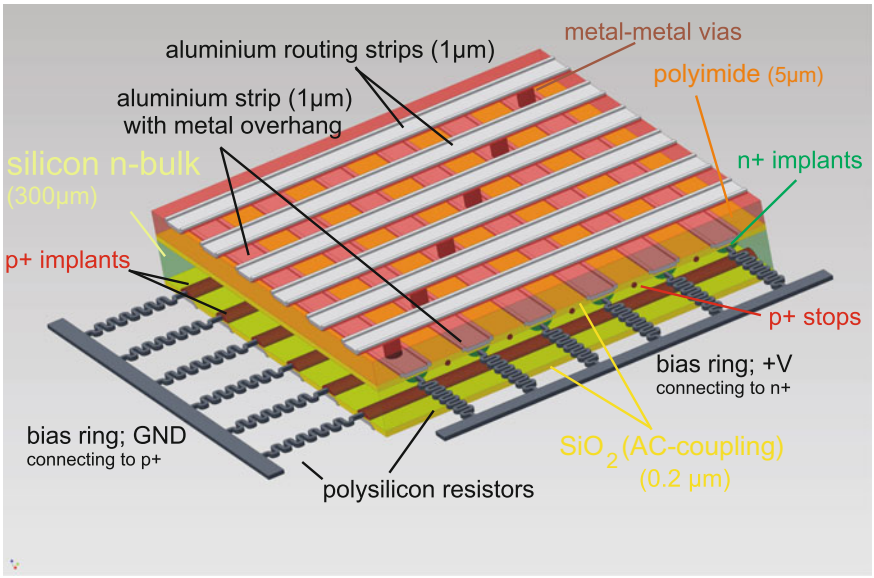


Fig. 4.12 The DELPHI double-sided, double metal sensor scheme. The sensors contain novel integrated coupling capacitors, the bias is applied via high-resistive polysilicon resistors, the *n*-side strips are routed via a second metal routing layer and the innovative field plate configuration guarantees 100 MΩ *n* strip isolation

strips (up) of sensor A are connected to the *n* strips (up) of sensor B. The configuration is shown in Fig. 4.13 and visible in Fig. 4.6, the different reflecting colours of the sensors in the module indicate the sensor side. As a first improvement, the different noise levels of *p*- and *n*-side are balanced.[11] Secondly, since *p*-side collects holes, while *n*-side collects electrons, the signals come in different polarity and therefore the polarity tells on which sensor the particle traversed, resolving ambiguities. The signals are also displayed in Fig. 4.13. This mix of technologies is applied to achieve the best possible point and impact parameter resolution and a very good pattern recognition. The effort finally paid off, when tracks like the ones shown in Fig. 4.16 could be resolved to figures like Fig. 4.19 to allow high precision particle physics in a collider experiment.

Hans Dijkstra, the DELPHI Vertex Detector coordinator commented on the time of design and construction with the words:

> These times were crazy, a small group of people had to deal on every corner with a new technology – new VLSI chips, beryllium hybrids, double-sided sensors with integrated coupling capacitances, high precision assembly, etc. The detector was like a huge puzzle for grown-ups. We had a lot of stress and fun.

[11] *n*-side is more noisy, e.g. due to the higher load capacitance.

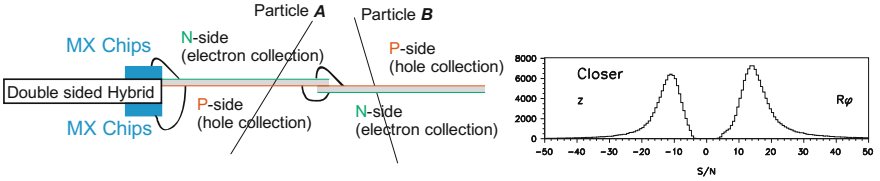


Fig. 4.13 Signals from flip module assembly. With the use of dipolar readout electronics, the signal polarity distinguishes if the particle traverses through sensor A or B of the module. Furthermore, with this connection scheme the different noise levels of n- and p-side are balanced

4.4 Implementation of Silicon Labs in Universities to Build a Large Device

In the LEP era, detectors became "large" and complicated, too large to be constructed at a single place. Several tasks were distributed to a number of institutes scattered over Europe. Assembly, connection and testing (quality assurance) facilities were established at many places. The right part of Fig. 4.14 shows a 3D measurement machine with a completed module under geometric survey while the left part shows the assembly jig where all modules were assembled and manually aligned. The full module construction, sensor testing, assembly (alignment and gluing), micro-wire bonding and final testing, was done at the universities. At this time the initial sensor testing was done manually strip by strip, while automatization was introduced in a later step. One famous finding of the sensor quality control is shown in Sect. 1.14. Modules were tested with laser, test pulses, β source ^{90}Sr and γ from ^{241}Am source for calibration[12] of ADC counts to electrons.

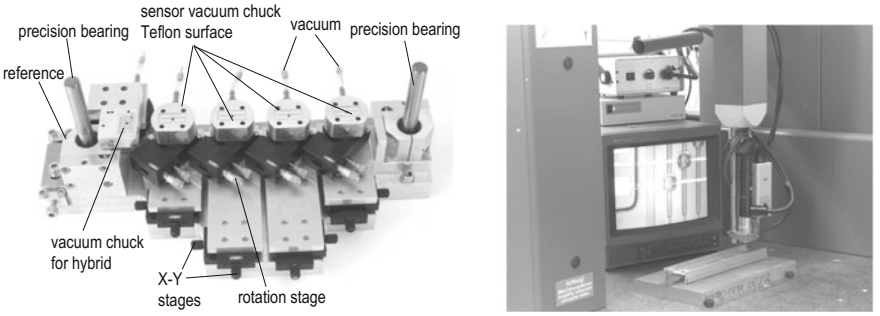


Fig. 4.14 The DELPHI assembly jig and one silicon module at final survey. The homemade assembly jig of the DELPHI outer layer modules is presented on the *left*. Hybrid and sensors are placed, fixed by vacuum and then positioned with the help of the 3D coordinate measurement machine. This procedure is done for the $R\Phi$ and for the Rz side separately. Then both halves were connected together with the help of the high-precision bearings. All these steps were fully manual. In a final step all modules were extensively surveyed [170]

[12] Triggering on γs is not possible, but with a strong source and an integration time of 2 μs a readout frequency of about 3 Hz without trigger was achieved – enough for calibration.

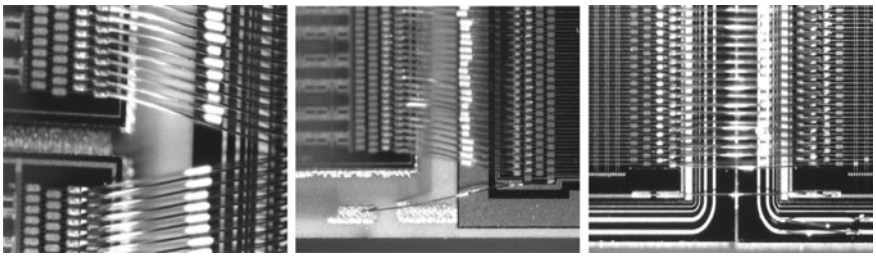


Fig. 4.15 DELPHI module bonding. The photos give an impression of the wire-bonding. The pitch is around 50 μm. The mass strip bonding was done automatically, while all special bonds, e.g. bias and guard bonds, were placed manually

Figure 4.15 shows the fine wire-bonding of the DELPHI sensors and hybrids. As an example, one DELPHI veteran from CERN, two Ph.D. students and three undergraduates constructed 16 DELPHI outer barrel half-modules with each consisting of double-sided readout hybrids plus four+four sensors mounted back-to-back. Furthermore, all buses on the Kapton cables of the pixel modules were tested and all the wire-bonding for the pixels was done by students. These activities were the seeds to establish expert silicon laboratories in university institutes.

4.5 Physics with the DELPHI Microvertex Detector

This chapter lists some examples of the impact of the microvertex detectors on elementary particle physics. It is by no means meant to be exhaustive. A good description can be found in [84]. The main improvements were achieved in the

- lifetime measurements of heavy quarks and the τ lepton with unprecedented precision
- the tagging capability of heavy quarks (flavour tagging)

In addition the overall transverse momentum resolution σ_{pT} improved dramatically. The Higgs boson search during the LEP2 run was heavily supported by the vertex b-tagging capability, since a Higgs in the accessible mass range would have been coupled predominantly to $b\bar{b}$. The absence of an excess of $b\bar{b}$ events therefore established a firm lower limit of the Higgs mass.

The basic mechanism of heavy quark flavour tagging is not very different to the NA11 example shown in Fig. 3.3.

Only track points in the individual sensors, in the different barrel planes, are available. These points have to be extrapolated to tracks and good algorithms are needed

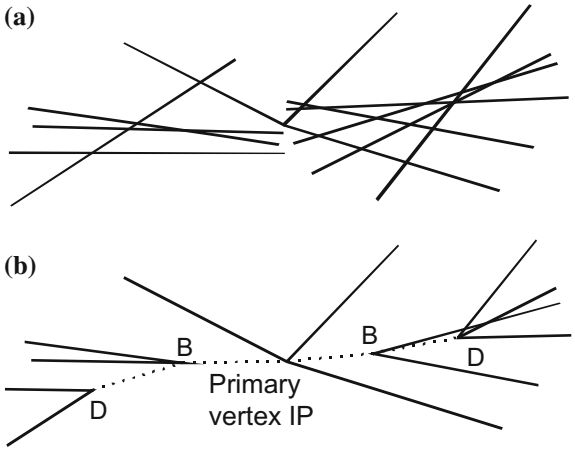


Fig. 4.16 A $b\bar{b}$ event, tracks and physics reconstruction. Two B mesons coming from the primary vertex are each decaying into a D meson, which shortly after decays into K mesons (b → c → s). A small number of tracks has to be resolved in this "simple" scenario. One should remember that tracks in silicon or also drift chambers are not continuous tracks like in bubble chambers. (**a**) Shows the reconstructed tracks using the space point coordinate information of the silicon detectors. (**b**) Shows the fully reconstructed $b\bar{b}$ event showing the two B mesons and their decay to D mesons, where also the different vertices are reconstructed [268]

to understand where the tracks start. In Fig. 4.16a $b\bar{b}$ event exemplarily shows what the real particle paths and the initially reconstructed paths from the space points look like. The situation becomes even more ambiguous once tracking imperfections and the effects of multiple scattering are introduced. A good point resolution is necessary, because the first space points (sensor locations) are several centimetres away, while the primary to secondary vertices are in millimetre regime and second to tertiary distances even less than a millimetre. However, the number of tracks not originating from the primary vertex is significant and even without full reconstruction the impact parameter d_0 delivers a good understanding of the process. Full reconstruction of all vertices from cascading D decays is therefore extremely difficult. In Fig. 4.19, a DELPHI b quark event is displayed.

Measure the Lifetime of Heavy Quarks and τ Leptons

Without quark flavour tagging the lifetime measurement of heavy quarks relies on the lepton tag, but full hadronic decays are unresolvable though having a larger branching decay ratio. Thus a larger statistical sample would increase precision. With the new vertex detectors tagging and flight path measurements the situation largely improved. The plots in Fig. 4.18 give profound evidence of this improvement. Actually two lifetime measurement methods were used.

The Impact Parameter Method

Full vertex reconstruction is not always possible. At the production threshold $b\bar{b}$ decay isotropically, as do their decay products. There is a large combinatorial back-

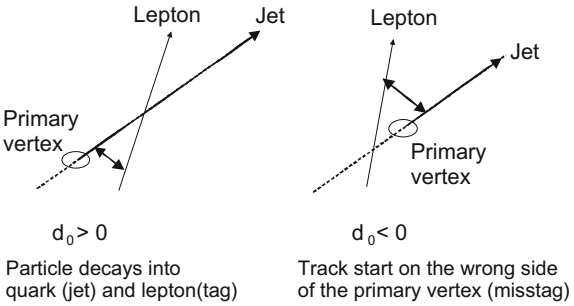


Fig. 4.17 Signed impact parameters. Particle paths not passing through the primary vertex definitively do not originate there. The impact parameter d_0 is the closest distance between track and primary vertex. The presence of a positive d_0 larger than the detector resolution therefore defines a secondary vertex and thus a lifetime tag. Negative d_0 results from detector resolution effects or a miss defined primary vertex. These signals can be used as control samples but they do not have physical significance

ground. Around 1995 the world sample of fully reconstructed B meson decays amounted to $<$2000. The additional b-tag via the impact parameter method (see also Fig. 4.17) without full reconstruction increases statistics and thus improves lifetime analysis. As described above, the impact parameter d_0 is defined by the shortest distance between a reconstructed track and the primary vertex; meaning $d_0 > 0$ defines the existence of a secondary vertex. At first order the average impact parameter d_0 is proportional to the lifetime τ_0 of the decaying particle: $< d_0 >=< \beta\gamma c\tau > \propto c\tau_0$. The first physics result obtained and published with utilizing the silicon vertex detector is the lifetime measurement of the τ lepton [7].

The Lifetime Measurement or Decay Length

The direct way to determine the lifetime, e.g. of the τ lepton, is the full reconstruction. This includes the determination of the decay path and decay length, which presents a direct lifetime measurement at known energies. Three-dimensional track points facilitate the reconstruction. Looking on the right part of Fig. 4.18, the significant improvements in lifetime measurement with the start of LEP and its silicon microvertex detectors become obvious.

Heavy Quark Tagging

Spectroscopy of particles naturally profits from a pure particle sample to obtain precision values about mass, lifetime and branching ratio. The left plot in Fig. 4.18 demonstrates the fundamental reduction of background in the $D^+ \to K\pi\pi$ decay. Flavour tagging without full reconstruction relies on the fact that the primary quarks (u, d, s) are light and have a very small probability for a secondary vertex

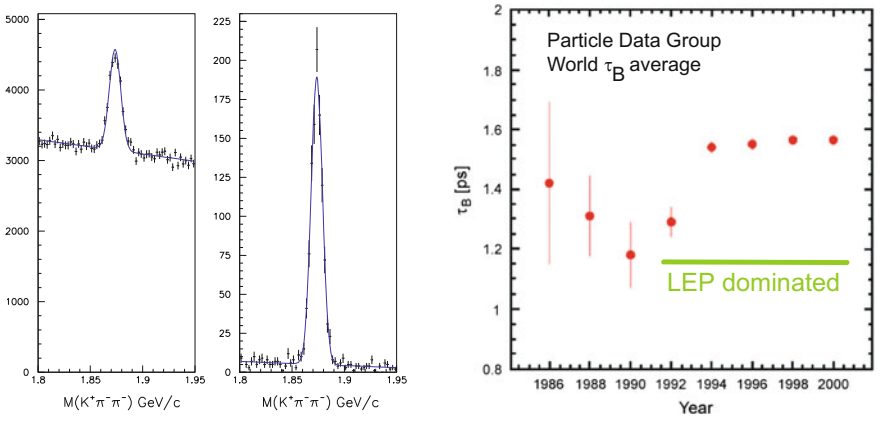


Fig. 4.18 Examples of physics potential improvement with the introduction of high-precision silicon vertex detectors in the LEP era. Plot (*1*) shows the $D^{\pm} \to K\pi\pi$ peak without and with 7σ vertex cut. The background suppression is impressive [148]. Plot (*2*) proves the clear domination of *b* lifetime measurement with the start of using vertex detectors at LEP. The physics results clearly support the necessity of silicon vertex detectors [68]

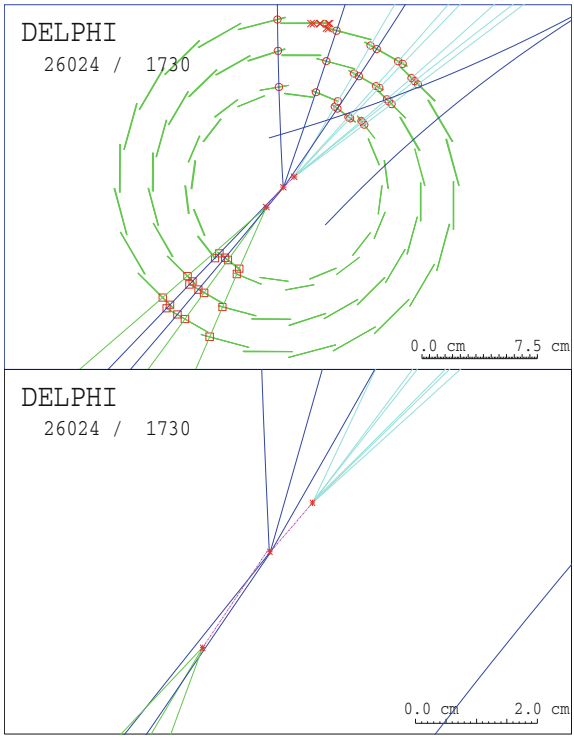


Fig. 4.19 A *b* event with the DELPHI vertex detector. The *upper* part of the figure displays the different track points of the three detector planes together with the reconstructed tracks. The *lower* part, an exploded view, shows the capability of track interpolation down to the primary vertex and to resolve the secondary *b*-vertex. This plot fully expresses the power of a microvertex detector to tag *b* quarks and determine the flight path of a short lived particle. [Courtesy of DELPHI, CERN]

close to the primary one. The decay products from heavy quarks with lifetimes of 10^{-13}–10^{-12} s originate from a secondary vertex. Furthermore c quarks decay directly into light quarks while b quarks decay in a cascade via c quarks to light quarks, thus producing twice as many vertices than pure c events. At the second step knowing the spectroscopic parameters, they can be used to identify/tag the decay particles which is especially interesting for cross-section measurements like the determination of $F(Z \rightarrow b\bar{b})$ which was limited beforehand by poor statistics. Another example is the study of jets originating from gluon radiation ($b\bar{b}g$) from b quarks by tagging the two jets from two b quarks and fixing the third jet to the radiated gluon. Figure 4.19 shows a DELPHI $b\bar{b}$ event in the vertex detector with an exploded view of the primary interaction point.

Chapter 5
CDF: The World's Largest Silicon Detector in the 20th Century; the First Silicon Detector at a Hadron Collider

The **C**ollider **D**etector at **F**ermilab CDF is situated in Batavia, Illinois near Chicago. CDF is one of the two multi purpose detectors[1] at the TEVATRON proton–antiproton collider at the **F**ermi **N**ational **A**ccelerator **L**aboratories FNAL. Figure 5.1 shows the CDF detector configuration for RUN II operating since 2000—a conventional multi purpose hadron collider detector. The beryllium beam pipe is naturally located in the centre followed by the inner tracking system consisting of three silicon sub detectors, Layer 00, SVX II and ISL, surrounded by the open cell drift chamber. The tracking volume rests completely inside the 1.41 T solenoid. The calorimeter, divided into the electromagnetic and hadronic part, is located outside of the solenoid. The drift chambers for muon detection are situated outside of the solenoid return yoke. The tracking system was largely increased for the RUN II upgrade, the silicon barrel detector's length increased to almost 2 m to cover a pseudorapidity[2] $|\eta \leq 2|$ without any endcap structure. This chapter describes solely the tracking detectors of CDF, starting with the historical evolution and focusing on the RUN II configuration.

5.1 Historical Evolution of the CDF Vertex Detector

Already at the time of NA11 and in the early design phases of DELPHI, the CDF collaboration had preliminary ideas on the use of silicon as a luminosity monitor inside the vacuum beam pipe or as forward detectors ([28, 214] 1983). The first CDF proposal of a silicon vertex detector to improve tracking resolution and to allow the tagging of heavy quarks was written up in 1985 [27]. Since the successful MX chips from the LEP experiments were tuned to low load capacitances they were

[1] The second detector is called D0 for its ring location at D0 [161, 317].

[2] Pseudo-rapidity $\eta = -ln(\tan \Theta/2)$ replaces the azimuthal angle Θ and is invariant vs. Lorentz addition. In hadron colliders the z-coordinate of the primary vertex varies. Therefore the quantity η is used in calculations and also to define and construct the detector elements.

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_5

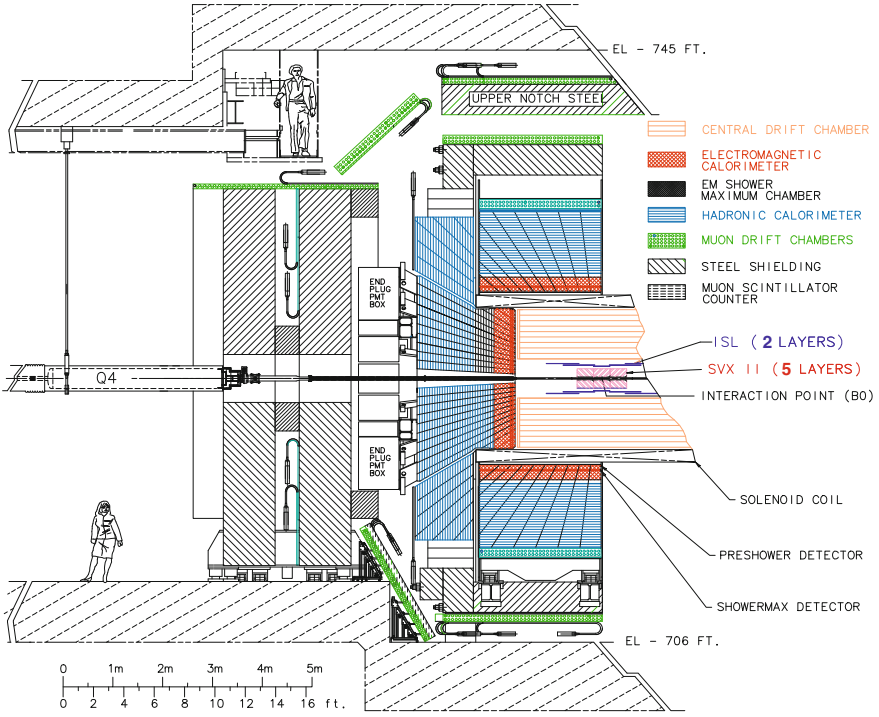


Fig. 5.1 Elevation view of half of the CDF II detector [61, 332]

not directly applicable for CDF with longer module design thus higher load capacitance. Also, bunch crossings every 3.5 µs made pulsed powering impossible. This led 1989 – 1990 to the development of the first SVX chip – refer also to Sect. 1.10 on p. 96f. It contained 128 channels with double-correlated sampling and sparsification. The first chip version worked also in quadruple sampling mode since the sensors were DC coupled and the constant leakage current had to be subtracted. More radiation-tolerant chip versions with integrated ADC and optimized for shorter beam crossings followed, with later chip iterations culminating in the SVX3 chip used for the upgrade in 2000. Cooling and mechanics of the detector were constructed in Pisa and FNAL, while all hybrids and modules were assembled at Berkeley. The expertise gathered in the operation of SVX in 1992 – 1993 led then to the SVX′, which was in operation from 1993 to 1996. The first silicon sensors were produced on 4 in. wafers with DC coupling since AC coupling had not yet matured during the design phase; it was introduced at LEP. To underline the pioneering aspect of the first vertex detector design and construction I like to cite Nicola Bacchetta:

> I think the SVX was never really a formally approved project for CDF, but rather a prototype to prove the concept. Vertexing with silicon sensors was not considered viable at hadron collider. I recall vividly people being very skeptical about all the effort …many others claiming at such radius the detector would be completely lit (100% occupied) all the time. It turned out not to be the case.

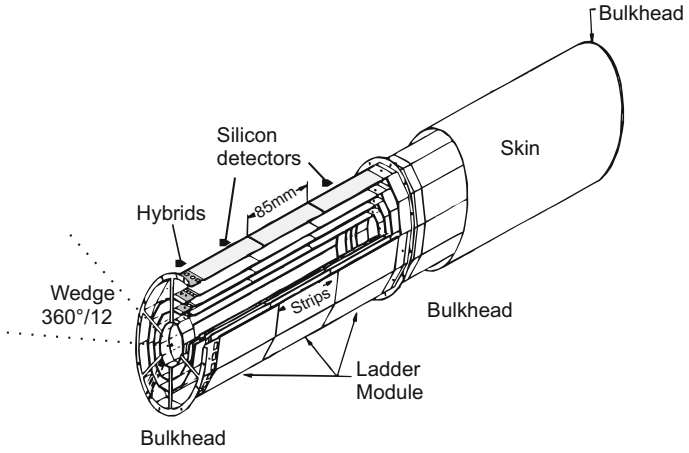


Fig. 5.2 SVX, the first vertex detector at a hadron collider, an isometric view. The consequent implementation of a wedge geometry made it necessary to have different sensor geometries per layer with different widths but same lengths. Two barrels with four layers cover 51 cm in length. Each ladder consists of three sensors with one beryllium hybrid carrying the SVX chips at the end [122]

Design criteria were similar to the LEP experiments but also additional aspects are important:

- to obtain the best impact parameter resolution, a high precision point was needed as close as possible to the interaction point together with a precision link space point or better track segment to the outer drift chamber[3]
- in contrast to the LEP experiments in the TEVATRON, a hadron collider, the interaction region is longitudinally stretched with a σ_Z of 35 cm. This requires a long detector to cover as much as possible of the interaction
- reducing multiple scattering to the bare minimum, the material budget must be minimized
- the mechanical tolerances internally must match the intrinsic detector resolution, i.e. about 10 μm, while the accurate placement with respect to the outer detector must match the resolution of the surrounding drift chamber

The SVX silicon detector (1992 – 1993) consisted of two barrels with four layers, single-sided, DC-coupled silicon sensors and non-radiation tolerant electronic chips, produced in 3.5 μm feature size. A full description can be found in [122] and a schematic view is shown in Fig. 5.2.

The four layers were situated at R =2.8, 4.3, 5.7, 7.8 cm. Information was $R\phi$ only. The barrels stretched to a total length of 51 cm covered around 60% of all $p\bar{p}$ collisions of the TEVATRON. Three silicon microstrip sensors, 8.5 cm long and

[3] CTC **C**entral **T**racker **C**hamber for RUN I and COT **C**entral **O**uter **T**racker for RUN II.

Table 5.1 Design parameters of the SVX detector in CDF at the TEVATRON, the first silicon vertex detector in a hadron collider experiment

	Width (μm)	Pitch (μm)	Number of strips	Number of chips	Sensor/layer length (cm)	At radius (cm)	Coverage $\eta\pm$
Layer 1	15.3	60	256	2	8.5/25.6	2.8	2.9
Layer 2	23	60	384	3	8.5/25.6	4.3	2.5
Layer 3	30	60	512	4	8.5/25.6	5.7	2.2
Layer 4	42.2	110	384	3	8.5/25.6	7.8	1.9

The detailed parameters are presented for the SVX from the conceptional point of view. There were no real changes towards the SVX′ detector with respect to part numbers of geometric placements. Of course, there have been significant changes in the silicon and chip technology

300 μm thick, were glued on a low-weight Rohacel bar with a readout hybrid at the end forming a 25.6 cm long module/ladder. To adapt circumference differences for different layers while keeping a perfect wedge layout, the sensor widths were adapted. In total 96 ladders with 288 sensors were implemented. A detailed list of detector parameters is presented in Table 5.1.

Pitches of 60 μm result in point resolution of 8 – 10 μm and therefore an impact parameter resolution of $\sigma_{IP_\phi} = \sqrt{13^2 + (39/p_T)^2}$ μm. Typical flight path lengths D are on the order of 350 μm. It was then possible to resolve B-decays which was a major success. Cuts on decay lengths improved invariant mass measurements greatly.

After around 30 pb^{-1} of data between 1992 and 1993 in RUN Ia with around 30 krad of radiation, the increase in sensor leakage current saturated the preamplifier and in addition significant threshold changes in the transistors deteriorated gain and increased noise.

The basic geometric parameters for SVX and SVX′ are comparable to those in Table 5.1, just layer 0 was moved 2 mm closer towards the beam pipe.

The main improvements were the technology choices of electronic chips and sensors. The new chip SVXH being produced in 1.2 μm technology was radiation tolerant up to 1 Mrad. It also featured higher gain with less noise and contained an 8-bit ADC on chip. The sensors contained integrated coupling capacitors to avoid dark current into the chip which therefore was operated in double-correlated sampling instead of in quadruple mode. The SVX′ detector was successfully operated from 1992 to 1996. During this time algorithms were already developed to have a silicon stand-alone tracking. As in DELPHI basically all sensors and modules were different for each layer which gives the detector the predicate to be a "work of art". During operation, the detector had an efficiency of >99% with an average position resolution of ∼12 μm and an asymptotic impact parameter resolution of 13 μm.

The silicon upgrade for RUN II is described in detail in the next section but first design considerations are given here – the lessons learned from the SVX and SVX′. The longer shutdown between RUN I and RUN II made a major upgrade possible. Also with the success of SVX and SVX′ the project was more fundamentally supported. The design parameters were defined by the need for a large coverage, a

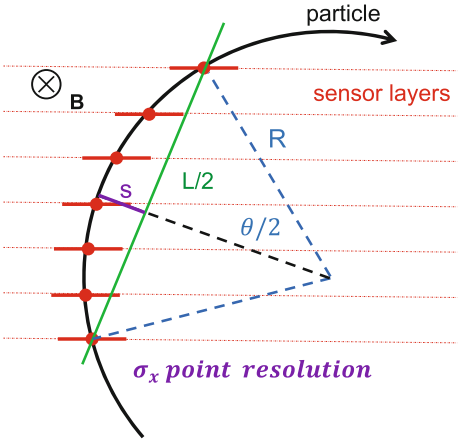


Fig. 5.3 Transverse momentum resolution p_T. The momentum resolution of a moving charged particle in a B field is given by its curvature path. Points on the path are only measured where the particle cross sensor layers with a point resolution of σ_x. With $s = L^2/8R$ and $B \cdot R = p/q$ one gets the momentum resolution as $\frac{\Delta p}{p} \approx \frac{\Delta s}{L^2 B} p$

good impact parameter d_0 and transverse momentum p_T resolution $\sigma_{p_T}(p_T)$ is often also expressed as $p_\perp$).

The transverse momentum resolution σ_{p_T} is defined by

$$\frac{\Delta p_T}{p_T} \approx \frac{\Delta s[\mu\mathrm{m}]}{(L[\mathrm{cm}])^2 B[\mathrm{T}]} p_T[\mathrm{GeV}] \tag{5.1}$$

with sagitta $s = L^2/8R$, lever arm L, magnetic field B, curvature radius R and transverse momentum p_T. The equation immediately tells that (a) intrinsic position resolution has to be good to resolve s and that (b) the B field strength gives a linear improvement, while (c) a larger lever arm improves momentum resolution quadratically. An explanatory scheme is given in Fig. 5.3. With increasing p_T the resolution gets worse again and with an error of 100% not even the charge of the particle can be identified anymore.

A layer around $R = 20\,\mathrm{cm}$ was necessary to improve σ_{d_0} and p_T resolution. The short lever arm of 5 cm realized in the SVX′ detector resulted in a poor p_T resolution and the track extrapolation to the drift chambers was not precise. An overall increase in length was needed to increase coverage. This led initially to a five layer SVX II detector with layer 4[4] at $R = 10.6\,\mathrm{cm}$. The length of SVX II was increased to 96 cm by adding a third barrel. To improve pattern recognition and allow for a 3D vertex reconstruction double-sided, double metal silicon sensors were introduced. To bridge between vertex and drift chambers an **I**ntermediate **F**ibre **T**racker IFT or an **I**ntermediate **S**traw **T**racker IST was proposed. Both concepts suffered from the necessary extensive R&D in the short available time period. Both were finally cancelled and additional **I**ntermediate **S**ilicon **L**ayers were introduced, the ISL detector. To make it affordable, a large fraction of the sensors were produced

[4]Layer 4 is the fifth layer, since counting starts at layer 0 with an additional layer 00 further in and the outer ISL layer count starts at 6.

on 6 in. wafers[5] – the first time for HEP experiments. The only remaining flaw in the concept of this beautiful detector is the unavoidable amount of material disturbing the impact parameter resolution, thus the b-tagging capability. This was solved by adding even more material but at a very low radius, namely adding a layer 00 at $R = 1.6$ cm directly onto the beam pipe. The closer space point recovers σ_{d_0} for the latter multiple scattering, refer to formula (4.1).

With an expected integrated luminosity of $2\,\text{fb}^{-1}$ the front electronics radiation tolerance needed to be improved. This led to a third generation of the SVX chip series – the SVX3 chip (Sects. 1.10 and 5.2.1). Somehow the SVX detector noise increased above expectation during RUN Ib and this was attributed to the FOXFET biasing, being less radiation tolerant than anticipated. As a result for RUN II polysilicon bias resistors were introduced. Last but definitively not least the SVX II information was introduced into the Level-2 trigger – the CDF **S**econdary **V**ertex **T**rigger SVT.

5.2 Design, How to Cover $|\eta \leq 2|$ Without Endcap

In 2000 the TEVATRON was upgraded for RUN II to higher luminosity and a slightly higher energy 1.8 to 2 TeV. The corresponding tracking system is described in detail in this section.

5.2.1 Tracking System

Efficient precision charged particle tracking is extremely important for the CDF analysis technique. Reconstruction of both high $p_T(m_W)$ and low $p_T(B \rightarrow J/\Psi K)$ is required. The combination of track, calorimeter and muon chamber information, with an excellent purity at both the trigger and offline level, is possible. Precise and efficient b-tagging is essential for top t-quark physics and new phenomena searches. The goal is to guarantee precise 3D impact parameter resolution with an enhanced coverage up to $|\eta \leq 2|$. The CDF II tracking system consists of an inner silicon vertex tracking system and a large drift chamber. The inner tracker consists of a minimum radius inner layer (Layer 00 at 1.35 cm) glued to the beam pipe and a five-layer silicon detector (SVX II at $R = 2.7 - 10.7$ cm) with two-dimensional readout in each layer. It is surrounded by a third two-layer silicon detector (ISL $R = 20$ and 28 cm) and finally an eight-layer open cell drift chamber (**C**entral **O**uter **T**racker COT at $R = 44 - 132$ cm). It is worth mentioning that the $|\eta \leq 2|$ coverage was achieved with a consequent long barrel geometry and without any forward structures. Layer 00 and SVX II are ∼1 m long, the ISL spans even ∼2 m.

[5]The 6 in. wafer contained two sensors in one processing step.

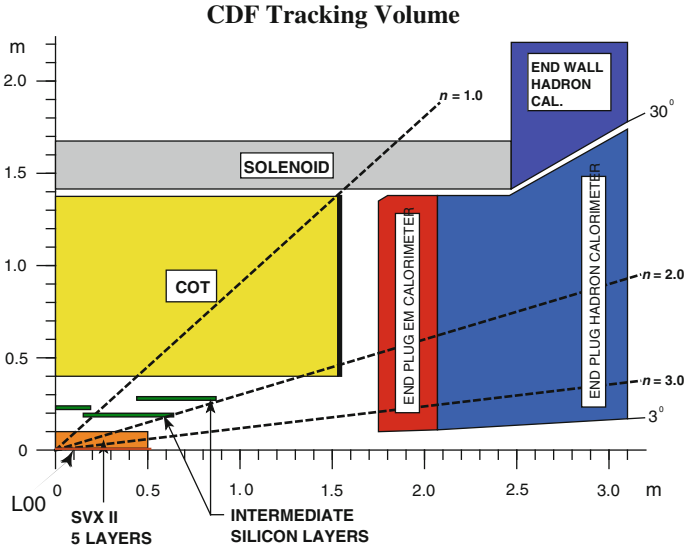


Fig. 5.4 Longitudinal view of the CDF II tracking system [272]

The full tracking volume can be seen in Figs. 5.4 and 5.5, the latter giving an expanded view of the inner part. The inner and outer tracking systems are capable of stand-alone tracking. This enables track–track matching instead of the normal track space point fit. The COT gives Level-1 trigger information and the inner silicon tracking system provides Level-2 information. The trigger acts on displaced vertices.

Layer 00

With a silicon layer placed at very small radius, the first precision space point is recorded without earlier multiple scattering thus improving impact parameter resolution. A 6-wedge layer is mounted directly on the beryllium beam pipe with single-sided-silicon microstrip sensors, each being 7.8 cm long and 0.8 cm or 1.5 cm wide. The length of Layer 00 is approximately 94 cm, consisting of six modules with two sensors per module, summing up to a total of 144 sensors. Sensors are placed at $R = 1.35$ cm and $R = 1.6$ cm (see Figs. 5.5 and 5.6).

The caveat of a very high radiation environment is overcome by using radiation hard silicon sensors following early design recipes of the freshly developed LHC sensors and silicon material proposed by the RD50 collaboration. The sensors were single-sided only with p-on-n $R\phi$ readout. They were produced on $\langle 100 \rangle$, 300 μm thick silicon, AC coupled with a multi-guard ring structure to guarantee high voltage operation, which is the key to radiation-tolerant operation. Some sensors were even oxygenated (see Sect. 2.1). Strip implants were spaced 25 μm while every second strip is connected to the readout. The sensors are actively cooled to freeze radiation damage and reduce leakage currents.

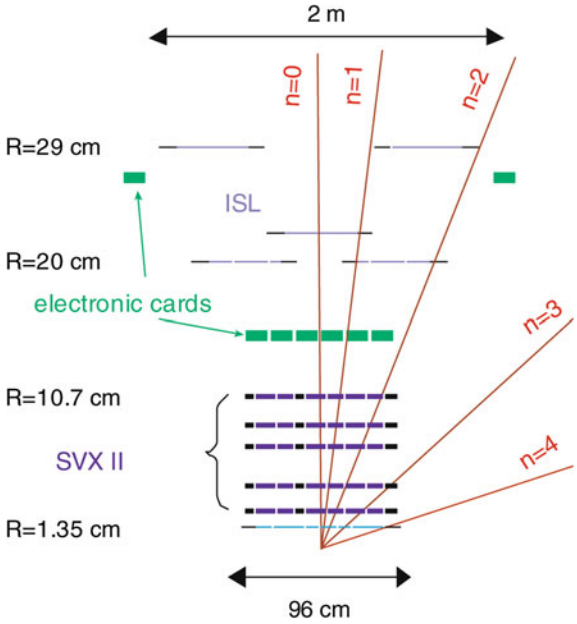


Fig. 5.5 CDF silicon tracking system. Full tracking is possible up to $|\eta \leq 2|$ [272]

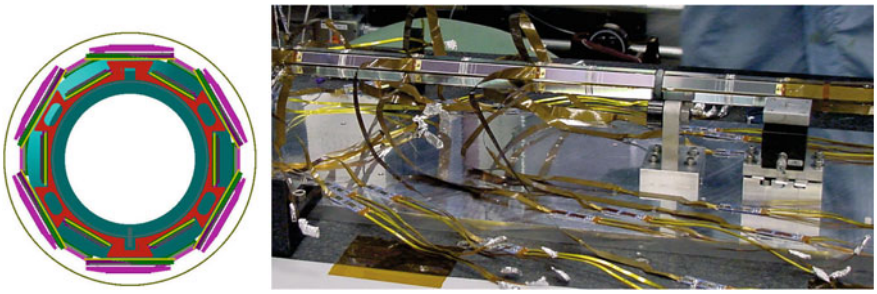


Fig. 5.6 End view of the layer 00 detector and a rare view during assembly. The *left* side expresses the tightness, allowing only 0.8 cm (1.5 cm) wide sensors at a radius of 1.35 cm (1.6 cm). Layer 0 of the SVX II detector sits already at R = 2.45 cm. On the *right* Layer 00 is visible with silicon sensors not yet covered by the thin long fine-pitch cables. The main importance of Layer 00 is the first high precision space point at very low radius to seed the impact parameter calculation before multiple scattering [154]

The radiation-tolerant SVX3 chips are mounted along the line separated from the sensors by longer cables at larger radii. Signals are carried out by thin, fine-pitch cables up to a length of 47 cm. Free cables can be seen in Fig. 5.6 while final assembly of Layer 00 during insertion with cables strapped tightly together is displayed in Fig. 5.7.

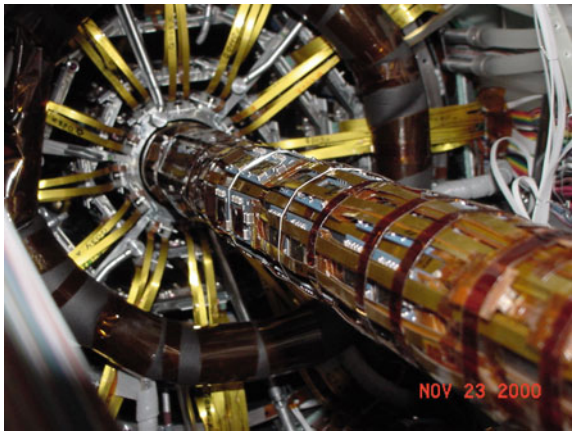


Fig. 5.7 Layer 00 during insertion. The clearance towards the SVX II detector is about 300 μm only. Insertion finally succeeded in November 2000 [154]

A summary of parameters is given in Table 5.2. The Layer 00 was proposed very late as a "beyond baseline" detector component, a detailed description of the early idea can be found in [189, 221].

Silicon Vertex Detector SVX II

The **S**ilcon **V**ertex **D**etector II SVX II is divided into 12 wedges in $R\phi$ and in 3 barrels in z (beam axis) with a length of 29 cm each. This results in a total length of 96 cm, almost twice as long as the former SVX or SVX′. With the length of almost 1 m SVX II covers $\approx 2.5\sigma$ of the interaction region providing track information up to $|\eta \leq 2|$. It has five double-sided silicon layers measuring the $R\phi$ and Rz coordinates. Three layers (0, 1 and 3) have a 90° stereo angle allowing high-resolution Rz measurements, while layers 2 and 4 have a 1.2° stereo angle. This design provides good pattern recognition and 3D vertex reconstruction with an impact parameter resolution $\sigma_\phi < 30$ μm and $\sigma_z < 60$ μm for central high momentum tracks. Figures 5.4 and 5.5 show the location of the SVX II detector and the ISL. The SVX II replaces the former SVX – a four-layer, single-sided silicon microstrip detector. A photo of one SVX II barrel is displayed in Fig. 5.8. Half-ladders contain two sensors plus a hybrid mounted directly to the silicon surface at the end. Two half-ladders are daisy-chained together to form full ladders of four sensors each. A total of 720 sensors form 360 half-ladders or 180 full ladders. A photo of an SVX II module is presented in Fig. 5.9.

A 12-fold ϕ symmetry makes it possible to treat each 30° wedge as an independent tracker. As for the SVX, the sensor widths had to be adapted for each layer to make a perfect wedge geometry possible. The data from SVX II are used at the Level-2 **S**ilicon **V**ertex **T**rigger SVT [215], which identifies displaced vertices coming from B fragmentation. A placement precision of better than 100 μm with respect to the beam was necessary to use track information in the Level-2 trigger.

All SVX II sensors are AC coupled, double-sided, 300 μm thick with polysilicon bias resistors. The $R\phi$ side contains the p^+-implants, while the Rz or stereo side is composed of n^+implants isolated by a p^+-stop configuration (see also Fig. 5.17). For

Table 5.2 Parameters of the CDF II Silicon Sensors, SVX II, ISL and Layer 00. The large variety of silicon sensor types gives an impression about the complexity of the device. Only the large area outer ISL detector is fabricated with a single sensor type

Sensor parameters	Layer 00			SVX			ISL	
	Layer 00	Layer 0	Layer 1	Layer 2	Layer 3	Layer 4	ISL L6 (4 in.)	ISL L7 (6 in.)
Radial distance (cm)	1.35/1.6	2.45	4.67	7.02	8.72	10.6	19.7/20.2 F	28.6/29
							22.6/23.1 C	
Stereo angle°	–	90	90	+1.2	90	−1.2	1.2	1.2
$R\phi/z$ readout channels	128/256	256/512	384/576	640/640	768/512	896/896	512/512	512/512
$R\phi/z$ readout chips	1/2	2/2	3/3	5/5	6/4	7/7	4/4	4/4
$R\phi/z$ strip (μm) pitch	–	60/141	62/125.5	50/60	60/141	65/65	112/112	112/112
Total width (mm)	0.85/1.48	17.14	25.59	40.30	47.86	60.17	59.3	59.3
Total length (mm)	7.84/7.84	74.3	74.3	74.3	74.3	74.3	69	74,8

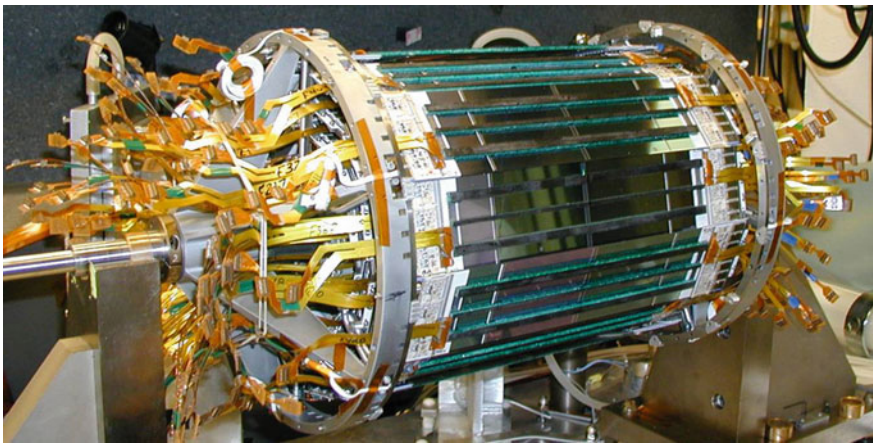


Fig. 5.8 The SVX II detector, one barrel [42]

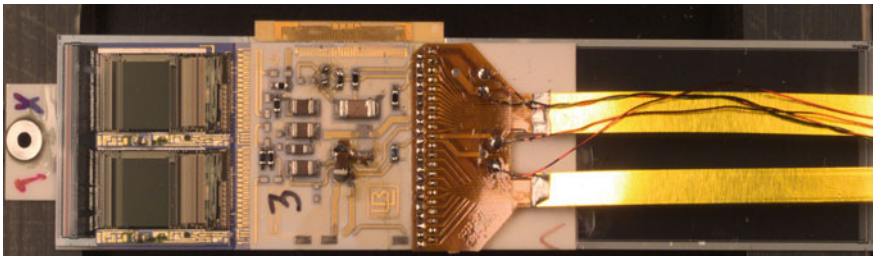


Fig. 5.9 A layer 1 hybrid with sensor [42]

the 90° Rz strips, a double metal layer routes the signals to the readout electronics. A photo of the double-sided sensor layout is given in Fig. 5.10. For the small stereo angle sensors no extra routing is needed. In Fig. 5.17 one can see how layer 2 and layer 4 sensors are processed on a single wafer with the new 6 in. sensor technology. A more detailed description of the advantages of 6 in. technology is presented in Sect. 5.3 and in the summary of sensor parameters in Table 5.2.

Intermediate Silicon Layers ISL

The intermediate silicon layers 5 and 6, both consisting of double-sided silicon strip detectors, measure both the $R\phi$ and Rz coordinates. The strips on one side of the wafers are parallel to the beam axis (z-axis) while the strips on the other side are tilted by 1.2°. This allows the measurement of the Rz coordinate with low ambiguity[6] and no additional double metal layer. The ISL layers are mounted at the radii 20 and 28 cm. The overall length of the ISL is 195 cm, covering the pseudo-rapidity $|\eta| \leq 2$.

[6]The angle of 1.2° allows the connection of strips of several consecutive sensors to one single electronic readout channel, as opposed to layers with a stereo angle of 90° where multiple strips are connected to one routing line, hence to one readout channel.

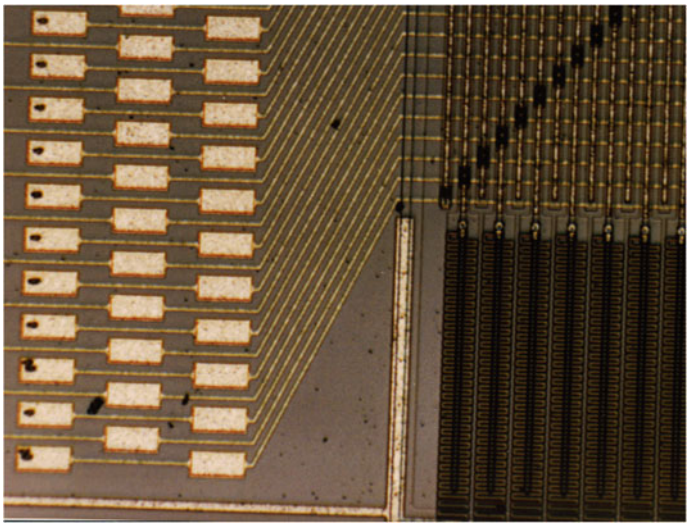


Fig. 5.10 The double-sided double metal layer 90° stereo sensor of the SVX II. The bonding pads are located on the *left*, they are connected to the Al strips on the double metal and are bonded to the readout electronics. The polysilicon bias resistors are located in the *lower right part* of the picture, the meander structures. The *black spots* on the *upper right* are the vias connecting *signal lines* to readout lines

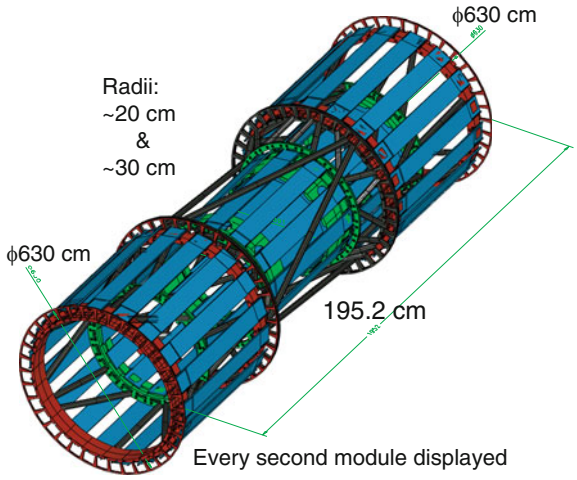


Fig. 5.11 The intermediate silicon layers detector—ISL. The schematic shows the design drawings of the carbon-fibre spaceframe. The geometrical concept of the single additional layer in the central region and the two additional layers in the forward direction can be seen. The carbon-fibre rings are connected with hollow carbon-fibre rods. The total weight of the structure is about 6 kg before mounting the modules [138]

This special long configuration enables "forward" tracking without forward structures. The whole support structure – the spaceframe – was designed with respect to weight, stiffness and material budget, which led to a carbon-fibre structure. The mechanical arrangement can be seen in Fig. 5.11 and an impressive view is shown in Fig. 5.12, a photo of the ISL before it was closed and inserted into the COT. The

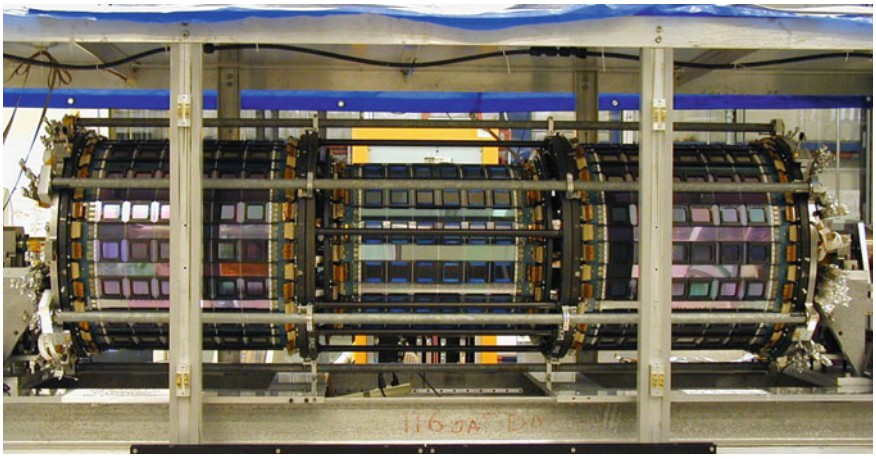


Fig. 5.12 The intermediate silicon layers detector. [Courtesy of CDF, Fermilab]

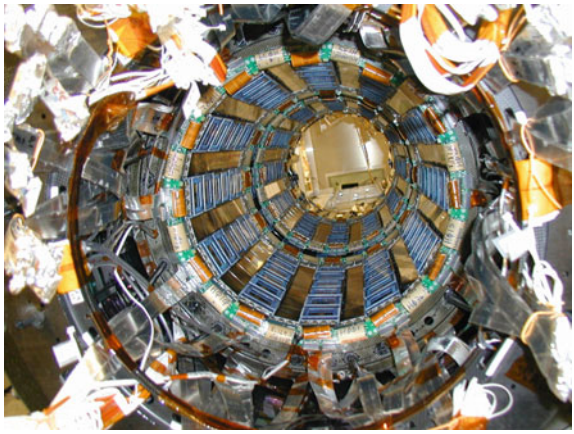


Fig. 5.13 Into the ISL. The full ladders, six sensor long with readout electronics on both sides, are visible in a barrel configuration, a rare view during construction. [Courtesy of CDF, Fermilab]

early design is described in [322]. Figure 5.13 allows a glimpse into the detector, the space where finally the SVX II was located. In contrast to SVX II, ISL was designed with simplicity in mind. A more detailed description of the ISL can be found in [136, 138]. The ISL sensors are also AC coupled, polysilicon biased and double-sided, with p^+-stop configuration on the n-side. For the larger radii, occupancy and radiation damage are lower. It is therefore possible to use longer strips and pitches are relaxed to 112 μm on the $R\phi$ and stereo side. This is also necessary to reduce the number of readout channels, thus the costs of front-end electronics and DAQ equipment. Sensors from both old 4 in. and new 6 in. technologies [43, 136] are used. In the 6 in. case, two sensors are processed on a single wafer, see Fig. 5.16. Pitch adapters are used to bring the signals from the strips to the more closely spaced inputs of the SVX3 chips. The ISL ladders are composed of six sensors arranged as

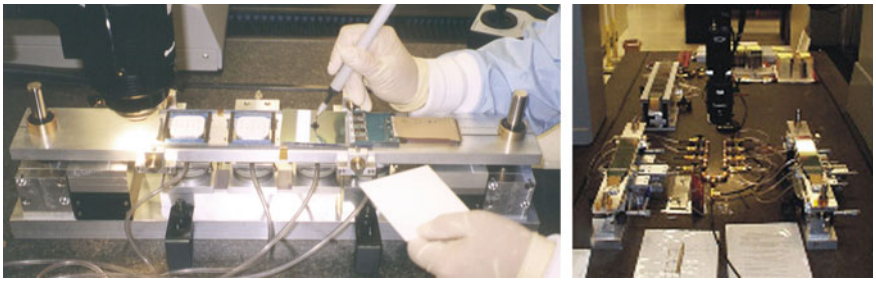


Fig. 5.14 Manual assembly of the CDF ISL ladders. The sensors resting on vacuum chucks were aligned and glued to the supports. The *right picture* show three half-ladders to complete the curing of the glue

half-ladders of three sensors each with a double-sided hybrid at each half-ladder end. A total of 296 half-ladders form 148 full ladders summing up to a total of 888 sensors with 2368 chips. The ladders are composed of carbon fibre. The assembly fixture is displayed in Fig. 5.14. Relaxed space constraints in the high radius regions allow overlapping of ladders even for the z-coordinate.

Central Outer Tracker COT

The COT is an open drift chamber with an inner radius of $R = 44\,\text{cm}$ and an outer radius of $R = 132\,\text{cm}$ and a coverage in η up to $|\eta| \leq 1$. It replaces the **C**entral **T**racking **C**hamber CTC used in RUN I. Four axial and four stereo "superlayers" provide 96 measurements[7], resulting in a total of 2520 drift cells and 30240 readout channels. The different wire angles with respect to Rz in each "superlayer" are $+3°\ 0°\ -3°\ 0 + 3°\ 0°\ -3°$. The location is shown in Figs. 5.1 and 5.4. The main changes in the upgrade are the small drift cells with a maximum drift distance of 0.88 cm and fast gas to limit drift times of less than 100 ns. The readout is realized via a pipelined TDC[8] system. The COT is also equipped with the possibility to measure dE/dx for particle identification. The measurement is complementary to the **t**ime-**o**f-**f**light TOF method. A more detailed description is given in [61].

The SVX3 Chip [61, 115, 248]

All silicon sensors are read out by the SVX3 chip, a radiation-tolerant CMOS custom-integrated circuit. Each chip has 128 parallel input channels. The chip consists of two parts built into a monolithic structure: the front-end for the analogue functions and the back-end part for the digital functions. The input amplifiers, integrators, the 46 cell pipelines needed for dead-timeless operation and the pipeline acquisition logic is located in the front-end. Up to four groups of cells can be queued for digitization and then read out at one time. The back-end consists of an 8-bit Wilkinson ADC, a

[7] One "superlayer" provides 12 measurement points.

[8] **T**ime to **D**igital **C**onverter: TDC.

readout FIFO[9] and differential output drivers. The chip is able to work with positive and negative signals to handle double-sided silicon detectors. It has a sparsification mode, a calibration input and is capable of pedestal subtraction at the chip level. The chip is programmable to adjust for different TEVATRON running conditions, e.g. bunch spacings (132 and 396 ns). It can handle strip capacitances from 10 to 35 pF. The pipeline depth, threshold level and bandwidth limit are adjustable. The chip has been irradiated with a ^{60}Co source up to 4 Mrad and was evaluated to perform well[10] [115, 248]. The amplification is 15 mV/fC which corresponds to 60 mV for a minimum ionising particle. The chip provides dead-timeless readout with Level-1 trigger rates up to 50 kHz. The maximum delay between the Level-1 trigger decision and read out without overwriting the pipeline cell content is 5.5 μs at 132 ns between beam crossings.

Silicon Vertex Tracker SVT – Secondary Vertexing

The ability to use impact parameter information in the trigger to detect secondary vertices can substantially increase the physics reach of a hadron collider experiment. Background can be quite substantially reduced e.g. for the process $Z \longrightarrow b\bar{b}$. The B-decay studies will also be greatly enhanced. Some physics processes not involving b quarks will also profit from the SVT e.g. the high p_T inclusive muon trigger, needed for the W-mass measurement, has a high Level-2 rate. The SVT can both reduce the accidental rate by demanding an SVX II track pointing to the primary vertex and remove lower p_T muon background by using improved momentum resolution to tighten the p_T threshold. In order to obtain impact parameter information at Level-2 the silicon tracker is read out after each Level-1 trigger. The SVT combines data with the Level-1 tracking information from the COT and computes track parameters (ϕ, p_T and impact parameter d_0) with a resolution and efficiency comparable to full offline analysis [61, 215]. The SVT also introduced several hard constraints on the detector design and assembly:

- wedge geometry – compatible with hardware regional track reconstruction
- very tight mechanical tolerances – placement precision of 100 μm and better with respect to beam
- dead-timeless readout with onboard digitization, sparsification and buffering
- dedicated DAQ pipeline for processing and control only for the SVX

The implementation of a silicon Level-2 hardware trigger on impact parameter information is unprecedented in a hadron collider detector.

[9] **F**irst **I**n **F**irst **O**ut **FIFO**: Register where the data entered at the input, appears in the output in the same order.

[10] During RUN II the chip will be exposed to about $7 \cdot 10^{13}$ minimum ionising particles per cm^2 over a period of 2 years, corresponding to an absorbed dose of 2 Mrad at the inner layer of SVX II [23].

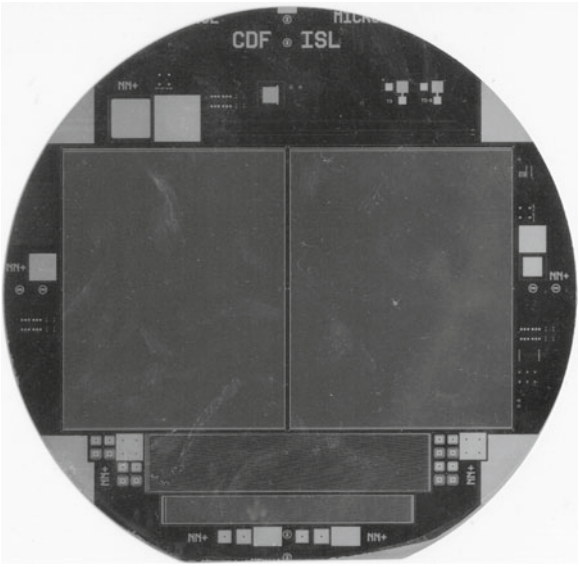


Fig. 5.15 Photo of a full 6 in. ISL wafer – real size *0.5. Two ISL sensors plus two smaller study sensors and many small test structures like diodes and resistors can be seen. The four rounded fully metallized triangles at the perimeter are for handling purposes

5.3 Six Inch, a New Technology Step for Large Silicon Applications

The fabrication of microstrip detectors on 4 in. high resistivity wafers allowing a maximum processable area up to 42 cm^2 has been established at LEP and RUN I at the TEVATRON. The workable area using 6 in. wafers increases up to 100 cm^2. Figure 5.15 shows a photo of a full wafer.

Since this is twice the area of a 4 in. wafer a larger number of sensors can be processed at the same time on the same wafer resulting in significant reduction of cost. CDF is the first HEP experiment using 6 in. sensors [43, 136]. Figure 5.16 show the mask layout of the 6 in. wafers for ISL and SVX II sensors, respectively. In the case of the ISL wafer both sensors are identical in size. In the SVX II wafer the sensors are produced for layers at different radii and hence have different geometries, specifically widths. In contrast to the production on 4 in. wafers, the cost of one mask set is saved which is a substantial fraction of the total cost, especially for double-sided sensors and small volume production. The ISL sensors on 6 in. design are even 5.75 mm longer than their counterpart in 4 in. design. Figure 5.17 shows the different electrical structures of an ISL sensor. A photo of the sensor surface (n-side) is shown in Fig. 5.18.

All sensors for SVX II and ISL were AC coupled, polysilicon biased, double-sided, produced out of ⟨100⟩ wafers. For 90° Rz readout strips a double metal was implemented over a 5 μm thick isolation to route the strips to the hybrid at the

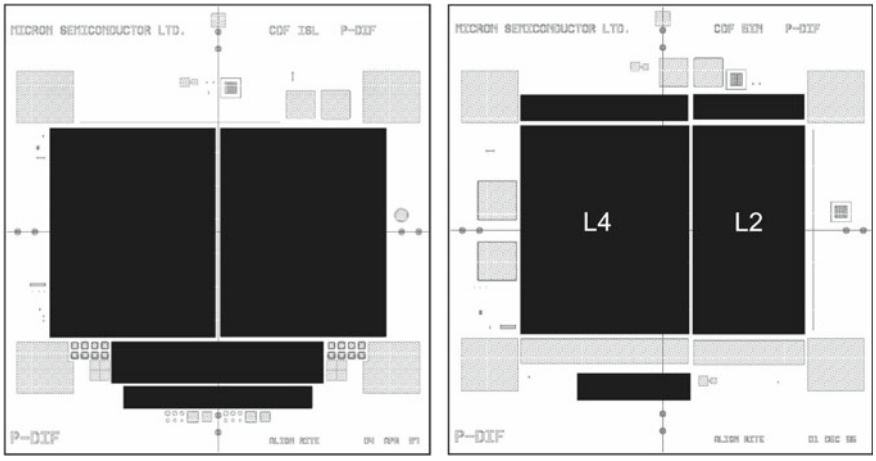


Fig. 5.16 Mask layout of an ISL and a SVX II wafer. Two sensors fit on one wafer. One ISL wafer is shown with two identical sensors, while on the right the SVX II wafer caries sensors with different layout – a layer 2 (L2) and a layer 4 (L4) sensor. At the *bottom*, mini sensors and test structures are placed

end of the ladders. Large thickness[11] of metal-to-metal isolation is mandatory to suppress cross-talk and keep capacitance low, especially parasitic capacitance. Also the aluminium trace widths are reduced to 8 μm to reduce capacitance. Two or four sensors were connected to one hybrid for SVX II and ISL, respectively. Coupling capacitor isolation was achieved with a pattern of silicon oxide and silicon nitride. The aluminium strip width was 6 μm smaller than the implant width below to suppress micro-discharge. All corners and edges have a radius of >10 μm to avoid high fields and any doped region has a concentration of at least 10^{14} ions/cm^2 to ensure radiation hardness. An active n^+ edge field shaper structure is implemented, refer to p. 49. Strip implants continue below the polysilicon resistors to maximize active surface. More details can also be found in [43, 227].

Initial Problems with First Sensors from 6 in. Wafers

In conclusion switching from 4 to 6 in. sensors was not completely straight forward, it was a bit painful, but feasible. In the end, fabrication on 6 in. sensors is cheaper and performance is as good as for 4 in. Some significant issues related to size increase are shown here, some other problems are presented in Sect. 1.14.

[11] Note that the double metal scheme was only applied to 4 in. wafers; for 6 in. wafers mechanical stress imposed by the thick oxide would have violated manufacturer design rules. Sensor bow would have been large, probably large enough to break processed structures.

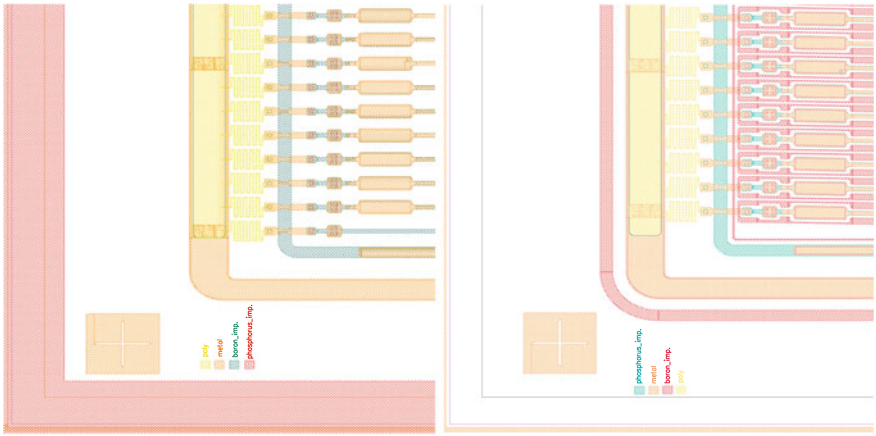


Fig. 5.17 Structure of an ISL sensor – p-side (*left*) and n-side (*right*). All 6 in. sensors for ISL and SVX II are 300 μm thick on ⟨100⟩ oriented Float-Zone silicon. The sensors are double-sided, single metal, AC coupled, polysilicon biased, with a common p^+-stop structure on the ohmic side for SVX II and isolated p^+-stop structure for ISL to guarantee inter-strip isolation

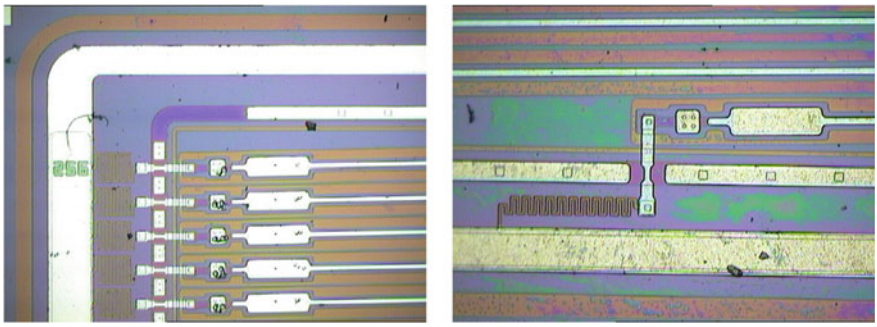


Fig. 5.18 Photo of the stereo n-side of an ISL sensor. The following structures are displayed from *left* to *right*, an outer guard ring, the bias ring with meander bias resistors with *metal lines* passing through the inner main guard ring ending on the small DC pads. The DC pads are then contacting the p^+-implants by a vias. The n^+ strips are fully surrounded by a p^+-stop structure. The mentioned inner guard ring is mostly covered by aluminium, the small squares mark the vias. The strips are arranged with a stereo angle of 1.2° with respect to the strips on the junction side. In the *right* picture a middle part can be seen where, due to the angle, the strip does not reach the end of the sensor. Therefore the bias resistor angles in from the side

Early Low Capacitances at the Wafer Edges

During the prototype phase, coupling capacitance values decreased significantly towards the wafer edge, dropping below specification limits. The cause was identified to be a non-uniform aluminium sputtering process. The process was not yet fully adapted to larger wafer sizes. This resulted in a substantial over-etching of the Al strips, thereby decreasing aluminium width, thus coupling capacitances. A cartoon of the process problem plus some measurement results are presented in Fig. 5.19.

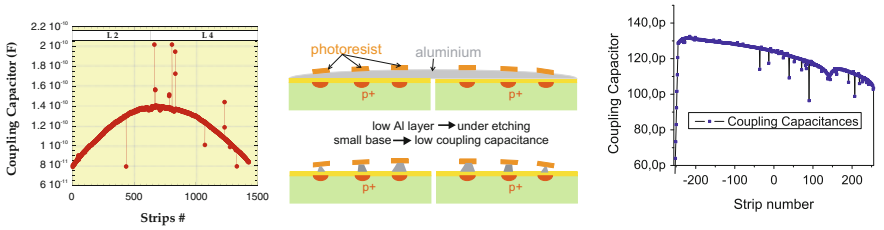


Fig. 5.19 Low coupling capacitances at the edges [43, 136]. On the *left* a measurement on a full wafer is shown while on the *right* measurements of an already cut sensor, representing half a wafer is shown. The cartoon in the *middle* explains how the over-etching happened

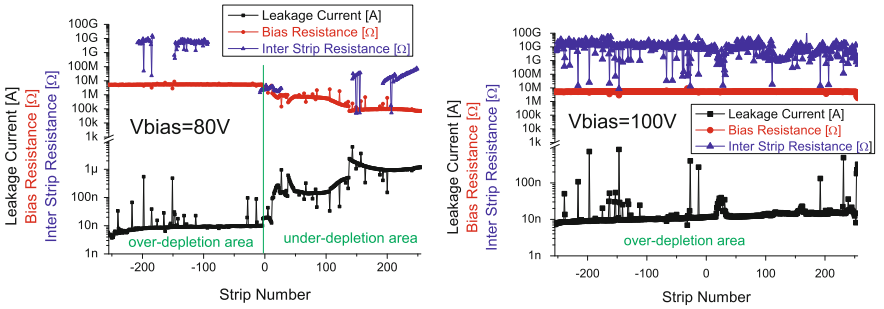


Fig. 5.20 Inter-strip resistance problems due to inhomogeneous wafer material. Half of the sensor shows a low inter-strip resistance on the n-side of the sensor. The effect was identified by leakage current and R_{poly} measurement and confirmed with a full R_{int} measurement. Low R_{int} connects strips, therefore current measurement also collects components from neighbour strips and bias resistors are now arranged in parallel configuration. Different bias voltages were applied during the *left* ($V_{bias} = 80$ V) and *right* ($V_{bias} = 100$ V) measurement. While only half of the sensor was full depleted with $V_{bias} = 80$ V, the whole sensor was fully depleted at $V_{bias} = 100$ V. This bias or full depletion characteristic can be explained by an inhomogeneous resistivity of the silicon wafer. Since, in the ISL case, the sensor only covers half of the wafer the decrease of resistivity towards the centre of the wafer translates to a decrease in one sensor to one side

Two Cases of Low Inter-Strip Resistance

Several sensors were received, showing a defined pattern of several clusters of five consecutive strips with low inter-strip. The problem was tracked down to a PECVD[12] machine. The pattern matches gas outlet positions, where an increased susceptibility to charge-up was introduced. Correct cleaning cured the problem.

A more severe problem is shown in Fig. 5.20 where low inter-strip resistance values were identified on only half of a sensor. The effect was present only on the n-side. Increasing the bias voltage cured the effect completely, leading to the suspicion that the depletion voltage on one half of the sensor is higher, thus the resistivity is lower. This explains why the effect is only present on n-side, because the depletion zone reaches n-side only after full depletion, however CV measurements on diodes

[12] **P**lasma-**e**nhanced **c**hemical **v**apour **d**eposition: PECVD.

contradicted this hypothesis. It was found later that the 6 in. wafer had a resistivity drop in the centre[13] of the wafer. A good fraction of sensors had to be rejected.

Another problem with a low inter-strip resistivity encountered during the CDF quality assurance is presented in Sect. 1.14 where the sensor surface was charged up during transport.

All the problems encountered during prototype phase and first production runs were identified and solved.

5.4 Lessons Learned from Operation

In this chapter some incidents that occurred during commissioning and operation will be listed. It is a tale of unexpected incidents to be expected in future experiments. The CDF silicon detector was installed in February 2001. Commissioning of the CDF II silicon detector took place until spring 2002 [45].

Initially, the cooling lines to the central ISL barrel were blocked by remains of epoxy. This blockage was later solved by introducing a boroscope with a surgical laser to bore the glue out.

Failing power supplies and pick-up noise on the layer 00 had to be countered by a special readout arrangement.

A much more severe and unexpected phenomenon was identified when a substantial fraction of modules failed (14/704) long after the commissioning phase. The rate was not compatible with the expected infant mortality rate. The failure was strongly correlated with the trigger frequency. The wire-bonds, perpendicular to the 1.4 T magnetic field carry power for the SVX3D chips from the $R\phi$ to the Rz side of the SVX II hybrids. Any current fluctuations induce a changing Lorentz force on the wires. During consecutive, synchronous chip readout, these forces drive the wires to resonate and cause them to break. A new VME board was introduced to prevent wire-bond resonances by stopping the readout when the trigger rate exceeds a certain frequency. Since installation of the board no further wire-bond failures have been encountered. In addition, the digital current consumption of the SVX3 chips has been lowered by reducing the power output of their digital drivers. Further detailed discussions about this topic can be found in [46].

Another set of very serious incidents were beam incidents from the TEVATRON where the detector was subjected to high dose rates of $\geq 10^7$ MIPs/cm^2 in a period of time $\leq$150 ns. During the first weeks after an incident the affected readout chips returned data with bit errors. After some months at least some of the affected readout chips recovered. As a consequence a fast interlock, aborting the beam before it has time to de-bunch, has been implemented. In addition, collimators have been installed in order to intercept deflected particles and minimizing the dose rate during an incident. After these implementations no additional damage has been recorded

[13] With two sensors on one wafer, the effect is present on just half of one sensor.

despite the occurrence of several RF failures and kicker pre-fires. Further details can be found in [45].

A small concern, since non-destructive, is the regular loss of communication to the non-radiation tolerant power supply mainframes in the collision hall. These crates need a reset from time to time, a so-called "hockerization" [145]. During the shutdown 2007/2008 another cooling problem was solved. An acidification of the cooling liquid caused a leak in the cooling joints. In an heroic effort again a boroscope and some catheters plus a small custom-made brass wire were utilized to deposit epoxy from the inside of the tube to the corroded areas [96].

An expected degradation derives from radiation. Signal is constantly decreasing while shot noise is increasing due to the increase of leakage currents. Especially in the inner layers, changes of depletion voltages can be observed. The data follow roughly the expectation given by the Hamburg model. Actually, the effect of radiation damage is being monitored and preliminary results of 2008 suggested the detector

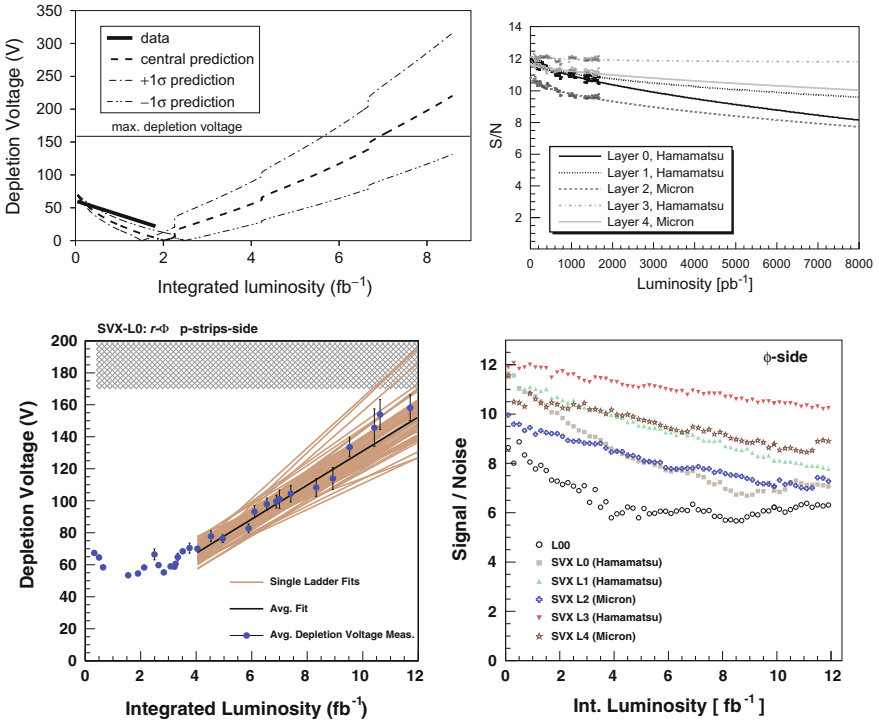


Fig. 5.21 Depletion voltage and signal/noise vs. fluence at CDF. The *upper* plots show the situation as presented in 2007 with predictions up to $\mathcal{L} = 8\ fb^{-1}$ (from [199]) and *lower* the final result evaluated after the end of operation with a total of $\mathcal{L} = 12\ fb^{-1}$ (from [29]). It is interesting to see predications to hold, although, fortunately the more positive extrapolation came true. SVX II layer 0 depletion voltage evolution is shown in the *left* plots. Expectations are met. In the *right* plots signal/noise values are plotted

will survive 8 fb^{-1} [96, 199, 348] – and it did – it even did survive $\mathcal{L} = 12 fb^{-1}$ of integrated luminosity. Depletion voltage and signal/noise changes vs. fluence is shown in Fig. 5.21. An interesting point in this discussion is the determination of the depletion voltage when a CV scan is not possible. By varying the bias voltage of the silicon sensors, noise values and signal when beam is present are recorded. When the signal maximizes and the noise minimizes or better plateaus their full depletion voltage has been reached. For CDF this is a manual operation and does not work in parallel to data taking.

In 2009 the CDF Run II silicon detector was running with 92% of all sensors and with a data-taking efficiency greater than 97%. Maintaining the detector at a high efficiency level requires a significant effort compared to other CDF sub detector systems, especially in terms of human resources. At least two of about 10 silicon operations group members were on 24-h call.

CDF II operations ended in September 2011, the TEVATRON delivered $\mathcal{L}$=12 fb^{-1} of integrated luminosity of $p\bar{p}$ collisions at $\sqrt{s}$=1.96 TeV – way above design! A more detailed overview of the operation experience of the CDF II silicon vertex detector can be found here [29].

5.5 The *t* Discovery, CP Violation in the *b* Quark Sector

How can precise track and vertex information help to reduce background? An impressive example is given in [204] where the $B^{\pm} \rightarrow J/\Psi K^{\pm}$ mass spectra is given with and without SVX $c\tau > 100$ μm cut, see Fig. 5.22. The cuts allow for fully reconstructed exclusive decays of b-hadrons together with a precise hadron momentum measurement in the magnetic field. This capability is also a milestone for any attempt to measure CP violation in the B sector.

In 1994 the evidence of a top quark was published in [5] – a milestone in particle physics finally achieved. Detailed descriptions are written in [6, 123, 147, 341]. Some brief analysis descriptions are presented where the SVX was instrumental in the initial identification. In later stages the SVX II with its larger coverage played also a vital role to reduce background for precision measurements. At the TEVATRON top quarks are produced ∼90% of the time by quark–anti-quark ($q\bar{q}$) annihilation into a gluon which then decay into $t\bar{t}$. Within the Standard model, the dominant decay is then

$$t\bar{t} \rightarrow W^{+}b,\ W^{-}\bar{b} \tag{5.2}$$

with further decay of W

$$W \rightarrow \begin{cases} (u\bar{d}), (c\bar{s}) & \text{hadronic decay with branching ratio BR : 2/6} \\ (e\nu, \mu\nu, \tau\nu) & \text{leptonic decay with branching ratio BR : 1/3} \end{cases} \tag{5.3}$$

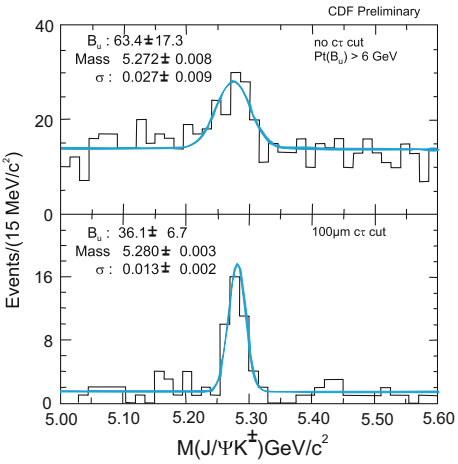


Fig. 5.22 Mass determination without/with vertex cut at CDF [204]. The cτ cut reduces background drastically thus reducing the statistical error. This is especially important for a hadron collider environment

The W decay products categorize the $t\bar{t}$ events into "full hadronic", "di-lepton" or "lepton + jet" channels, but there are always two additional b-quark decays which may be "tagged" by a displaced vertex and/or another lepton if the b or a secondary charm decay is semileptonic. In the initial "evidence" paper only "di-lepton" and "lepton + jet + b-tag" channels were analyzed.

The "lepton + jet" channel has a high p_T lepton, missing energy and two jets from the W decay plus two b jets. Initially, the kinematic cuts, number of jet cuts, plus jet mass cuts did not find enough events to significantly beat the background to claim "evidence". The identified sample of 52 events can be further constrained by a SVX b-tag down to six events with W plus three or more jet events with a b-tag above 2.3 ± 0.26 expected background events. An alternative b-tagging method by identifying "soft" leptons from ($b \to l\nu_l X$ or $b \to c \to l\nu_l X$) with the calorimeter and CTC will not be elaborated here. In summary the "lepton + jet" event type, particularly with a b-tag, is considered the most powerful signature due to a combination of cleanliness and statistics. To improve this method a longer SVX is needed as realized in the RUN II upgrade. An event display is shown in Fig. 5.23 – The Golden TOP event.

Thomas Müller said, when asked about the physics success of the CDF experiment

> On a conference in 1983, the theorist John Ellis from CERN instructed us experimental particle physicists to find ways for identifying experimentally beauty quarks in particle jets. Ten years later, first evidence for top quark signals was obtained in the CDF experiment using the reconstruction of beauty hadron decays. Any claims for the observation of top quarks would have been very doubtful without the excellent performance of the silicon vertex detector.

There are many more exciting physics subjects like CP violation in the B sector to determine the CP violating phase β_s or B_S^0–$\bar{B}_S^0$ oscillations [8, 289]. Spectroscopy (mass, lifetime, decays) of B_c, B_s, Λ_b, Ξ_b, Σ_b, orbitally excited B-mesons, etc. The SVT enhanced especially the B spectroscopy. It highly enriched the statistical sample

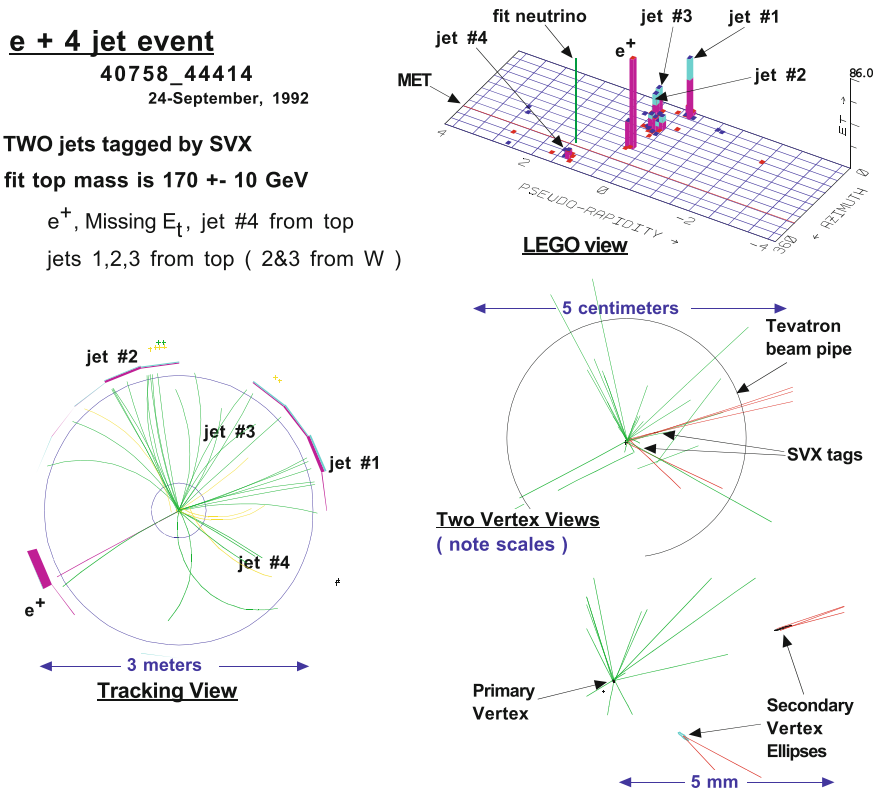


Fig. 5.23 A "Golden" t event. $t\bar{t}$ decaying into W^+b, $W^-\bar{b}$, where one W decays leptonically with the signature lepton ID plus missing energy, the second W decays into $q\bar{q}$ resulting in two jets together with the initial two tagged b-jets. In total one lepton, four jets, two tagged b jets and missing energy were reconstructed [332]

on hadronic B-decays, which have an intrinsic good p_T resolution (no neutrino). The SVT was the key for the B_s measurement. The field is rich. First observation of charged Σ_b and first direct observation of Ξ_b at the TEVATRON plus some more examples are discussed in [245].

Additionally, for some topics a simple b flightpath tag is not enough, e.g. for all the CP and oscillation studies. The correct flavour must be known. To "tag" the true flavour of the B-meson the charge must be known. Two methods were developed: (1) The "opposite side tagging" where the jet charge or the lepton charge ($b \to l^-$ but $\bar{b} \to l^+$) is evaluated on the "other" B-hadron to determine the B-hadron flavour of interest. (2) The "same side tagging" method where the charges of the particle of associated production are determined, e.g. $\pi^+(\pi^-)$ is associated with the formation of a $B(\bar{B})$ meson. The mechanisms were developed during RUN I and perfected in RUN II. As mentioned earlier, the purpose of vertex detectors is simply to measure precise tracks and tag the flavour of heavy quarks.

Chapter 6
CMS: Increasing Size by 2 Orders of Magnitude

The **C**ompact **M**uon **S**olenoid CMS operates at the **L**arge **H**adron **C**ollider LHC at CERN. It was conceived to study proton–proton (and nucleon–nucleon) collisions at a centre-of-mass energy of 14 TeV (5.5 TeV per nucleon for Pb-Pb) and at luminosities up to $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} (10^{27} cm^{-2}s^{-1}). The detector is characterized by high hermeticity and compactness. A 3.8 T superconducting solenoid surrounds a silicon tracker, an **e**lectromagnetic **cal**orimeter ECAL and a **h**adron **cal**orimeter HCAL. The return yoke is instrumented with muon detectors covering most of the 4π solid angle. Forward sampling calorimeters extend the CMS pseudo-rapidity coverage to high η values. The detector is fully described in [4, 290, 292, 299]. Its layout can be seen in Fig. 6.1. The CMS detector was fully constructed on the surface and was lowered 100 m down to the cavern in 11 slices plus 2 very forward calorimeters (**H**adron **F**orward HF). Individual slices weighted up to 2500 t summing up to a weight of 14,000 t. The inner detectors like calorimeters and tracking detectors were inserted and instrumented later down in the cavern. A schematic how CMS identifies (IDs) basic particles is placed in the Appendix in Fig. B.3. In the following the CMS silicon tracker (Phase 0 until 2016) will be described with an update section describing the pixel upgrade installed in Spring 2017.

December 1999, on the eve of the last millennium, the CMS Tracker Collaboration decided to construct its tracker fully out of silicon, summing up to 206 m^2 silicon surface.[1] The design goal of the tracker is to operate with a maximum of 1 – 3% occupancy at all layers and a good impact parameter resolution in $R\phi$ and Rz.

[1] Prior to this meeting about 70 m^2 silicon plus ∼200 m^2 **M**icro**S**trip **G**as **C**hamber MSGC were specified in the initial **T**echnical **D**esign **R**eport TDR [290].

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_6

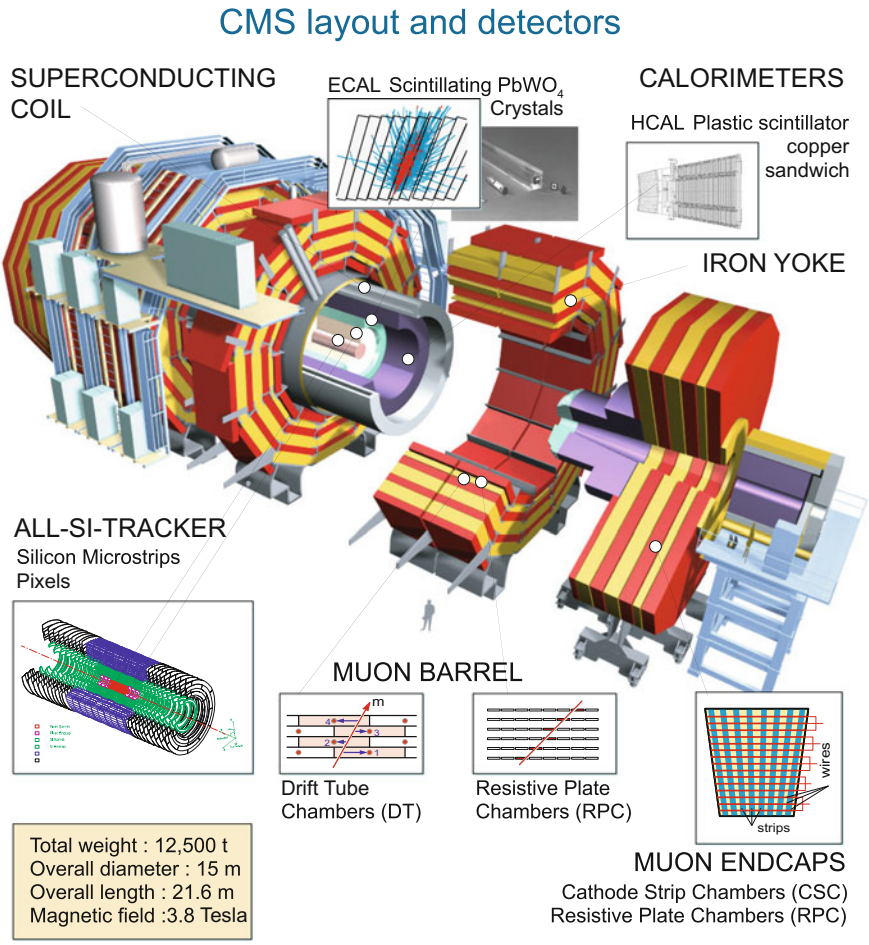


Fig. 6.1 The CMS detector and its components [Courtesy of CERN, 333]

The requirements are

- efficient and robust pattern recognition
- fine granularity to resolve nearby tracks; e.g. from 3-prong τ-lepton decays
- fast response time to resolve bunch crossings
- ability to reconstruct tracks from very low p_T in the order of 100 MeV up to 1 TeV
- ability to reconstruct heavy objects translating to $1-2\%$ p_T resolution at $\sim$100 GeV
- ability to operate in a crowded environment at 10 cm from primary vertex
- ability to tag b/τ through secondary vertices
- reconstruction efficiency: 95% for hadronic isolated high p_T tracks and 90% for high p_T tracks inside jets
- ability to operate in a very high radiation environment
- efficient and precise tracking in the pseudo-rapidity range $|\eta| \leq 2.5$.

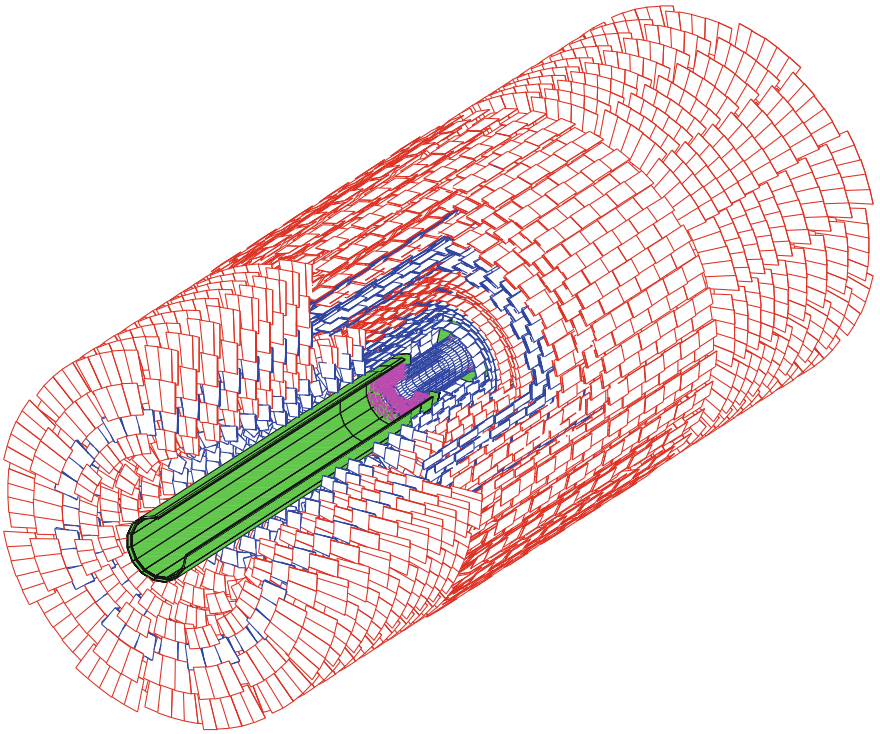


Fig. 6.2 The CMS all-silicon tracker layout (figure displays the detector throughout 2008–2016) [129]. The pixel detector is located in the centre. It is surrounded by 10 strip barrel layers supplemented by 3 small forward and 9 large forward disks. The barrel layer 1, 2, 5, 6 and endcap rings 1, 2, 5 consist of double modules, mounted back-to-back with a stereo angle of 100 mrad. The overall length and diameter is 5.4 and 2.4 m, respectively. 1632 pixel (1440 in phase 0) and 15148 strip detector modules build up the tracker. All in all 127 million pixel cells (66 million in phase 0) and 10 million strips are available. The tracker directly surrounds the beam pipe and is enclosed by the electromagnetic calorimeter; no additional inner tracking device like a drift chamber or transition radiation tracker is installed [128]

The result is a small pixelated detector in the centre with a large volume strip detector surrounding it. All in all a cylinder of 2.5 m diameter and 5.4 m length is instrumented with 206 m^2 of active silicon detectors. The general 3D layout can be seen in Fig. 6.2. At the design luminosity of $10^{34}\,cm^{-2}s^{-1}$ the tracker will see above 500 particles from about 20 overlapping proton–proton collisions per bunch crossing, every 25 ns. With a homogeneous field of 3.8 T, low momentum particles spiral at low radii and therefore the occupancy decreases faster than $1/r^2$ as it would without a magnetic field. Fast readout and high granularity are needed to achieve the low design occupancy, get a reasonable track separation, good primary vertex resolution and good impact parameter resolution σ_{d_0}.

The CMS tracker as of the year 2017, is composed of a pixel detector with 4 barrel[2] layers and a silicon strip tracker with 10 barrel detection layers extending outwards to a radius of 1.1 m. Each system is completed by endcaps which consist of three disks in the pixel detector and three small plus nine large disks in the strip tracker on each side of the barrel extending the acceptance of the tracker up to a pseudo-rapidity of $|\eta| < 2.5$. A schematic with detailed individual module positions can be found in the Appendix in Fig. B.5.

6.1 The CMS Pixel Detector – Phase 0 – 2008 – 2016

In order to keep the occupancy at or below 1%, pixelated detectors have to be used at radii below 10 – 20 cm. For a pixel size of $100 \times 150\,\mu\text{m}^2$ in $R\phi$ and Rz, respectively, which is driven by the desired impact parameter resolution, the occupancy is of the order of 10^{-4} per pixel at LHC bunch crossing rate. The pixel sensors are discussed in Sect. 6.4.2. With analogue[3] signal interpolation, a hit resolution of 10 $(R\phi) \times 20\,\mu\text{m}(Rz)$ can be achieved (see also Sect. 6.6.2). The magnetic field of CMS, inducing a Lorentz angle, enhances the charge spread to more than one cell, thereby increasing resolution by interpolation of charges. While the magnetic field is naturally perpendicular in the barrel region, the forward region is built in a turbine-like geometry to profit from the effect (see Figs. 6.3 and 6.4). Due to the very high number of channels, the pixel readout is zero-suppressed (sparsified) and only for pixels being hit, analogue data along with the pixel cell address is being sent to the data acquisition. The data rate is dynamic, depending on beam background and instantaneous luminosity. With limited bandwidth (inside the chips themselves) at very high rates not all information can be sent out and an inefficiency in the order of a couple of percent is expected at nominal LHC luminosity ($\mathcal{L} = 10^{34}\ \text{s}^{-1}\text{cm}^{-2}$) and the maximum Level-1 trigger rate of L1 = 100 kHz in the innermost layer.

Three pixel barrel layers are located at radii 4.4, 7.3 and 10.2 cm. The pixel barrel is 53 cm long. Two pixel forward disks at $z = \pm 34.5$ and $z = \pm 46.5$ cm guarantee 3D space points coverage up to $|\eta| < 2.5$. With a total area larger than 1 m^2 the CMS pixel detector is as large as the complete DELPHI silicon vertex detector. The pixel detector layout is displayed in Fig. 6.3. It is instrumented with 768 hybrid pixel modules in the barrel plus 672 modules (plaquettes) in the forward disks for a grand total of 66 million pixels.

Figure 6.5 shows a very interesting photo demonstrating the complexity of tracking detectors in the very tight spaces of a **H**igh **E**nergy **P**hysics HEP Experiment. It displays the so-called supply tube since it supplies the detector with all necessary services: cooling pipes, control fibres, power for electronics, and sensor high voltage, also the electrical to optical converters are situated here. The readout fibres are prominently visible since the other parts are buried beneath. A very complex object!

[2] It was 3 barrel layers and 2 forward disks from 2008 to 2016.

[3] Analogue data represents pulse height to allow a centre-of-charge analysis later.

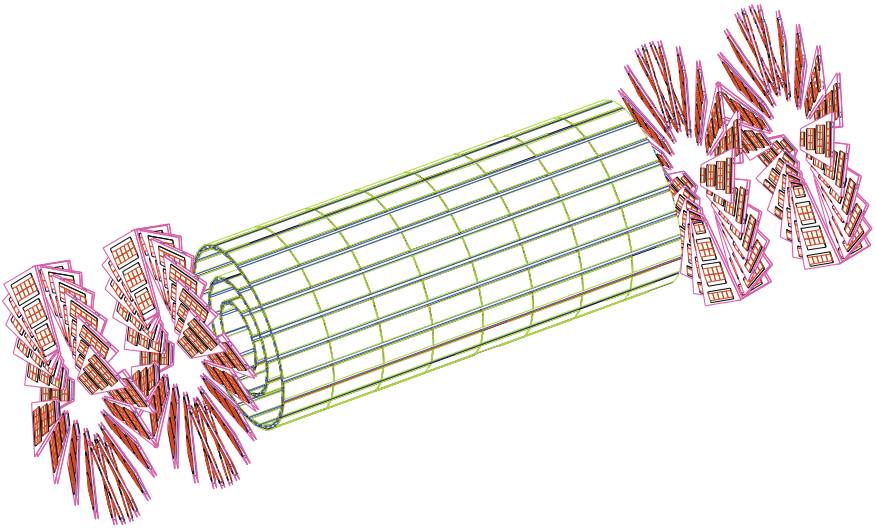


Fig. 6.3 The CMS pixel detector layout. Three 53 cm long barrel layers at radii 4.4, 7.3 and 10.2 cm plus two forward disks at z = ±34.5 and z = ±46.5 cm guarantee the 3D space points coverage up to $|\eta| < 2.5$. The barrel (forward) pixel detector contains 768 (672) pixel modules (plaquetts) with 48 million (18 million) pixel cells covering 0.78 (0.28) m^2 [87]

Fig. 6.4 One of the forward half-disks. Please note the turbine configuration emphasising the Lorentz angel spread. To achieve a wedge geometry seven different sensor geometries, assembled on so-called panels, haven been produced. Eight half-disks, forming 4 full disks, 2 at each end, finally surround the beam pipe [Courtesy of CMS/CERN/FNAL]

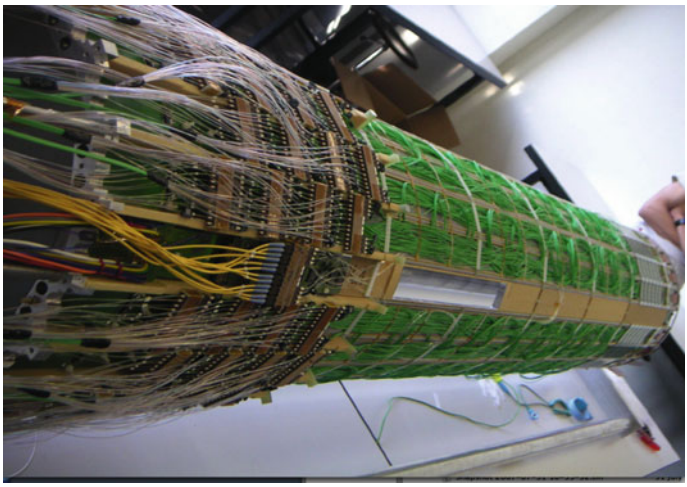


Fig. 6.5 The supply tube of the CMS Pixel Detector (2008 – 2016) [Courtesy of CMS/CERN/PSI]

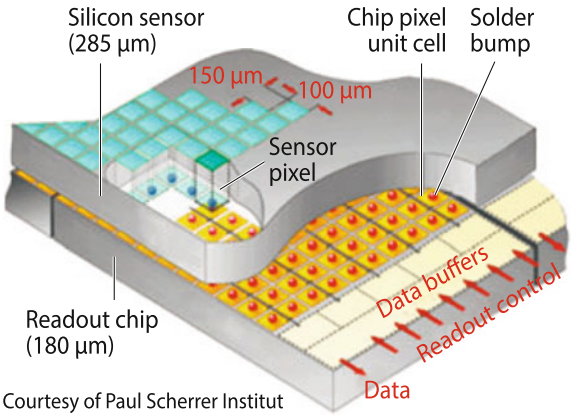


Fig. 6.6 Schematic of the CMS pixel with 100 $\times$150 μm^2 pixels per cell. A good example of a hybrid active pixel sensor *HAPS* providing intrinsic two-dimensional position information. The electronics amplifier chip has the same size and channel pattern as the sensor and is bump bonded (flip-chip bonded) to the sensor. A more detailed schematic can be found in the next Fig. 6.7. A photo of the bump lithography is presented in Fig. 1.73 on p. 108

To quote my colleague and CMS Pixel coordinator Roland Horisberger:

> The supply tube is the place where all troubles meet.

The very reliable and radiation tolerant Hybrid Active Pixel Sensor HAPS technology has been chosen. The schematic of the electronic-sensor stack can be seen in Fig. 6.6 and a full pixel module including a module assembly picture is shown in Fig. 6.7.

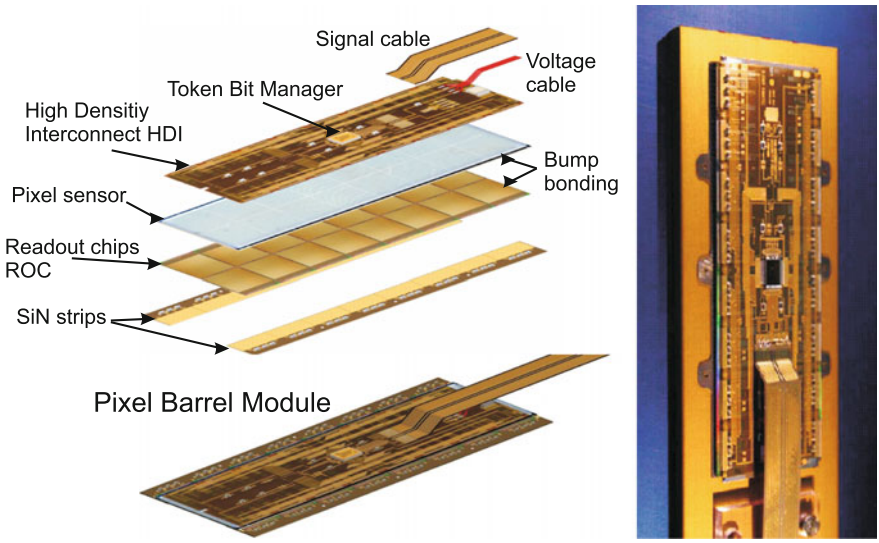


Fig. 6.7 CMS Pixel module – barrel type [Courtesy Paul Scherrer Institute, CERN]

The pixel detector is essential for b & τ-tagging and primary vertex reconstruction. It also determines the track seeds towards the outer strip tracker. The detector is designed and built in half-shells allowing installation with the central LHC beam pipe[4] in place. Also the services are done with fast extraction and installation in mind. The reason is to allow maintenance and repair in a typical LHC **Y**ear-**E**nd-**T**echnical **S**top YETS and the replacement of the innermost layer (of the phase 0 detector) was initially foreseen 2 – 3 years into the run due to high radiation damage at low radius. In reality, the detector had been removed and repaired in the YETS 2009 and also in long shutdown 1 2013/2014. This is illustrated impressively with the photos Figs. 6.8 and 6.9 of the barrel and forward pixel detector installed inside CMS. Instead of replacing the innermost layer, a completely new detector (Phase I, see later) has been installed in spring 2017.

6.2 The Pixel Phase I Upgrade – Installed February/March 2017

As indicated in the previous section, the pixel detector design allows a replacement of its innermost layer during the lifetime of CMS and the initial plan was to replace just the innermost layer when radiation renders it dis-functional. In the end, the reason to replace the full detector was not driven by radiation damage but by the

[4]Installation of the beam pipe requires a high temperature bake-out to allow a good final vacuum which takes several weeks.

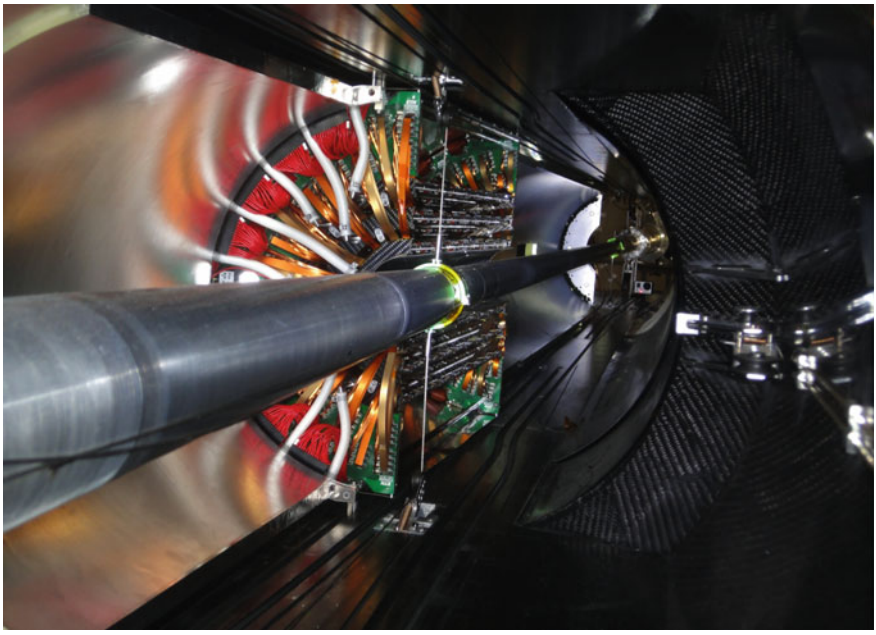


Fig. 6.8 Half of the Barrel Pixel detector just placed into the heart of CMS – December 2014 at the end of Long Shutdown 1. One can see the three layers and the end flange where the cooling pipes fan out. The half-shell concept allows fast extraction and insertion WITH the central beam pipe of the LHC in place [Courtesy of CMS/CERN]

"beyond-LHC-design" increase of instantaneous luminosity achieved by the LHC. As described earlier, the numbers of on-chip buffers is limited and would lead to too high inefficiencies at very high luminosities, and also the data transfer from detector to the back-end is strictly limited. Due to these reasons, CMS decided to replace the full pixel detector instead of only the innermost layer. At the time of writing, the new detector has just been installed.

This is done in a conservative way but still using optimised design and readout electronics. The newly optimised readout chips are performing much better and are also more radiation tolerant – even the innermost layer will withstand an integrated luminosity of $\mathcal{L}$ =500 fb^{-1} instead of the originally anticipated 250 fb^{-1}. The full information can be found in [309] and [304] while only a very brief summary is presented here, focussing on interesting facts demonstrating the evolution in the field. With a new smaller beam pipe[5] in place, it was possible to move the inner layer closer to the interaction point. A fourth layer has been added at higher radius improving the track link towards the outer strip tracker. Layer 2 and 3 radii are almost unchanged. The four barrel layers (BPIX) are now located at radii 3.0, 6.8, 10.9 and 16.0 cm; they span ±274.4 mm. The forward disks (FPIX) are moved further in

[5]New outer diameter D = 45 mm with respect to old D = 59.6 mm.

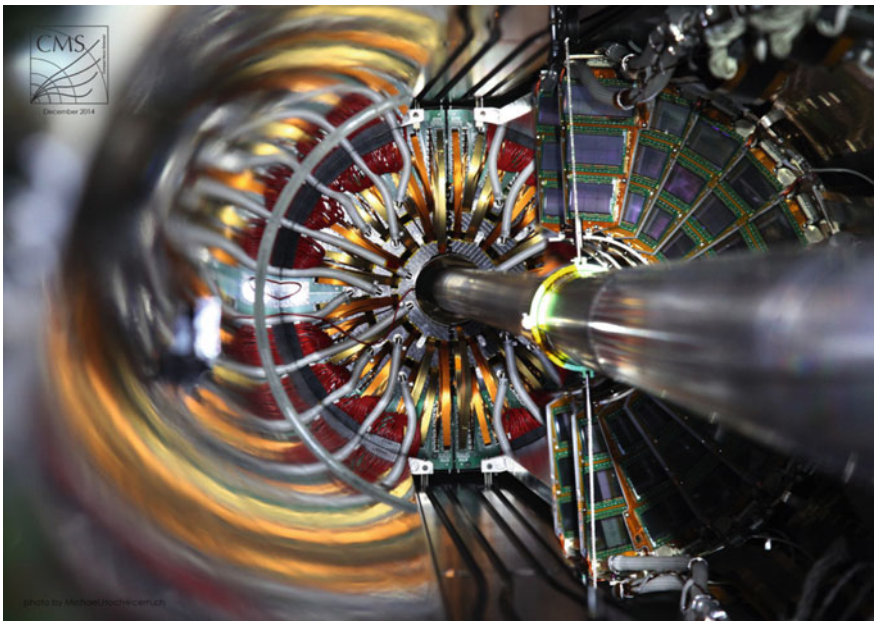


Fig. 6.9 The full Barrel Pixel detector is in place and half of one end of the forward pixel detector is in place (December 2014). The individual pixel modules are visible. The module wedge form has been achieved using sensors of different sizes, housing a different number of ROCs (6, 8 or 10) [Courtesy of CMS/CERN]

and are now located at z = ±29.1, ±39.6 and ±51.6 cm; they span the radius from 4.5 to 16.1 cm. The whole detector (barrel and forward) features now one single module geometry for simplicity. The forward modules are still placed into a turbine structure to profit from the Lorentz angle. Due to this, the two mechanical structures for the positive and negative detector end, are mirrored instead of simply rotated, otherwise the Lorentz angle would shift in the wrong direction in one end. The pixel cell count increases from 18 to 48 and 48 to 79 million in the FPIX and BPIX respectively. Table 6.1 gives details about barrel and endcap configurations also compared to the first CMS pixel detector – Phase 0.

The layout of the new detector is displayed in Fig. 6.10. A photo of the forward and barrel part can be seen in Fig. 6.11.

The new detector design addresses the following points:

- increase data transmission by changing from 40 MHz analogue transmission to 320 MHz digital transmission
- increase on-chip buffers to limit the inefficiency and therefore to cope with $\mathcal{L} = 2 \cdot 10^{34}\ \mathrm{s}^{-1}\mathrm{cm}^{-2}$
- improve radiation tolerance of the readout chips (ROCs)

Table 6.1 The table expresses the main difference between the Phase 0 and the Phase I detector. Refer also to Figs. 6.3 and 6.10 displaying the corresponding layouts. The 2nd number in the Phase 0 module column refers to the half-modules. Impressively, the material budget at $\eta = 0$ is lower for phase I despite having one more layer

	Phase I, since Spring 2017			Phase 0, 2008 – 2016		
Cooling	CO_2 (bi-phase)			C_6F_{14} monophase		
Powering	DC-DC powering			direct power		
Readout	320 MB/sec (digital coded)			40 MHz (analogue coded)		
Pixel size	$100 \times 150\,\mu m$			$100 \times 150\,\mu m$		
Material $\eta = 0$	$X/X_0 = 5.5\%$			$X/X_0 = 6\%$		
Material $\lvert\eta\rvert = 1.6$	$X/X_0 = 20\%$			$X/X_0 = 40\%$		
Layer	Radius (mm)	Facets	# Modules	Radius	Facets	# Modules
Beampipe	22.5			29.8		
4	160	64	512	–	–	–
3	109	44	352	102	46	320/32
2	68	28	224	73	30	224/32
1	30	12	96	44	18	128/32
total			1184			672/96
Disk	Radius 45 – 161 mm			Radius 60 – 150 mm		
	z pos (mm)	Mod outer	Mod inner	z pos	Sensors/blade	# Blades
1	$\pm$291	2*34	2*22	$\pm$345	7	2*24
2	$\pm$396	2*34	2*22	$\pm$465	7	2*24
3	$\pm$516	2*34	2*22	–	–	–
Total		448				96

- improve the hermiticity of the detector and establish an overall 4-hit coverage to increase redundancy and allow for a quadruplet track finding algorithms and therefore better track seeding capability. This highly improves the track quality/resolution and pattern recognition
- reduce mass wherever possible especially in the tracking volume $|\eta| \leq 2.5$
- reduce the innermost radius to improve impact parameter resolution σ_{d_0}
- add a layer at higher radius to significantly improve the impact parameter resolution in z direction due to the larger lever arm (the strip detector at higher radii has only a good resolution in φ not in z)

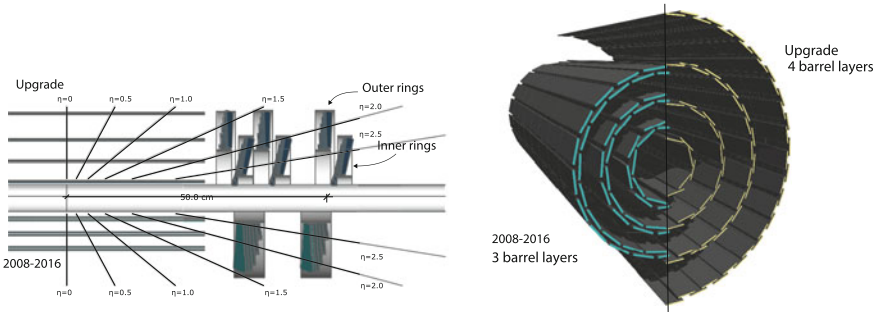


Fig. 6.10 Layout of the CMS phase I pixel detector. One can see the increase from 3 to 4 layers in the barrel and from 2 to 3 disks in each endcap. In addition the inner layer moves *lower* in radius thus closer to the primary interaction point [304]

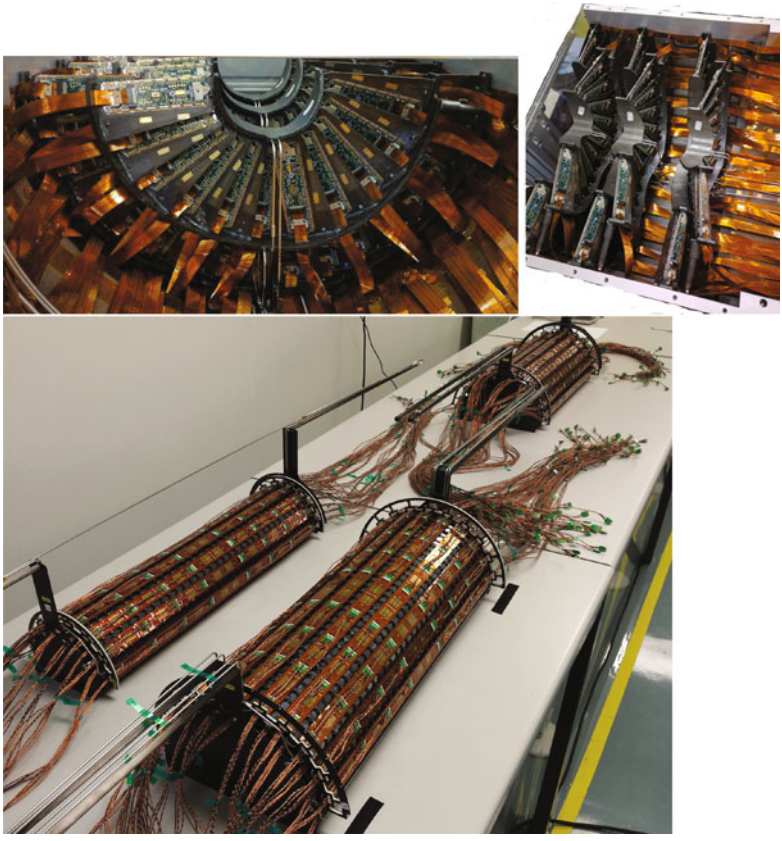


Fig. 6.11 The *top left* photo shows the first assembled forward disks of the Phase I FPIX system. The turbine arrangement as in the first detector has been kept but the modules are of one single type; compare with Fig. 6.9. The *top right* photo displays three half-disks installed in a half-cylinder. In the *lower* part, you can see three out of four layers of the barrel detector during their assembly [Courtesy of CERN, Fermilab and PSI]

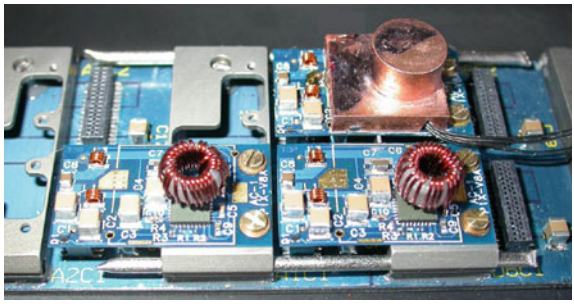


Fig. 6.12 Photo of a DC-DC converter of the CMS phase I pixel detector [304]. The Figure displays the DC-DC with final shield (*upper right*) and with bare air coil (*lower part*) for illustration

The challenge is to send more than double the amount of data through the same number of optical fibres and to supply more power through the same number of power cables – it was not possible to lay more services.[6] The data throughput challenge is solved by changing from 40 MHz analogue transmission to 320 MHz digital transmission. The new chip features a fast 8-bit ADC retaining the ability to send pulse-height information to allow centre-of-charge calculation offline. As said earlier, the data and time stamp buffers have been increased to drastically decrease inefficiencies. DC-DC power converters are located in the supply tube. They allow higher input voltages, thereby increasing power without increasing electric current thus avoiding a higher cross-section of the cables [100]. The concept is that of buck converter[7] but dedicated radiation tolerant chips have been developed by CERN. A picture of the DC-DC converters (prototypes) with and without shielding the air coil is presented in Fig. 6.12.

The main ingredients to reduce mass, thus minimizing multiple scattering, are:

- move all service electronics, e.g. opto-hybrids out of the tracking volume into the supply tube
- use ultra low mass mechanics
- use bi-phase CO_2 cooling utilising smaller pipes with thinner pipe-walls

As result the total mass of the pixel barrel is halved despite placing about twice the amount of modules. The result of all mass reduction efforts can be seen in Fig. 6.13.

The n-in-n DOFZ sensors, used already by the first CMS pixel detector, are radiation tolerant enough – the exact same concept has been re-used. Interestingly, the 4 in. masks for the barrel sensors are physically *the same* as for the first production and have simply been re-used at the company. The forward sensor cell and full sensor design did also not change conceptually but the production was done using 6 in. instead of 4 in. technology; fitting eight instead of three sensors onto one wafer. The decision to have a single sensor/module geometry simplified the system a lot.

[6] New cooling pipes, compatible with the high pressures of a CO_2 cooling system (triple vacuum jacketed steal pipes) have been installed during long shutdown 1; two years in advance of the pixel installation.

[7] A buck converter is a voltage step down converter. It is a switching converter storing the energy in an inductor. Due the CMS magnetic field only air coils can be used.

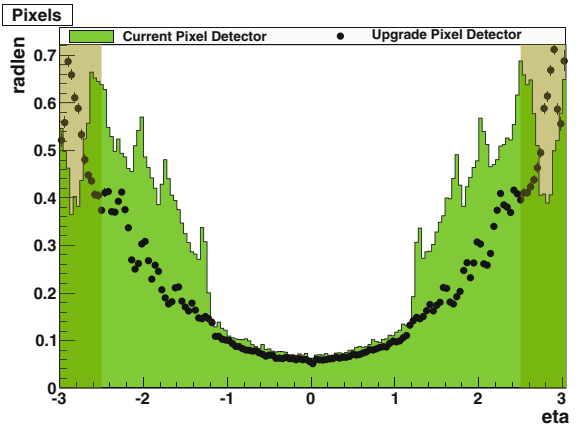


Fig. 6.13 Radiation Lengths of the CMS phase I pixel detector. Clear reduction of mass in the relevant (*non-shaded*) region below $|\eta| = 2.5$ [304]

Section 6.3 gives more details about the sensors, where also Fig. 6.25 shows a photo of the real wafers scaled down by a factor of two.

6.3 The CMS Silicon Strip Tracker – SST

The CMS **S**ilicon **S**trip **T**racker SST spanning radii 20 cm $< r <$ 110 cm plus a substantial forward region is subdivided into **T**racker **I**nner **B**arrel TIB, **T**racker **I**nner **D**isk TID, **T**racker **O**uter **B**arrel TOB and **T**racker **E**ndcap TEC. The sub detector designs are different to match the requirements at the different radii—a bit like the manifold solutions as for DELPHI and CDF II. The different sub detector locations can be seen in Fig. 6.14, where a quarter section of the detector is presented.

Tracker Inner Barrel TIB

The **T**racker **I**nner **B**arrel TIB has four barrel layers assembled in shells, layers 1 and 2 are double-sided DS plus three inner Disks on both detector ends. The TIB is complemented by two disk-shaped forward detectors TID each composed of three small disks. The TIB/TID spans the radii 20 cm $< r <$ 55 cm with a length of 2.2 m($|z| <$ 110 cm). The TIB alone is 1.4 m long ($|z| <$ 65 cm). All modules are mounted on "strings" which are directly and individually placed on the structures. The reduced particle flux density allows the use of strip sensors with a typical cell size of 10 cm $\times$ 80 µm and 10 cm $\times$ 120 µm resulting in an occupancy of up to 1 – 3% per strip at LHC bunch crossing at nominal instantaneous luminosity. Modules are limited to one single sensor to accommodate for the maximum strip length. A photo of the inside of TIB can be found in Fig. 6.15 and a picture of a TID part in Fig. 6.16.

Tracker Outer Barrel TOB

The **T**racker **O**uter **B**arrel TOB structure where the modules are assembled in six concentric layers (TOB layers 1 and 2 are DS; globally Layer 5 and 6) surrounds

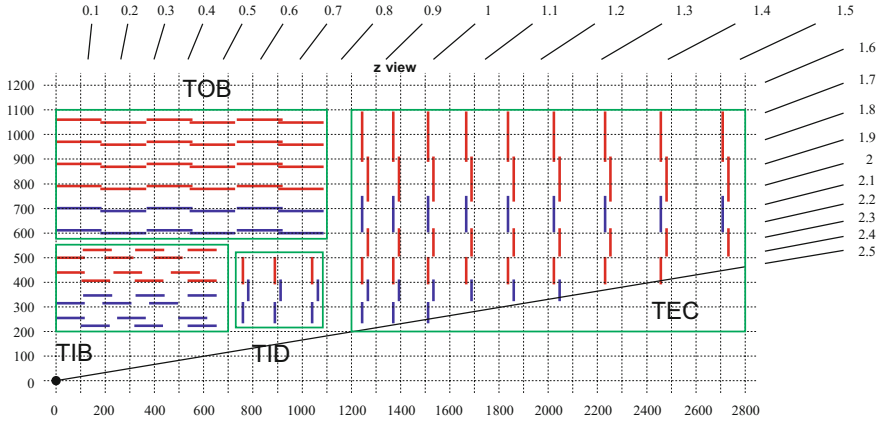


Fig. 6.14 The CMS strip tracker layout. At intermediate radii (20 cm $<$ r $<$ 55 cm) the reduced particle flux allows the use of silicon microstrip detectors with a typical cell size of 10 cm $\times$ 80 μm, resulting in an occupancy of up to 1 – 3% per strip at LHC bunch crossing. In the outer region (55 cm $<$ r $<$ 110 cm) the strip pitch and length can be further increased to have cell size 20 cm $\times$ 180 μm. The detector is consequently instrumented with barrel geometry (*TIB* and *TOB*) including some stereo modules, complemented with the forward inner disks (TID) and the large endcaps on both sides (TEC). Barrel layer *1*, *2*, *5*, *6*, TID ring *1*, *2* and TEC ring *1*, *2* and *5* are instrumented with stereo double-sided modules

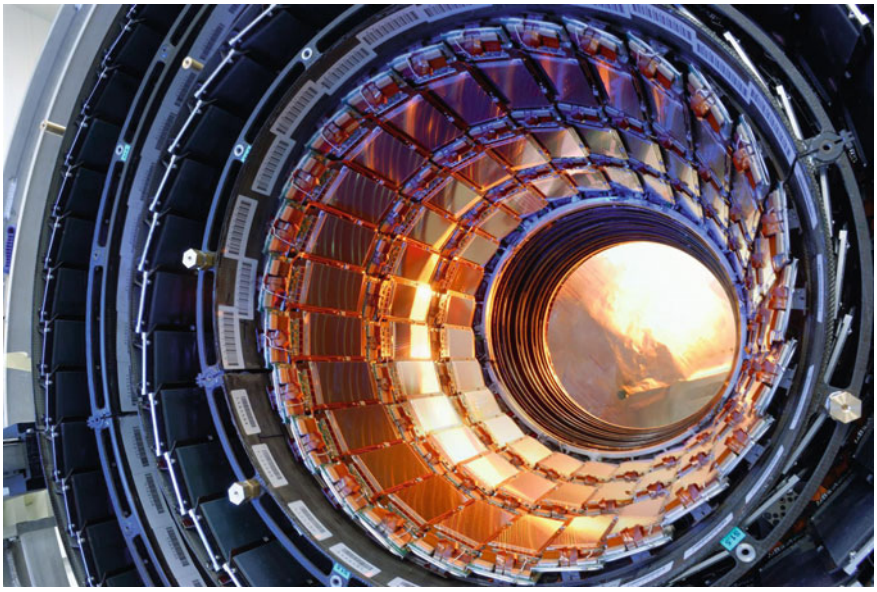


Fig. 6.15 **T**racker **I**nner **B**arrel *TIB*. Three layers of tracker inner barrel detector are visible. In Fig. 6.18 it is inserted into the TOB. Finally, it will house the pixel detector. In the background the inner edges of the three **T**racker **I**nner **D**iscs can be seen [Courtesy of CERN]

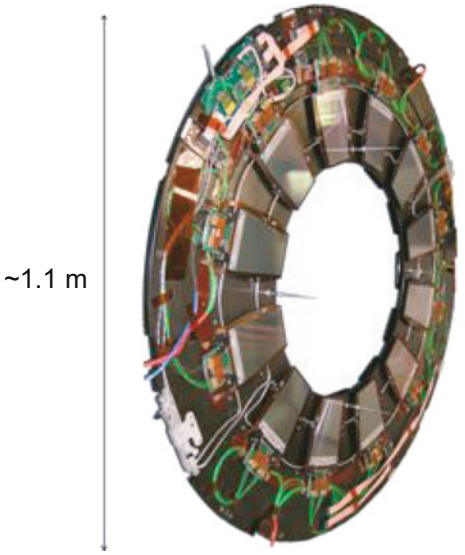


Fig. 6.16 One disk of the **T**racker **I**nner **D**isk TID. Three of these disks on both TIB sides complete the inner detector fully embedded in the TOB [269]

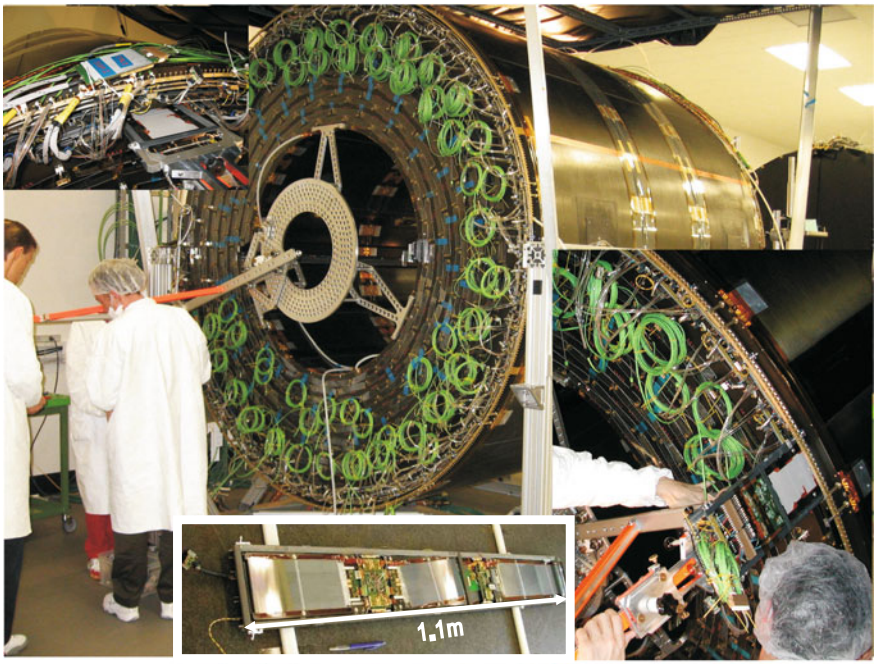


Fig. 6.17 Rod insertion into TOB. The main picture gives a good overview of the size and complexity of the TOB. A special tool guarantees a safe insertion. A rod is displayed at the *bottom* [128]

Fig. 6.18 TIB inserted in TOB, a rare view. The four layers of the tracker inner barrel are visible inside the tracker outer barrel. The TOB layers 5 – 10 are completely hidden by the necessary cabling. This is a rare view during the time of construction

the TIB. It is located inside the electromagnetic calorimeter. The TOB spans radii 55 cm $< r <$ 110 cm and a length of 2.2 m($|z| <$ 110 cm). For a detector of the size of the TOB an individual module placement is not practical anymore. A structure was designed where 688 one metre long "rods" equipped with modules are inserted. At these radii the pitches and strip lengths can be increased, which helps to moderate the number of readout channels, given the large area. The cell size is 20 cm $\times$ 122 μm and 20 cm $\times$ 183 μm with two sensors daisy-chained to one readout hybrid. The daisy-chaining has obvious advantages but also increases the channel capacitance load, and thus the noise. This is compensated by increasing the thickness from 320 μm ("thin") for TIB/TID to 500 μm ("thick") for TOB sensors. The same argument and design holds for TEC where inner rings 1 – 4 are "thin" and ring 5 – 7 sensors are "thick". An important detail is the smaller pitch in the two outermost layers reflecting the need to have a precision space point at a large lever arm (Fig. 5.3 on p. 199) and a precision link point to the calorimeters. For TIB and TOB, having a barrel-like structure, sensors are rectangular and strips run parallel to the beam z-axis. A double-sided module is composed out of two single-sided sensors mounted back-to-back where one is tilted by an angle of 100 mrad with respect to the z-axis. A photo of the TOB and a rod can be seen in Fig. 6.17. Figure 6.18 displays how the TIB is embedded in the TOB.

Tracker End Cap TEC

Two **t**racker **e**nd**c**aps TEC ensure a pseudo-rapidity coverage of $|\eta| \leq 2.5$. The strip orientation is radial, to achieve a continuity of all strips running radially through all rings from ring 1 to 7-sensors at different radii. Seven different types are necessary. All sensors are wedge type and for modules with two daisy-chained sensors, the outer pitches of e.g. sensor W5A must match inner strip pitches of sensor W5B for all individual strips, refer to Table 6.2. The detectors of ring 1, 2, 5 are made of double-sided modules. Modules of ring 1 – 4 contain a single sensor while ring 5 – 7 have a larger pitch and two sensors are daisy-chained together. As for the TOB the modules are arranged on substructures – on wedge-shaped "petals". Petals hold up to 7 rings of modules with the different geometries. One TEC consists of 9 disks populated with petals, each covering 1/16 of 2π. There are eight different petal types varying in radial length, hence number of populated rings, to adapt for the location in z. Petal types differ for disks 1 – 3, 4 – 6, 7 and 8 and 9, the arrangement can be seen in Fig. 6.14. A total of 288 petals form both endcaps. The detector spans $120\,\text{cm} < |z| < 280\,\text{cm}$ and $20\,\text{cm} < r < 110\,\text{cm}$. An impressive photo of the TEC can be seen in Fig. 6.19.

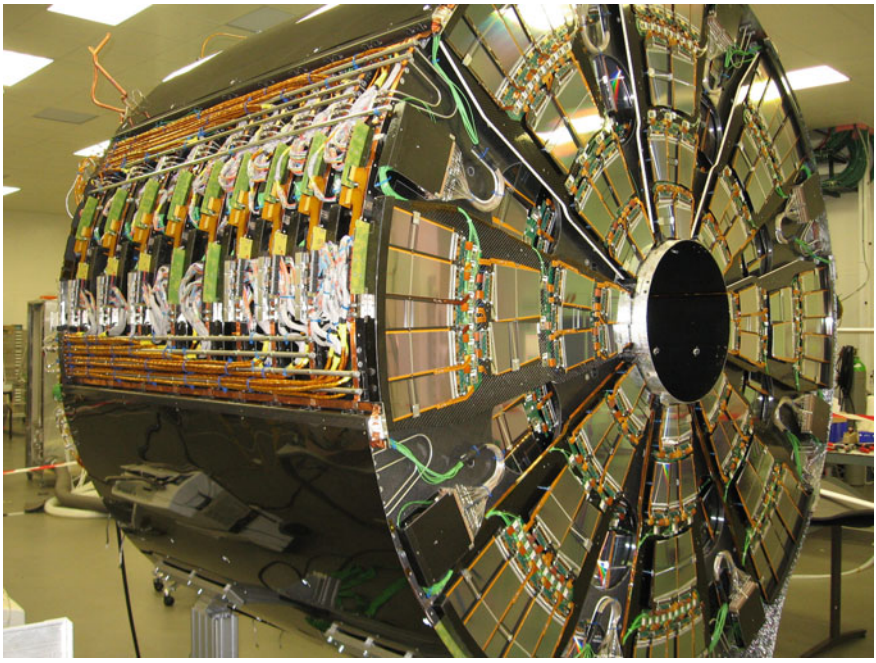


Fig. 6.19 The **t**racker **e**nd**c**ap TEC, fully equipped and open, a rare view. Eight petals are visible on the front, the space in-between is covered with petals on the other non-visible side of the disk. Nine similar disks form one tracker endcap. The 96 visible wedge-shaped sensor modules placed in a wedge geometry on the petals guarantee a continuous-radial strip orientation. On the *left* an opening shows all the services for one sector, out of eight. The diameter of one TEC is 2.5 m while the length stretches to 1.7 m. One TEC holds 144 petals. All electrical and optical cables and cooling pipes were installed in autumn 2005. Both TECs were fully ready, with all petals installed and inserted into the tracker support tube early 2007 [Courtesy of CMS]

Table 6.2 The different sensor types

Sensor type	@ Layer	Trapezoidal a b h (cm)	Stereo module	Pitch (μm)	No. of strips	Thick (μm)	Resistivity (kΩcm)	Sensors module
IB1	1, 2	63.4 63.4 119.2	YES	80	768	320	1.5 – 3	1
IB2	3, 4	63.4 63.4 119.2		120	512	320	1.5 – 3	1
OB1	9, 10	96.4 96.4 94.5		122	768	500	3.5 – 7.5	2
OB2	5, 6, 7, 8	96.4 96.4 94.5	L5,L6	183	512	500	3.5 – 7.5	2
W1TID		63.6 93.8 112.9		80 – 119	768	320	1.5 – 3	1
W1		64.1 88.1 89.5	YES	81 – 112	768	320	1.5 – 3	1
W2		88.2 112.4 90.3	YES	113 – 143	768	320	1.5 – 3	1
W3		65.0 83.2 112.8		124 – 158	512	320	1.5 – 3	1
W4		59.9 73.4 117.4		113 – 139	512	320	1.5 – 3	1
W5A		99.0 112.4 84.0	YES	126 – 143	768	500	3.5 – 7.5	+W5B
W5B		112.4 123.0 66.1	YES	143 – 156	768	500	3.5 – 7.5	+W5A
W6A		86.1 97.5 99.0		163 – 185	512	500	3.5 – 7.5	+W6B
W6B		97.5 107.6 87.8		186 – 205	512	500	3.5 – 7.5	+W6A
W7A		74.1 82.9 109.8		140 – 156	512	500	3.5 – 7.5	+W7B
W7B		82.9 90.0 98.8		157 – 172	512	500	3.5 – 7.5	+W7A

The IB stands for **i**nner **b**arrel, the OB for **o**uter **b**arrel and the W for **w**edge, the sensors in the endcap. a, b defines the width and h the height of the sensors. Note that the OB1 sensors with the narrower pitch are located in the outermost layers to have a precise measurement with a long lever arm. All OB and W5–W7 sensors are 500 μm thick while all IB and W1–W4 are 320 μm thick

Detailed descriptions of the SST are also to be found in [128, 129, 292, 315]. Table 6.2 informs about all important sensor geometry parameters.

Appreciation of the full CMS detector size and the relation to the tracker can be gained by looking on Fig. 6.20 where the tracker is on the way to its final position in the heart of CMS – December 2007.

As a result from all necessary services, like cables, cooling pipes, support structures and electronic boards, the material budget is much higher than optimal and it is one of the issues which needs to be improved for any future detector. Figure 6.21 shows the material budget of the CMS tracker in units of radiation length. It increases from 0.3 X_0 at $\eta \approx 0$ to about 1.7 X_0 at $|\eta| \approx 1.4$, beyond which it falls to about 1 X_0 at $|\eta| \approx 2.5$. It is quite obvious that cables and support structures carry most of the weight, also the interface region between sub detectors, between barrel and forward region are prominently too massive.

Nevertheless, the designs described above lead to excellent performance underlined here with the track resolution plots for single traversing muon tracks in Fig. 6.22.

The issues of radiation and appropriate designs are discussed in the next sections. The front-end chip technology choice is presented in Sect. 6.4.1. The sensor design and strategies to achieve the necessary radiation tolerances are expressed in Sect. 6.4.2.

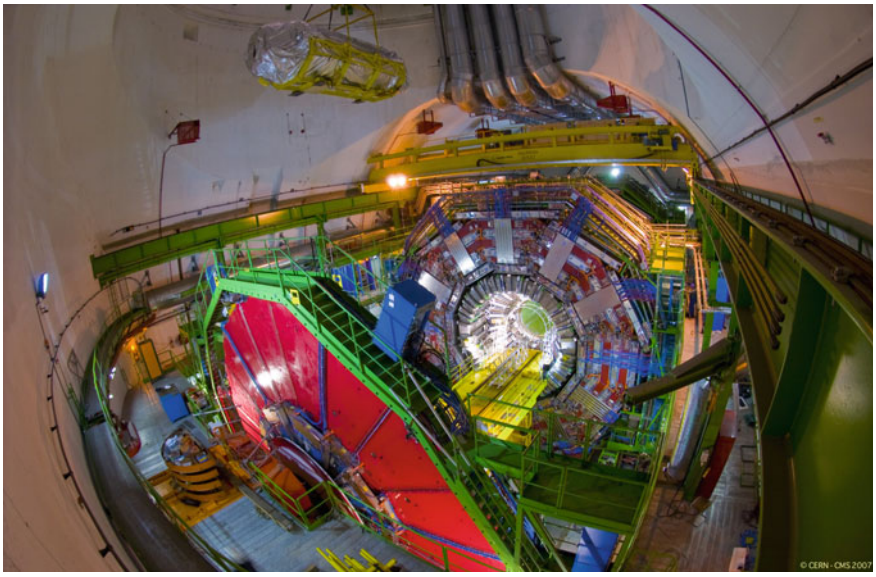


Fig. 6.20 The CMS tracker on its way to the heart of CMS – 2000 man years on a hook. In December 2007 the CMS tracker finally descended down the 100 m deep shaft to be inserted into the heart of the CMS experiment to measure the particle tracks at between 15 and 22 positions in the detector with a precision of better than 20 μm [Courtesy of CERN]

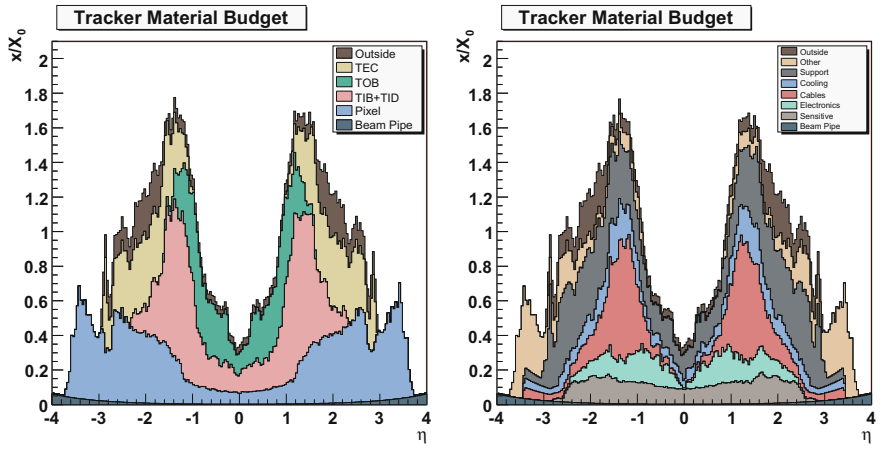


Fig. 6.21 Material budget is displayed in units of radiation length as a function of pseudo-rapidity η for the different sub detectors (*left*) and broken down into the functional contributions (*right*). [292]. A lot of design effort was spent to achieve a "light" tracker, Kapton cables were introduced directly on the flex hybrid with no extra connector; thin-walled cooling pipes were used; low mass power cables (aluminium) run inside the active CMS tracker volume, while outside low resistivity copper cables are used; structural elements are made of light weight carbon fibre and honeycomb structures. The beam pipe is made of beryllium. But despite all these efforts, at the interface of TIB and TOB to TEC all inner cables and pipes have to pass resulting in the two "too" high material peaks at both detector ends at $|\eta| = 1.4$

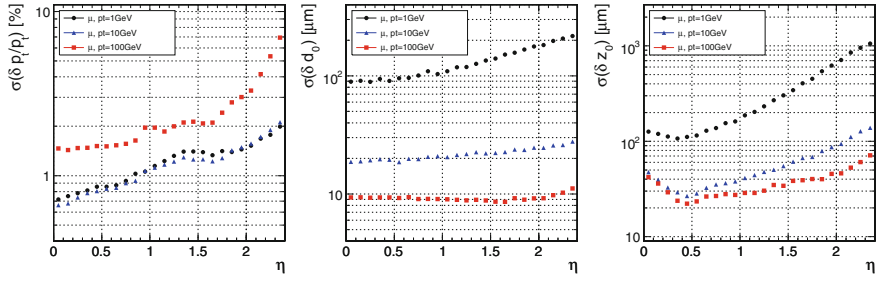


Fig. 6.22 Resolution of several track parameters for single muons with transverse momenta of 1, 10 and 100 GeV: transverse momentum (*left*), transverse impact parameter (*middle*) and longitudinal impact parameter (*right*) – simulation [292]

Different from CDF II the CMS tracker is not participating in the hardware-level trigger but tracking information plays a major role in the **H**igh-**L**evel software **T**rigger HLT (computer farm). The Level-1 hardware trigger reduces the 40 MHz event rate to 100 kHz. The HLT further reduces the data rate to 100 Hz to be stored[8] – refer to Sect. 6.7 later.

[8] 100 Hz HLT rate is the original design, today 1 kHz rate is standard while several kHz can be stored for dedicated calibration studies, e.g. timing scans.

6.4 Design, How to Survive 10 Years in the Radiation Environment of LHC

The radiation environment at luminosities of $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ can be translated to integrated fluences[9] on the order of $\Phi_{eq} = 10^{13}\ n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ in the outer strip tracker, $\Phi_{eq} = 1 \cdot 10^{14}\ n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ in the inner strip tracker volume for the full operation period of 10 years. The pixel detectors expect to be subjected to even $\Phi_{eq} = 3 \cdot 10^{14}\ n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ per year. These numbers are derived from simulations. Detailed numbers per volume cell including the real charge particle/neutron ratio are calculated. This radiation affects the gates in the ASIC transistors as well as the crystal lattice of the sensors. Early ASIC chips of the LEP or CDF chips would not have survived in the LHC environment. The basics of bulk and surface radiation-induced damages are discussed in Sect. 2. In the next sections the dedicated survival strategies of the CMS tracker are discussed.

6.4.1 Electronics – Quarter Micron Technology

In the late 1990s radiation studies on chips proved that the thinner the gate oxides the more radiation tolerant the transistor, thus the electronic device. This was already seen for the different chip versions of LEP and CDF where the decrease in feature size exhibited an increased radiation tolerance with respect to ionising radiation levels. Within a "thin" oxide there is simply not enough lattice space for an effective defect concentration of holes to form and accumulate. In addition, the probability of electrons "tunnelling" into the oxide from the silicon bulk beneath is higher for a thinner oxide layer. These electrons can then recombine with the trapped holes near the interface. The detailed effect and technology exploitation is described in [155, 156]. The technology is also called "deep sub-micron" or "quarter micron[10]" and is much more radiation tolerant[11] than any former chip generation. Unfortunately the use of standard ASIC design libraries from the manufacturer is not sufficient to achieve the necessary radiation tolerance. The feature size defines the gate oxide thickness but there are always neighbouring oxides where surface damage can form and the resulting increasing surface leakage current again renders the transistors useless. The missing link is the full enclosure of all n^+ diffusions at different potentials including N-wells – fully enclosed NMOS transistors. A schematic is shown in Fig. 6.23. For CMS any larger feature size than 250 nm would not have been radiation

[9] Reminder: fluence numbers are always given in 1 MeV neutron equivalent/cm^2 $[\Phi_{eq}] = n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$.

[10] At a feature size of 0.25 μm and below, chips start to be enough radiation tolerant for LHC operation.

[11] Another radiation-tolerant technology called DMILL existed but there were some production yield problems.

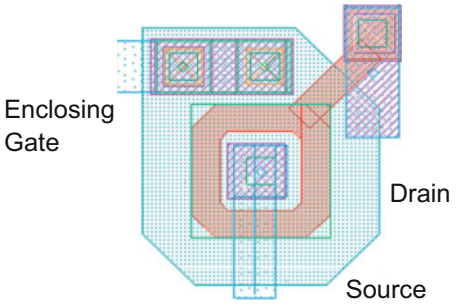


Fig. 6.23 Radiation-tolerant transistor, 0.25 μm, enclosed layout. The gate encloses all n^+ regions completely avoiding any thick transistor-relevant oxide structures. This method prevents any radiation surface damage in parasitic thicker side oxides. The thin formed gate oxide is intrinsically radiation tolerant

tolerant enough. Hybrids equipped with CMS readout chip APV25[12] were irradiated to fluences of more than 30 years of LHC equivalent and exhibit no degradation.

Gigi Rolandi, the CMS Tracker Project Manager (2000 – 2006) said,

> The Quarter Micron technology arrived just in time, without it the realization of our plans would not have been possible!

Both, the pixel readout chip (ROC) and the strip detector chip (APV25) are produced in quarter micron technology.

In a high radiation environment there is also a chance of a state change in the chip microprocessor, the memory cells or the transistors caused by an ionising particle. These single event upsets have only minor effects in the front-end electronics used LHC experiments and are beyond the scope of this book. They were investigated thoroughly and are considered "soft", while their effects are of temporary nature. They do not permanently degrade the functionality of chips.

6.4.2 Silicon Sensors

The pixel detector, being subjected to fluences up to $\Phi_{eq} = 3 \cdot 10^{14}\, n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ per year[13] at nominal LHC luminosity, utilizes different sensor designs and technologies than the strip detector, facing "only" a fluence of $\Phi_{eq} = 1 \cdot 10^{14}\, n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ for the full 10 years of operation. In reality, the LHC is exceeding design luminosity; still the margins of the Tracker system and its sensors, especially with the Pixel Phase I upgrade (2017), are comfortable.

[12] 0.25 μm version of the APV chip – final version.

[13] With an initially lower luminosity an integral fluence of $\Phi_{eq} = 6 \cdot 10^{14}\, n_{1\,\mathrm{MeV}}/\mathrm{cm}^2$ was expected after 4 – 5 years of initial running.

Pixel Sensors

The pixel detector's sensor and chip design is described in detail in [87, 292]. The sensor technology did basically not change from Phase 0 to Phase I, only the sensor geometry did for the forward part. The CMS pixel sensors are designed and processed in n-in-n technology on high resistivity oxygenated Float-Zone silicon (Diffused Oxygenated FZ – DOFZ), the most radiation hard concept available at the time (pre-LHC), and as proposed at the time by RD50. With the n-in-n concept (n^+ pixelated implants on n-bulk) the depletion volume after type inversion develops from the segmented pixel face thus allowing operation at bias voltages lower than full depletion voltage (under-depleted operation[14]) and still sufficient charge (electron) collection. The high oxygen concentration also reduces acceptor creation during irradiation thereby reducing final depletion voltage.

With an initial resistivity of about 3.7 kΩ and a thickness of d = 285 μm full depletion voltages are about 50 – 60 V. Guard rings on the edge of the sensors, being kept on relative ground, guarantee a stable running up to bias voltages of 600 V without breakdown. A perfect homogeneity for the barrel part has been achieved by having all wafers out of one single ingot. The n-in-n concept requires a double-sided process where the multi-guard ring structure are situated on the backside of the sensor, cascading the high negative voltage down to zero towards the side of the sensor. Historically, the barrel and forward pixel consortia acquired sensors from two different vendors with different design rules. Quality and performance are equally excellent but pixel cell isolation is solved with p^+-spray technique and p^+-stops for barrel [168] and forward [44] sensors, respectively. Photographs of the different pixel cells are shown in Fig. 6.24. Cells are DC-coupled; n-implants are connected by "vias" to the metal bump bonding pads and then connected with flip-chip technology to the chip cells (see Fig. 6.6 and Fig. 1.73). Both types are using punch-through as biasing techniques allowing IV characteristics at testing stage.[15] In the forward pixels, a simple opening in the p^+-stop presents a low resistance path until full depletion is reached changing to a high resistance path at over-depletion (volume below pixel cells depleted) to assure pixel isolation. The barrel pixel sensors, using moderated p^+-spray, utilize a dedicated punch-through structure with a well-defined bias dot for every pixel (small distance to GND potential and maximal area).

For Phase 0, two different sensor geometries have been realized for the barrel part; the standard rectangle corresponding to 2×8 read-chips and a smaller number of half-modules only 1×8 chips worth and necessary to instrument the connecting part of the two half-shells. The forward part is more complicated, where 7 different sensor geometries 1×2 up to 2×5 chips have been implemented to achieve a wedge geometry. Sensors plus chips sandwiches are called plaquettes; several plaquettes (3 or 4) form a module called a panel; two panels form a blade. The inner panel

[14] In n-in-n under-depletion mode, the non-depleted zone stays on the backside only reducing the active volume but with depletion zone at the segmented pixel face.

[15] GND potential is finally connected via the individual chip cells but for sensor testing purposes a bias grid has been implemented.

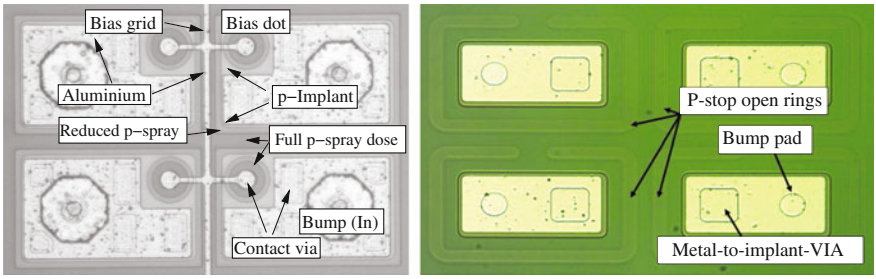


Fig. 6.24 Barrel and forward pixel sensor cells. The differences of pixel isolation and biasing methods are shown. Indium bumps on the barrel one are already deposited but not yet reflown [87, 292]

features 1×2, 2×3, 2×4, and 1×5 type plaquettes, the outer one 2×3, 2×4 and 2×5. The outer panel can be checked in Fig. 6.4. The forward system has a total of 672 plaquettes on 96 blades. All sensors have been processed on 4 inch wafers; clearly with several different geometries on one wafer.

For Phase I, all sensors are rectangular, also for the forward disks. Figure 6.25 shows a photo of the real wafers of the Phase I detector scaled down by a factor of two. The barrel pixel sensor cells are exactly the same as shown in Fig. 6.24. Interestingly, the 4 inch masks for the barrel sensors are physically *the same* as for the first production and have, several years later, simply been re-used at the company. The forward pixel cells underwent a small optimization from Phase 0 to Phase I.

With the initial PSI46 analogue chip, test beam runs [258] with un-irradiated and irradiated detectors have shown that the barrel sensors still have 99.0% efficiency for tracking charged particles after an exposure of $\Phi_{eq} = 3 \cdot 10^{14}\ n_{1\mathrm{MeV}}/\mathrm{cm}^2$ and the forward sensors maintained an efficiency of 98.8% after $\Phi_{eq} = 8 \cdot 10^{14}\ n_{1\mathrm{MeV}}/\mathrm{cm}^2$. With the Phase I, PSIDIG and PROC600 digital chips with much lower threshold capability, the system becomes much more radiation tolerant and will be able to operate for the full until Long Shutdown 3 foreseen integrated luminosity of $\mathcal{L} = 500\ fb^{-1}$ corresponding to $\Phi_{eq} = 3 \cdot 10^{15}\ n_{1\mathrm{MeV}}/\mathrm{cm}^2$ accumulated in the innermost pixel layer at $r = 3$ cm. Barrel Layer 2 gets four times less. The system, as extra precaution, allows to extract the pixel detector and exchange the innermost layer in a normal Year-End-Technical-Stop. The strategy and mechanical structure allows, though, to replace layer 1 easily after about half the expected fluence.

Strip Sensors

All silicon strip sensors are single-sided p-in-n with AC-coupled readout and p^+ strips biased through polysilicon resistors, based on the planar process. Pitches range from 80 to 183 μm (up to 205 μm in the end of a wedge-shape sensor) without any intermediate strips. The substrate is non-oxygenated Float-Zone n-type silicon made from 6 in. diameter wafers. All barrel sensors are consequently rectangular with strips running parallel to the beam axis, a stereo angle for layers 1, 2, 5 and 6 is achieved by rotating the complete sensor inside the module by 100 mrad. Double-sided modules are composed of a normal and a stereo module back-to-back.

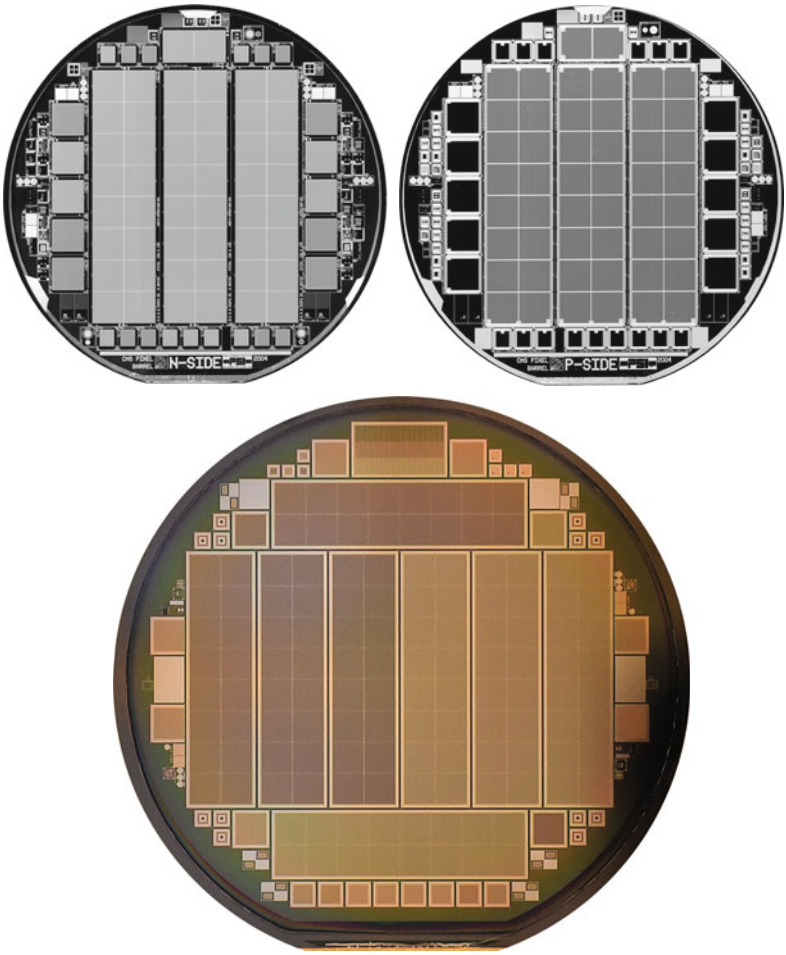


Fig. 6.25 The photos show the wafers of the pixel sensors. All pictures are scaled down by a factor of two. The *first* shows the n-side (pixel cell) side of the barrel sensors. n-in-n sensors need double-sided processing featuring guard rings on the backside and full backplane metallisation – see second (*right*) picture. The *last* photo shows the n-side of a forward pixel sensor wafer processed on 6 in. The resulting individual sensors have the same size but eight instead of three sensors fit. Photos of the individual pixel cells can be seen in Fig. 6.24 [Courtesy of Paul Scherrer Institute and Fermilab]

All sensors in the forward part are trapezoidal with strips running in the radial detector direction. As discussed earlier outer modules contain two daisy-chained sensors. In order to maintain a good signal-to-noise ratio of well above 10 (after irradiation) CMS uses thicker silicon sensors for the outer tracker region (500 μm thickness as opposed to the 320 μm in the inner tracker) with correspondingly higher signal. These thicker sensors would in principle have a higher depletion voltage, since $V_{FD} \propto D^2$. But since the radiation levels in the outer tracker are smaller, a higher

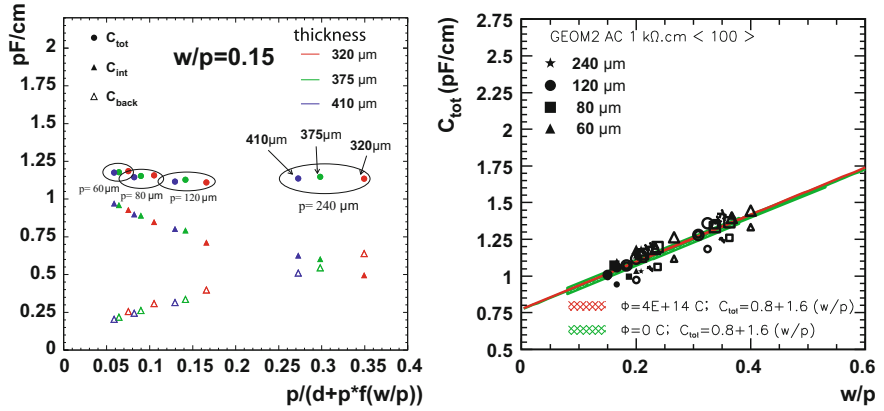


Fig. 6.26 Varying width and pitch of the microstrip sensor, CMS finds a constant total capacitance for a constant width/pitch (w/p) ratio. The *left* plot shows the constancy of C_{tot} together with its two components C_{int} and C_{back} for a constant w/p = 0.15, capacitances are plotted versus formula (6.2). The *right* plot shows the linear increase of C_{tot} with increasing w/p. The test holds true for several sensor thicknesses and for ⟨111⟩ as well as ⟨100⟩, but only ⟨100⟩ shows also robustness versus radiation, refer also to Fig. 6.30 and [34, 47, 49, 77]

initial resistivity can be chosen such that the initial depletion voltages of thick and thin sensors are in the same range of 100–300 V, since $V_{FD} \propto N_{eff}$. The different sensor parameters, electrical as well as geometrical, are presented in Table 6.2. The load capacitance is the main source of noise for fast signal shaping, therefore detailed studies were conducted to establish a correlation between pitch p, strip width w, sensor thickness D and load capacitance composed of inter-strip and backplane capacitance. Multi-geometry sensors were processed in the R&D framework of CMS together with the vendor where several combinations of w/p were tested on ⟨111⟩ silicon crystal orientation as well as ⟨100⟩ silicon, as well as for different resistivities. The result proved that for a constant w/p the total capacitance C_{tot} stays constant, inter-strip capacitance decreases in the same order as backplane capacitance increases (see Fig. 6.26). Furthermore, to achieve robust coupling capacitors a multi-layer dielectric was used. A good process stability of polysilicon resistors was achieved by saturating by heavily doping the sheet resistance. Further information about the CMS sensor design can be found in [34, 47, 49, 77]. Formulas (6.1) – (6.3) parameterize the results

$$C_{int} = \left[0.03 + 1.62\frac{\mathrm{w} + 20\,\mu\mathrm{m}}{\mathrm{p}}\right]\frac{\mathrm{pF}}{\mathrm{cm}} \tag{6.1}$$

$$C_{back} = \varepsilon_0\varepsilon_{\mathrm{Si}}\frac{p}{D}\frac{1}{1 + \frac{p}{D}f\left(\frac{\mathrm{w}}{\mathrm{p}}\right)} \tag{6.2}$$

$$C_{tot} = \left(0.83 + 1.67\frac{\mathrm{w}}{\mathrm{p}}\right)\frac{\mathrm{pF}}{\mathrm{cm}} \tag{6.3}$$

A constant w/p = 0.25 is used for all CMS sensors resulting in a load capacitance of 1.2 pF/cm strip length.

Within the high radiation environment of LHC all sensors will undergo type inversion, as discussed in Sect. 2. The main strategies of CMS to ensure radiation hardness of silicon sensors consist of

1. delaying the bulk type inversion and achieve V_{FD} always below 400 – 500 V
2. the use of stable sensors with respect to high voltage and
3. the reduction of surface damage

Delaying Type Inversion – Guarantee $V_{FD} < 400 - 500\ V$

A tailoring of sensors and of the operation plus maintenance scenario is possible to withstand a substantial amount of radiation by taking into account all theory (see also Sect. 2) of radiation damage. Especially important is the evolution of full depletion voltage versus fluence, time and temperature. The three main strategies are the following:

- freeze out all reverse annealing by maintaining all sensors at sub-zero temperatures during operation and as much as possible during maintenance
- exploit beneficial annealing with a controlled warm up during the maintenance periods
- start with relatively low resistivity silicon to reach the inversion time late and end with about the same depletion voltage as the initial one

The employment of low resistivity silicon and therefore high initial V_{FD} delays the type inversion point in time resulting in an operational depletion voltage after 10 years of LHC operation. Figure 6.27 (left part) shows the simulated evolution of V_{FD} with time including all operation and maintenance periods, hence radiation and annealing periods. The different time constants for beneficial and reverse annealing have to be exploited. During operation at sub-zero temperature all defect annealing processes are practically suppressed. Figure 6.50 on p. 278 shows the individual sensor temperatures in situ compared to the original design value $T_{silicon} = -10\,^{\circ}\mathrm{C}$; there is margin to reduce temperatures by another 10 degrees. Operating at low temperature is also necessary to limit leakage current and avoid thermal runaway. During maintenance periods the tracker will be "warmed up" to benefit from the beneficial annealing component. Figure 6.27 (right part) shows the resulting depletion voltages depending on the "warm up" temperatures and it clearly shows that these periods are needed, but with meticulous temperature control. During the whole production period, sensors and test structures have been irradiated and results were compared to the Hamburg model – Fig. 6.28.

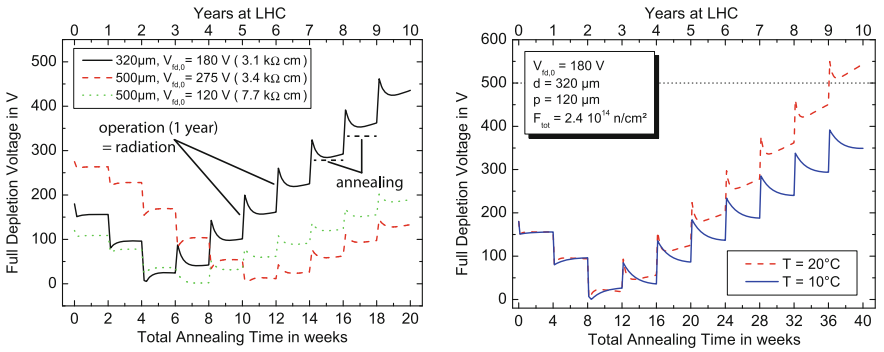


Fig. 6.27 The CMS sensors are subjected to a high radiation field. To freeze the radiation damage and to decrease leakage current the operation temperature is −15 °C. During shutdown and maintenance periods the tracker will be "warmed up" for required interventions or to make controlled use of the beneficial annealing effect (see Sect. 2.1) [79]

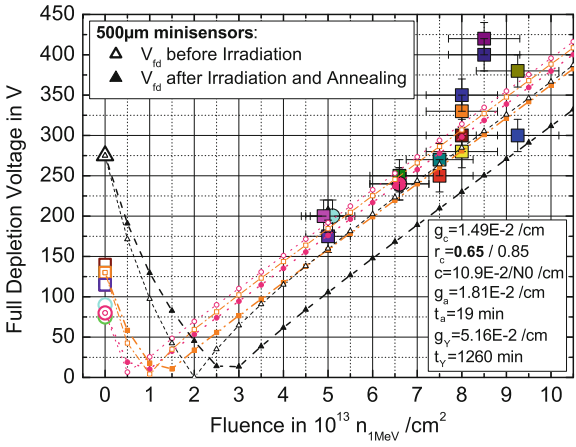


Fig. 6.28 Depletion voltages of several CMS sensors for several fluences. These measurements were used to determine the Hamburg model (see Sect. 2.1) parameters for the CMS sensors and to evaluate operation parameters during the 10 years running period [79]

High Voltage Robustness

As a second measure, the sensors have to be extremely robust with respect to operation voltage. For several regions, bias voltages of around 500 V will be necessary after 10 years of LHC operation. The basic strategy is to shift high fields into the isolation oxide and consequently avoid high local fields:

- all implants are deep and round shaped
- no sharp corner on any metal structure, only rounded shapes
- active n^{++} edge to have a defined homogenous field at the outer region (p^{++} for n-in-p)
- no chips/breaks larger than ±20 µm at the cut edges, not penetrating the active edge
- metal overhang to distribute field to the SiO_2–metal region for strips, guard and bias ring

- implementation of a floating guard ring with a metal overhang to continuously smooth the high potential (HV) from the edges to GND applied on the inner bias ring. This is an improved version of a multi-guard ring which levels the potential discretely
- after type inversion, high field region is located at the sensor backplane

The use of round structures and avoidance of breaks and chips together with an active edge[16] are more or less standard methods to avoid any high local field. The implementation of a metal overhang is a CMS innovation and improves voltage robustness significantly. High fields[17] at the p^+ implants were shifted towards the metal strips into the SiO_2 insulation where the breakdown voltage is 20 times higher ($V_{break}(Si) = 30\ \mathrm{V/\mu m};\ V_{break}(SiO_2) = 600\ \mathrm{V/\mu m}$). In addition, it was possible to reduce a multi-guard ring to a single one with an overhang operating with a continuous potential shift of the high voltage at the sensor edge to GND at the bias ring. Many layout simulations and final testing of the implementation were necessary. The metal overhang ranges between 4 and 8 μm. At best, the aluminium electrode should even have a slightly more negative potential than the p^+ implant. All sensors were tested during standard quality control up to 550 V, many up to 800 V and some sensors were tested even up to 1500 V without breakdown.

Surface Damage Reduction

As a third measure CMS uses silicon with ⟨100⟩ crystal orientation, with less dangling bonds than standard ⟨111⟩ silicon (Fig. 6.29). Formula (6.4) lists the surface density of dangling bonds.

$$\text{Surface density} \begin{matrix} 11.8 \\ 9.6 \times 10^{11}\ \frac{\text{atoms}}{\text{cm}^2} \\ 6.8 \end{matrix} \begin{matrix} \langle 111 \rangle \\ \langle 110 \rangle \\ \langle 100 \rangle \end{matrix} \tag{6.4}$$

The use of ⟨100⟩ with the lowest number of dangling bonds naturally leads to a suppression of surface damage resulting in reduced increase of inter-strip capacitance after irradiation. Figure 6.30 illustrates the capacitance differences of radiated ⟨111⟩ and ⟨100⟩ sensors. In addition the use of ⟨100⟩ also decreases the undesired oxide charges therefore decreasing micro-discharge probability (Sect. 1.6.3).

The interested reader will find more exhaustive information about the CMS sensors and their quality control program in [13, 129, 183]. The process control, a vital part of the quality assurance, is described in detail in [35, 197].

[16] Was already implemented in CDF II, refer to Sects. 1.6.2 and 5.3.

[17] Main field from the high bias voltage applied to reverse bias the *pn*-junction.

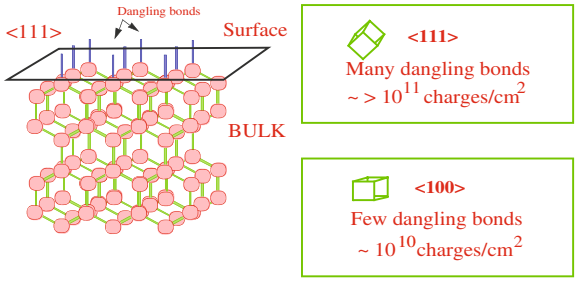


Fig. 6.29 CMS uses ⟨100⟩ instead of ⟨111⟩ crystal orientation, exploiting the smaller number of dangling bonds at the Si – SiO_2 interface allowing less charge accumulation due to irradiation

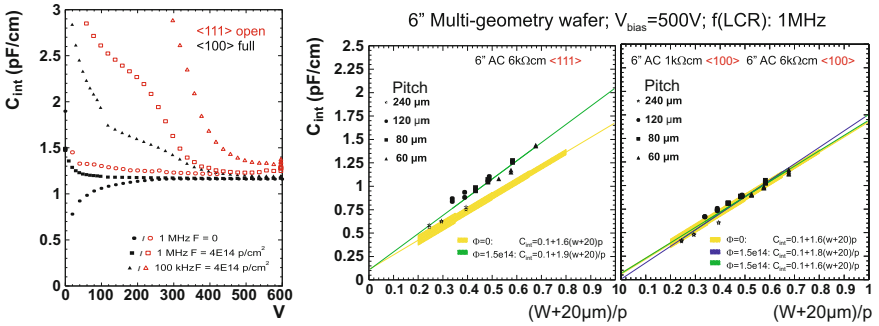


Fig. 6.30 In the *left part*, the full bullets show the beneficial effect of ⟨100⟩ crystal orientation. Differences at over-depletion are better seen in the *right part*. The study is consistent for several different pitches p and strip widths w. A clear beneficial effect of the ⟨100⟩ orientation is shown [34, 47, 49, 77]

The following items summarize the CMS tracker strategies to guarantee a safe 10 year operation:

- relatively "low" silicon resistivity is used resulting in high initial depletion voltage shifting type inversion to a later time ending up with voltages that are not too high
- operation at low temperatures limits leakage currents, avoids thermal runaway and freezes reverse annealing
  - to freeze the reverse annealing, the detector must remain cold, also in the maintenance periods
- the beneficial annealing effect is exploited by a controlled temperature rise during shutdown periods
- voltage robustness of the sensors is improved to allow for high voltage running especially with the use of a metal overhang design
- ⟨100⟩ silicon was chosen to minimize Si – SiO_2 interface deterioration

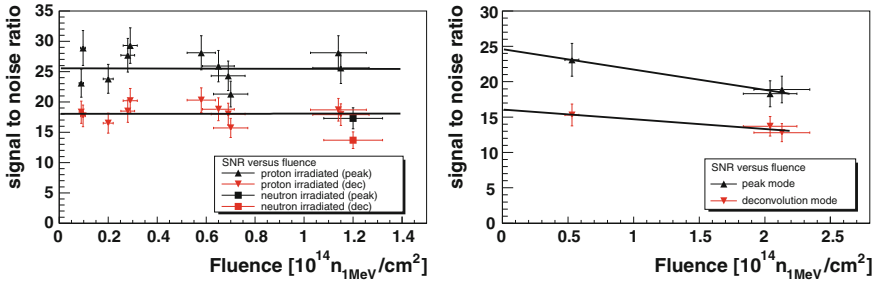


Fig. 6.31 Signal/noise of full CMS modules after neutron and proton irradiation (*right plot*: protons only). The values far exceed the expectation and technical design specification [351]. A substantial number of full modules were irradiated to several fluences to guarantee full functionality for the 10 years of operation with a sufficient signal/noise value

Figure 6.31 depicts the estimated functionality over the complete operation period of the CMS tracker with higher than marginal signal/noise ratio. All sensors are operational at 400 V and above, a clear sign of a successful design. A number of modules were thoroughly tested for beam incident robustness which they passed successfully, refer to [98, 99, 130].

The Thermal Screen

To guarantee the sub-zero temperature described above during the whole operation and maintenance period of 10 years, an active thermal screen was designed and built. This is an innovative and unprecedented feature of a tracking detector. The thermal screen is an "intelligent" insulation, responsible to keep the inside below $-10\,^{\circ}\mathrm{C}$ and the outside at $18 \pm 1\,^{\circ}\mathrm{C}$. This is achieved with 32 cold panels inside a Rohacell insulation layer and several dedicated heating foils outside. Two dedicated and redundant cooling stations guarantee the cold operation, while the heating foils are controlled by a dedicated system via PID[18] feedback loops. This engineering masterpiece is one of the keys for the silicon sensor to survive in the harsh environment. The thermal screen is located inside the tracker support tube surrounding the TOB and TEC sub detectors.

6.5 Construction Issues for Large Detector Systems with Industry Involvement

Describing all the logistics, quality assurance steps, technology choices, failure types, etc. of the CMS tracker could fill a book by itself and is the subject of a larger number of publications and conference presentations, e.g. [12, 72, 128, 130, 269]. The challenge to equip 206 m^2 with silicon detectors up to radii of $R = 110$ cm is unprecedented. The achievement was made possible by adopting and developing new production and quality assurance methods and by adapting industrial techniques.

[18] **P**roportional plus **I**ntegral plus **D**erivative PID control, an industry control standard.

Table 6.3 Number of individual components built into the CMS strip tracker

Thick sensors	18,192
Thin sensors	6052
Front-end hybrids/full modules	15,148
FE ASICs	75,376
Electronic channels	93,16,352
Petals	288
Rods	688
Analogue optical links for 65 m transm.	>40,000
Front-**E**nd **D**rivers (ADC boards)	440
Peak module assembly per day	90
Optical fibres	3000 km
Voltage channels (LV+HV)	8000
Temperature and humidity sensors	1000
Slow control computer	10

To stress the fact even a bit more, a petal consists of about 400 sub parts each, a module has about 10 items with about 1000 – 3000 wire-bond connections. A hybrid has 4 – 6 chips in combination with three additional ASICs and about 40 passive components

These procedures were novel in the field of particle physics detectors. A large collaboration of 51 institutes[19] with almost 500 physicists and engineers succeeded over a period of 12 – 15 years to design, develop and build this unique device.

Table 6.3 presents an example of the numbers of components and Fig. 6.32 visualizes the different locations, distributed all over the world, of the different assembly levels. In-between every shipment there are defined acceptance protocols, e.g. visual inspection and/or electrical testing. Also in-between each assembly step a quality assurance test is mandatory. A consequence of the increase in detector size is a change in the personal attitude towards the individual sensors and modules. With the number of sensors exceeding several thousand elements these have to be tested, handled and used in a nearly industrial manner. The time is over when each individual silicon sensor received special attention and was "brought to life" in many hours of careful handling. The silicon detector community had to develop standard procedures to systematically test sensors, efficiently perform the module construction and to use them in large detector systems.

6.5.1 Quality Assurance and Problems During the Process

A detector of this size and complexity calls for unprecedented meticulous quality control to identify and solve problems as early as possible in the production. Some

[19] Today 2017, there are 69 Tracker Institutes.

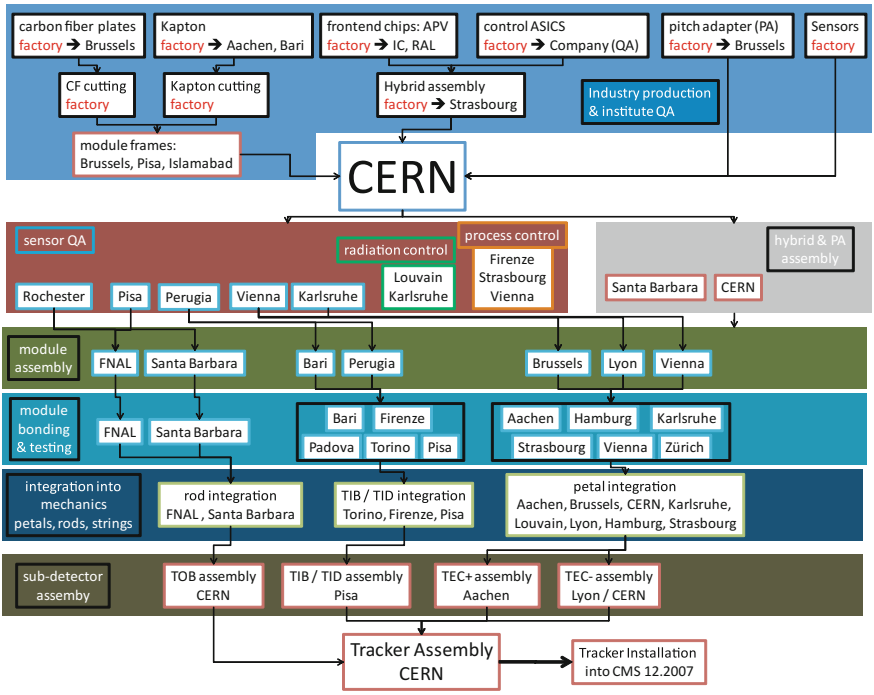


Fig. 6.32 The CMS tracker logistics

examples are listed. Basic **Q**uality **A**ssurance QA strategies for silicon sensors are described in Sect. 1.7.1 including the steps introduced for CMS. In addition to the standard sensor quality control, CMS introduced a special "process monitoring" necessary for the long procurement time of 3 years for the sensors. Due to the stringent radiation requirements a fraction of sensors and test structures were irradiated to 1.5 times the nominal fluence levels. The result is an overall percentage of good strips in the accepted sensors of 99.94%. The early qualification steps are described in [13, 129]. Along the way all kinds of faults were encountered and all were subsequently solved, mostly in close collaboration with the vendor. In an early stage of delivery coupling capacitors consisted only of SiO_2. This insulation layer technique proved to be fragile and several pinholes developed during initial quality control. The issue was solved with the introduction of additional layers of nitride [183]. The stringent requirement of cut edge breaks/chips below 20 μm required several visits to one company to sensitize them to the problem. The thin sensor handling plus requested measurements "after" cutting was not a common standard for the manufacturer. Also scratches occurring on certain areas of the sensor was a frequent fault. On a wafer with 100 ASICs a scratch decreases the yield by a percentage but with a single sensor per wafer the yield drops to zero. Testing personnel were finally specially trained for CMS requirements, see also. Section 1.14 and [60, 130, 183]. In Sect. 1.14 a short

description of the low inter-strip resistance problem is presented. The problem was identified due to vigilance and the stringent process monitoring established, because inter-strip resistance was not a standard QA item of the manufacturer.

For a period of about 2 months a high flat-band voltage was experienced due to oxide contamination, oxide traps and/or interface states for a large quantity of ∼1000 sensors. An immediate irradiation campaign showed a resulting increase of inter-strip capacitance with radiation different to the non-contaminated ⟨100⟩ sensors [197].

Very prominent was the finding of corrosion of the aluminium bias and guard rings, described in Sect. 1.14 and elaborated in detail in [139].

Even after sensor QA individual channels exhibited an influence on common mode noise due to micro-discharges, observed in the module long-term testing. A single strip, inconspicuous during sensor QA, was able to significantly increase the noise of a full chip [60]. An additional voltage[20] had to be applied to the AC pad, while current was measured on the DC pad to mimic the operation condition in a module. In this configuration a strip vulnerable to micro-discharge exhibited a significantly higher current.

Glass pitch adapter series were delivered with varying metal line quality and bonding was not always possible. The ability of sensors and pitch adapters to be bonded was constantly monitored during the whole production. The final bonding failure rate stayed below 0.01%.

Another serious problem affected a series of hybrids. Vias were unstable under thermal stress. A cleaning step during processing was also etching away part of the glue between the Kapton layers. The metal conductor path was therefore stretching too far into the grooves and was breaking under thermal expansion. As a solution, an additional Kapton layer was introduced and the via size was increased, see also [85, 128, 181].

A new problem relevant only for sensors in a very high-radiation environment was detected during module production. The backside module contact was made with conductive glue between aluminium sensor backplane and metallized Kapton. Varying and ageing conductivity of the conductive glue was found between several Ω and several kΩ. Oxygen was diffused through the glue oxidating the aluminium surface and thus degrading conductivity. For a sensor with bulk resistance in the order of several GΩ, an additional series resistance of some kΩ is of no consequence at all. Since the bulk resistance drops significantly with radiation down to several kΩ, the voltage drop on the glue would be in the same order as for the sensor bulk and depletion could no longer be guaranteed [85, 128, 181]. All modules were therefore retrofitted with backplane bonding.

The prior mentioned active n^{++} edge introduces the feature that the edge is at a high potential and therefore wire-bonding with a low loop height resulted in a voltage spark. Module quality control detected this problem and bonding loop parameters had to be adapted accordingly.

[20] During operation together with the hybrid, the potential on the chip connections are +0.85 V higher than the sensor ground.

All components, hybrids, modules, rods, petals and sub detectors, were long-term tested and all went through several thermal cycles. The level of quality assurance was new for the HEP community but ultimately with zero fault tolerance and rigid vigilance, the final strip inefficiency remains below 0.5%. This includes all sensor, chip, hybrid and module faults and all missing connections from readout and power towards the tracker.

6.5.2 Assembly

To realize a tracker composed of 1440 pixel and 15148 strip detector modules, industrial quality control and assembly methods and logistic structures are necessary as well as a modular design. The smallest assembled element is a module, an example is shown in Fig. 6.33. Modules were assembled in six institutes worldwide using automated robotic "gantry" systems. A gantry system consists of a camera surveying the various components, a robotic arm which can hold vacuum pickup tools for placing sensors and hybrids and syringes to precisely dispense silicone epoxies. Pitch adapter to hybrid assembly was also fully automated on a gantry. A CMS gantry is shown in Fig. 6.34. With pattern recognition a precision placement of 10 μm and 0.1 mrad was achieved. A removable assembly plate holds components for three or four modules. After the assembly, the plates were moved, with module components fixed by vacuum, to complete the curing of the glue, while the gantry was available for further assembly work. With this procedure a worldwide peak assembly rate of 90 modules/day was achieved. Modules were then sent to bonding centres with high-throughput industrial bonding machines. Bond pictures are displayed in Fig. 6.35; 25 million custom wire-bonds were necessary to finish the module construction. As the next step, modules were assembled into TIB strings, TID disks (Fig. 6.16), TOB rods (Fig. 6.17) and TEC petals (Fig. 6.36) – in the shell, rod, and petal integration centres.

Overall, 7 sensor quality assurance centres, 3 hybrid production sites, 6 module gantry centres, 2 hybrid assembly centres, 13 bonding and module testing centres, 2 rod integration and 7 petal integration centres and 7 shell integration centres struggled through construction and logistics, however they were very successful. CMS has gained experience in logistics of a large production which is widely distributed over several institutes and countries, refer to Fig. 6.32. The petal integration center of Karlsruhe is presented in Fig. 6.37.

The remaining steps were (1) insertion of sub structures into the sub detectors (2) inserting sub detectors into tracker support tube (3) installation of the tracker into the CMS detector plus (4) the connection of all power cables, optical fibres and cooling pipes.

Rods were integrated into the TOB structure in a drawer-like fashion, see Fig. 6.17. The TOB was then inserted into the tracker support tube, resting on a dedicated rail system.

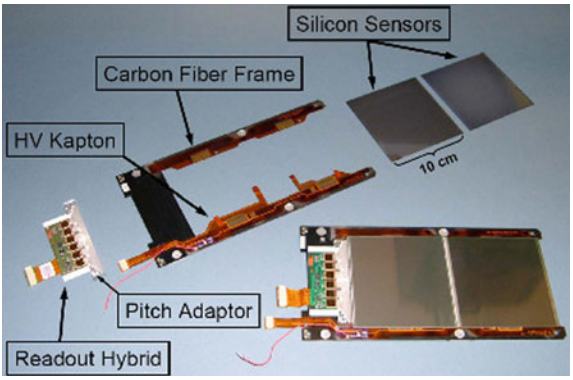


Fig. 6.33 A CMS module. The different parts forming a module are the frame of carbon fibre and Kapton, the hybrid with the front-end electronics and pitch adapter and the silicon sensors. Courtesy of colleagues from UCSB Santa Barbara, California

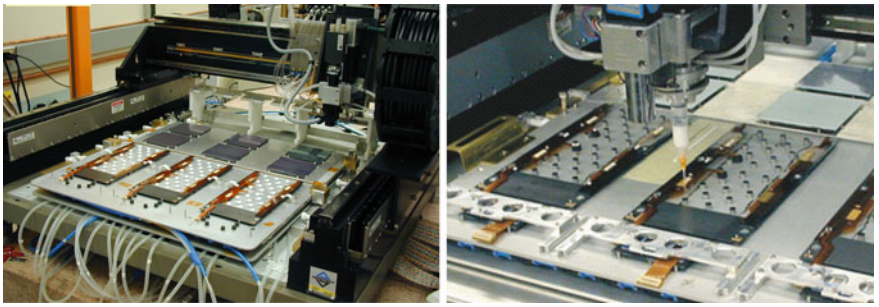


Fig. 6.34 The gantry is a robotic assembly for modules. The gantry systems consist of a camera surveying the various components, a robotic arm which can hold vacuum pickup tools for placing sensors and hybrids and syringes to precisely dispense silicone epoxies [Courtesy of CMS]

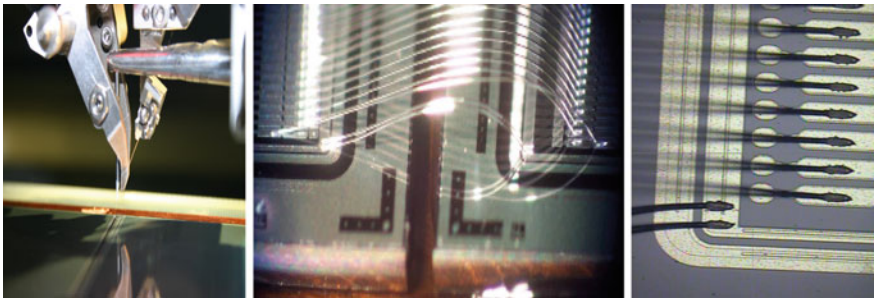


Fig. 6.35 CMS module bonding

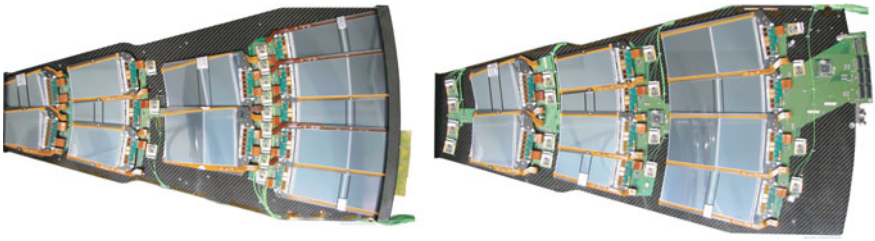


Fig. 6.36 The inner disk petals are the largest of the eight petal types, equipped with 28 modules and more than 1 m long. The pictures [292] show both sides of a petal. All modules are assembled with an overlap and space seen taken by electronics is covered on the other side with modules. A 100% coverage plus overlap is established. All strips are following a full radial path, but some modules are rotated by 100 mrad to have two-dimensional point information. Eight different petal types are needed to cover all cases on the endcap volume. One TEC is composed of 144 petals

TIB and TID modules were mounted directly and manually onto the different TIB layers and TID disks. TIB and TID were assembled together with cables, pipes and fibres, integrated then tested and inserted into the TOB, resting on another dedicated rail system, see Fig. 6.18. The picture also shows a fraction of the amount of necessary services to be connected.

Petals were mounted on the front and backside of the nine TEC disks. Insertion happened sectorwise on one disk face only, then the whole TEC was rotated to get easy access to the other disk face. For the TEC all services like fibres, cooling and power cables were integrated even before the first petal was inserted. A sector cabling and petal placement from the front can be seen in Fig. 6.19. The TECs were then inserted into the tracker support tube on the same rail system as the TOB.

Late December 2007, 8 years after the decision to construct an all-silicon tracker, it was finally inserted into CMS, inside the barrel electromagnetic calorimeter. Figure 6.38 catches the historic moment.

The CMS spokesperson (2008 – 2009), Tejinder Virdee about the CMS tracker:

> Constructing a scientific instrument of this size and complexity, designed to operate at the LHC without intervention for more than ten years, is a major engineering and scientific achievement. More than five hundred scientists and engineers from fifty-one research institutions world-wide have contributed to the success of the project.

The meticulous design and quality assurance paid off. The CMS tracker was installed inside CMS; it was connected and is ready to operate for 10 years. August 2008 (publishing date of the first edition), about 99.5% of the tracker was working properly.

Fig. 6.37 Petal integration in Karlsruhe. Three technicians plus two physicists were assembling two petals simultaneously. The petals were mounted on so-called petal grills which enable easy turnaround and thus access to both sides. After assembly, petals were long-term tested in fridges with additional active petal cooling

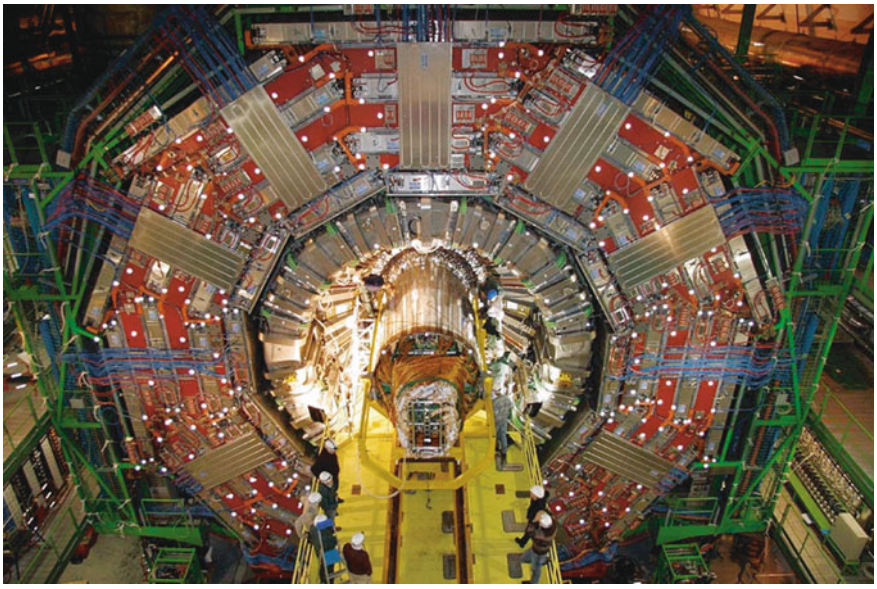


Fig. 6.38 Tracker insertion. After assembly and testing, the CMS tracker arrives in the heart of CMS. From this moment on another 3 months were needed to connect all fibres, cooling pipes and power cables before commissioning and data taking was possible. Around 130 km of cables and 18 km of cooling pipes were installed and tested [Courtesy of CERN]

6.6 Tracker Operation and Performance

During RUN I and RUN II of the LHC, the CMS Tracker and CMS in general showed an excellent performance. Several papers have been published and a collection can be found here [52, 75, 103, 228, 295, 306, 344, 359]. The uptime was always excellent $\gg$99% in 2012. The SST had an uptime of 100% in 197 out of 302 so-called physics fills of the LHC machine! The main reasons for Tracker specific downtimes were failures of the cooling system, power racks or primary power supply failures, problems with individual **F**ront **E**nd **D**river boards FEDs (and here specifically the service infrastructure; failing crate power supplies). In the early days of operation and during the increase of instantaneous luminosity firmware upgrades of the FEDs were necessary to cope with rare occurrences of problems inside the detector and data hiccups.

6.6.1 Lessons Learned from Operation and Maintenance

The key of successfully operating a complex and large detector such as the CMS Tracker is careful planning and early automation of regular tasks. Detector Control, Safety and Data Acquisition systems have been built defined by databases, meticulously filled with info about cabling, connection and sensor map, parameters, limits, etc. . The software then scanned all hardware items to determine the baseline parameters such as pedestals, noise, gain of laser, dead or flaky channels, etc. For the pixel detector several iterations of chip parameter tuning are necessary regularly to achieve the optimal thresholds for every individual pixel. Despite full automation, this procedure took several weeks and is being repeated after every LHC **Y**ear-**E**nd-**T**echnical-**S**top YETS and some partial aspects weekly. In general, all systems are behaving very well but some surprises and difficulties were encountered along the way.

Safety is essential for a detector, basically inaccessible and therefore not really repairable. All safety systems ran from day 0 and never failed. All relevant values like temperature, humidity, current, voltages are constantly monitored and many participate in a majority logic to interlock the detector in case of values exceeding predefined threshold values. In reality, warnings including SMSs are issued long before and a software system analyses trends and shuts down smaller parts of the system before any hardware safety interlock kicks in. With this compromise, safety is guaranteed but in case of over-temperature in a small section, the major fraction of the detector remains available for physics data taking.

In my personal experience

> It is always the low-tech that fails. It's the plumbing and cable connections. We are very good in getting all complex parts, e.g. ASICs, sensors, control and readout boards going but fail in the simple ones.

Four[21] dedicated cooling stations operate with C_6F_{14} in the experimental[22] cavern serve 180 cooling loops of the SST and 2 times 8 for the Thermal Screen. These are served by the so-called "brine circuit" via heat exchangers. In 2008 we experienced a rupture of one of those heat exchangers ending up having some brine in one of the cooling plants (fortunately not in the detector). As follow up, brine has been replaced with C_6F_{14} adding much to robustness but also reducing the cooling performance[23] making a large refurbishment of the plants in the first long shutdown necessary.

Later, an overpressure[24] incident during a maintenance operation weakened some joints and some of the over-pressured cooling lines started to leak a substantial amount of C_6F_{14}; subsequently, today, 5/180 lines are closed and the corresponding detector parts and are being only cooled by neighbouring lines via the carbon fibre structure. Luckily the margin in the system allows to operate them during the full design integrated luminosity despite changes due to radiation. And even without them the tracking efficiency would only be marginally affected – still something to avoid for future detectors. The built-in redundancy paid off. In the endcap, two cooling circuits serve alternatively the odd and even layers of the disks in one sector and simulation prove that one missing circuit would not jeopardize tracking/physics performance significantly. In some cases the overpressure moved the pipes slightly and in the very dense environment, they now touch the sensor backplanes short-circuiting the sensor HV voltage potential. Also the cooling contacts of some modules degraded resulting now in a spectrum of temperatures across the whole detector – temperature maps will be presented in Fig. 6.50 in the section about sensor evolution with radiation.

To reduce sensor leakage current and prevent reverse annealing the system was designed to keep sensor temperatures during operation at $T = -10\,°C$ and consequently the dew points below $T = -30\,°C$ to avoid any risk of condensation. These dew point specifications were met from day 0 inside the detector but not in the interface volume towards the detector where some bare pipework[25] was in contact with higher humidity values. In short, for the first years the CMS Tracker was operated about 25 °C higher than specified. Final operation conditions were only met in Run II, which started 2015 after two years of refurbishment [52]. Today, the environment conditions (dryness and cooling power) allow operation at $-25\,°C$ (coolant temperature). With these conditions the Strip Tracker is operational until long shutdown 3 LS3 (2024 – 2026) with respect to degradation due to radiation damage. The ingredients of success were (a) largely refurbished cooling plants, (b) an engineered vapour-sealing concept, (c) heater elements outside the cold volumes (bulkhead and

[21] Until 2017, an additional C_6F_{14} one was serving 18 lines for the pixel detector but has been replaced with a CO_2 system for the Pixel Phase I detector. The CO_2 is situated in the service cavern and has only dedicated passive manifolds on the experimental cavern.

[22] In CMS magnetic fringe field and radiation environment; with about 60 m long transfer lines.

[23] C_6F_{14} is much denser/heavier than water.

[24] Pipes had been closed on both sides with the liquid not yet thermalised to room temperature thus pressure increased during warm up. Later the possibility to close the return end was abandoned.

[25] With 180 pipes in a very small and thus crowded environment, thermal insulation was not possible.

cooling ducts), (d) highly granular[26] and precise temperature and dew point sensors also remote at the end of a gas extraction system[27], and (e) most importantly a very high flux of dry gas (installation of a dedicated membrane plant[28]. With the initial relatively low radiation levels during RUN I, no significant additional degradation has been encountered due to the elevated temperature – values have been closely monitored and some results will be presented in Sect. 6.6.2. In the first month of operating cold another unexpected phenomenon was encountered. The valves of the cooling lines, located in the experimental cavern thus inside the magnetic fringe field, are piloted by pressure of dry air. Despite a gas dryness with dew points below $-60\,^{\circ}\mathrm{C}$, after several weeks ice clogging was encountered inside pipes touching structural elements of temperature $\mathrm{T} = -15\,^{\circ}\mathrm{C}$. Water from the scarcely humid cavern air diffused through the plastic pipes! The then formed ice blocked the gas flow/pressure thereby closing the valve thus switching off cooling for parts of the detector. The solution is to avoid stagnant air by adding some micro openings in the pipes flushing out any water content.

Erik Butz, CMS SST Operations Manager and Technical Field Manager

> Seeing this detector in operation, one cannot help but marvel at how well this large and complex system works. On the other hand also after a number of years, it still manages to surprise us with new and unexpected problems

As for operation, in Run I, 97.7% of the barrel pixel and 92.2% of the forward pixel were operational. During LS1 the pixel detector had been extracted and repaired (see Figs. 6.8 and 6.9). The forward detector was fully recovered 100% minus one single chip. In the barrel part, all reasonably accessible modules have been treated arriving at 98.9% functional units. The detector has been repaired in LS1[75]. A small number of defective detector modules have been replaced. But the main problems were missing or flaky connections or some connections with inferior continuity thus showing signal degradations and slow signals. Finally one half of one layer in one sector of the barrel did not work with magnetic field, due to a tripping power channel, reducing the barrel efficiency to 98.33% in RUN II, a very respectable value.

About 97% of the SST channels are reliably operational; reason for failure are the previously described short-circuited sensor backplanes, malfunctioning sensor and electronics voltage connections, but mainly failure of the control chain circuits, called control rings.

The silicon systems are only switched-ON with so-called “STABLE BEAM” to avoid losses in the unlikely case of a beam accident. To maximise the collected integrated luminosity, every effort is undertaken to reduce the switch-ON-time. In principle two control mainframes can control the 3888 SST HV power lines but the read-back of the status values takes time thus initially four and then eight units have been installed achieving switch-ON-times of initially about 4 min and finally

[26] Several hundreds of temperature and dew point sensors have been installed to have good local understanding of the environment; something missing during RUN I.

[27] Most often called sniffer pipes.

[28] System reducing the oxygen content in air to a non-flammable value, often used in oil tankers to inert the environment preventing fire.

one minute. Several beam parameter need checking before the high voltages are raised. In the first months this was under strict expert[29] control, delegating then to regular Tracker shift experts then to the CMS central crew shifter, the so-called CMS Technical Shifter and finally, after 1.5 years of operations, all checks have been fully automated. On average the automatic mechanism is 60 s faster and prevents incidents like five minutes human reaction times. Still the system prevents switch-ON in bad or not fully understood conditions.

Another interesting surprise, during RUN I was the roughly 30% increase of radiation damage in one quadrant of the innermost pixel layer due to the local radial displacement of the detector with respect to the primary beam by a couple of millimetres. More about radiation dependence on radius will be discussed in the Sect. 6.6.2.

Despite all challenges and surprises the system is robust and redundant against failures. The meticulous quality assurance during construction and design paid off and the detector will operate longer than initially designed for and thought of.

6.6.2 Signal Processing, Some Key Figures and Tracking with the CMS Tracker

This section discusses the essential Tracking and Vertexing performance of the CMS Tracker in an exemplary, educational way focussing on dependencies of design and reality – how it all fits together. It is not meant as an exhaustive description of the CMS performance giving all or even the most recent details. These are shown in much greater details here [295, 310, 312]. The next paragraphs will, therefore give a descriptive overview of signal processing, essential parameters like position resolution, Signal/Noise S/N, tracking efficiency and momentum resolution. It will also introduce the alignment challenge, detector parameter evolution with radiation during operation plus give an insight of the role of the Tracker in the CMS High Level Trigger. It should be mentioned that in the end, all results are compared with dedicated detector simulation. Basic particle interaction with the detector material plus the individual sensor element response including particle drift, induction – signal formation on strips and pixels, Lorentz angle, etc., are simulated, taking into account also alignment, bad channel list, calibration values, etc. Without having the same level of precision in simulation and detector performance no analysis is possible since the data is not well understood. These simulation are called “MonteCarlo”.

Signal Processing

As a first step, the detector needs to be timed-in with respect to the LHC clock to ensure having all signals of all pixels and strips in the optimal range of the preamplifier. Due to cable signal paths (cable lengths, positions within the detector, time-of-flight, back-end readout sequences) every single module has its own time

[29] Two individuals took 14 h-shifts at the experiment seven days a week for 6 weeks.

delay. With cosmics particles and even more important with collision data the delays are being scanned until the optimised/maximised signal has been found.

In addition, despite only small fluctuations in number of bad channels (strips/pixels), these are automatically identified on a run by run basis thanks to a dedicated software. The corresponding channels are masked in the offline reconstruction database, to optimise offline reconstruction and tracking. The bad components are also treated in the same way in MonteCarlo samples to reflect reality and to allow the best MonteCarlo/Data matching.

As described earlier, the full CMS data can neither be read out nor stored, and only 100 – 1000 Hz of events[30] are finally saved for further physics analysis. In addition only '*zero-suppressed ZS*' data is being stored, where values below threshold are suppressed. This sparsification in the pixels happens already in-situ on chip level while the SST implements this step in the back-end hardware.

In general for the SST, the analogue signal from the silicon sensors is preamplified, shaped and deconvoluted (rf. Sect. 1.10) by the front-end chip APV25; then is amplified, multiplexed, converted to light by the **A**nalog **O**pto **H**ybrid AOH and sent via optical fibres to the FEDs. Clock, trigger signals and slow control communication with the front-end electronics are managed by the **F**ront **E**nd **C**ontroller FEC boards and transmitted via optical fibres to the **D**igital **O**pto **H**ybrid DOH for each control ring of the Tracker: signals are distributed by the DOH to all **C**ommunication **C**ontrol **U**nits CCU. Finally each CCU sends signals to the chips of a set of Tracker modules, in particular: clock and trigger; it receives the data from detector slow control measurements provided by each module by the **D**etector **C**ontrol **U**nit DCU[31]. The DCU chip also contains a unique electronic identifier prohibiting later ambiguities or wrong assigning.

For the SST, the full pulse-height data for every strip is transferred to the back-end FED boards where the zero-suppression happens – this is often being called synchronous read out. Every individual chip channel has its 192 capacitive buffer cells. The dedicated zero suppression algorithm runs in the **F**ield-**P**rogrammable **G**ate **A**rray FPGAs of off-detector FEDs allowing more complex and adaptable comparisons. So-called "*pedestals*" (the baseline signal level of a given strip when no particle is present) and "*common mode*" noise (event-by-event fluctuations in the baseline within each tracker front-end readout chip) are subtracted suppressing any 'zero-event'. Zero-suppression accepts a strip if its charge exceeds the expected channel noise by at least a factor of five, or if both the strip and one of its neighbours have a charge exceeding twice the channel noise. As a result, information for only a small fraction of the channels in any given event is being sent to the CMS central **D**ata **Acq**uisition DAQ and finally, after **H**igh **L**evel **T**rigger HLT selection of events retained for offline storage. This data treatment is illustrated in Fig. 6.39. In special occasions the data from all strips inside the SST can be stored. This mode is called **V**irgin **R**aw VR and is used for calibration and during the **H**eavy **I**on HI periods,

[30] Design was 100 Hz, since 2012 more than 1 kHz events of data are being saved.

[31] DCU; situated on each hybrid, the DCU measures temperatures of hybrid and sensor, several voltages plus the sensor leakage current.

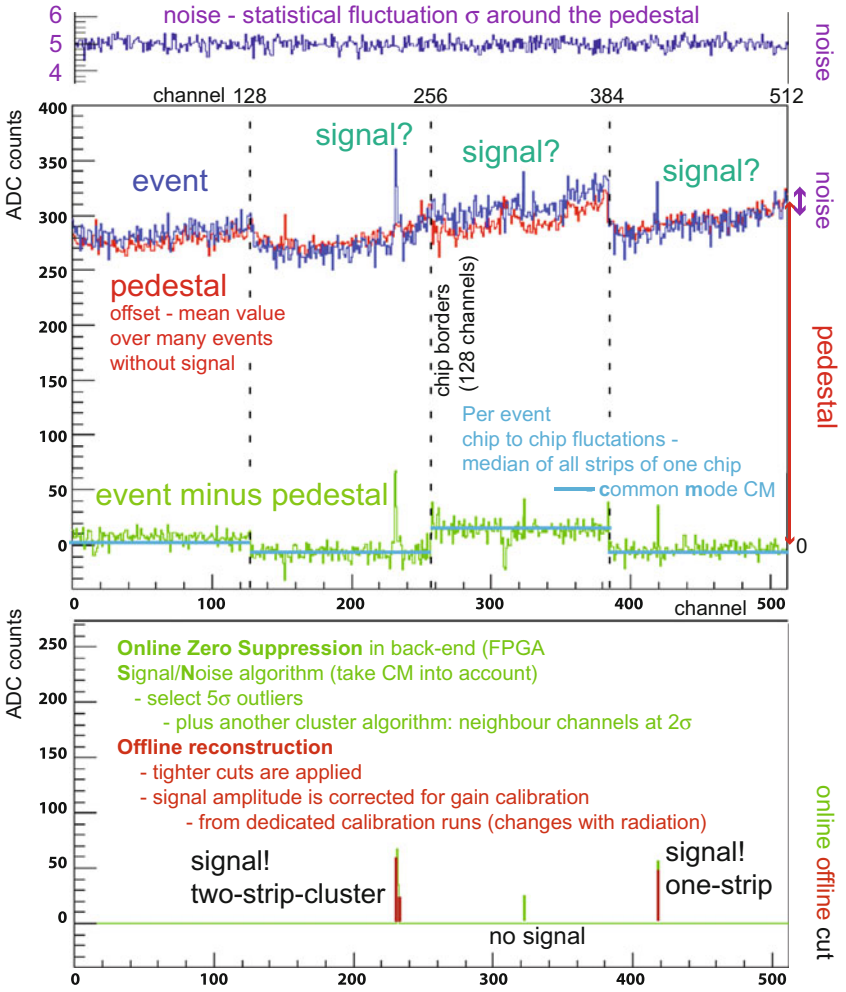


Fig. 6.39 Each strip/channel has a "*pedestal*", the mean value of the offset with a statistical fluctuation "*noise*" around it. The noise is displayed in the *uppermost part*. Below, the plot shows the raw analogue signal (dark blue), the "*pedestals*" (*red line*) and the signal after pedestal subtraction (*light green*). After pedestal subtraction, each chip has still an additional global 'common' offset. The *lower plot* then shows the result of the online (*light green*) zero suppression a cluster algorithm running on FPGAs on the back-end electronics. This is the signal information sent to offline reconstruction where gain calibration and tighter cuts, e.g. via a more refined "clusterizer" algorithm, are applied (*red*). Lastly the information is being translated into position, taking Lorentz angle and centre-of-charge into account

where the multiplicity is extremely high but the initial collision rate is strongly reduced. One SST event in VR sums up to 14 MB just for the SST while with ZS it depends on the occupancy/luminosity and sums to about 0.5 MB at design luminosity.

Regular calibration runs are necessary to determine pedestals and noise. Further cuts happen later and positions are Lorentz-Angle corrected, 10 μm (20 μm) in the TIB (TOB). The position of the hit corresponding to each cluster is then determined from the charge-weighted average of its strips positions. The uncertainty in the hit position is usually parametrized as a function of the expected width of the cluster obtained from the track angle.

With comparably low radiation levels in the SST, no sensor signal degradation is expected and noise degradation due to increase of currents can be factored in – signal-to-noise values will remain high throughout the lifetime of the detector. Still the effects need to be followed and corrected for in the offline reconstruction stages. The optical links, on the other hand, change with radiation and the edge-emitting laser diodes have four gain stages to compensate for non-uniformities and radiation effects. In addition the laser gain is temperature dependent. Also differences in voltages and temperature in the front-end have an offset effect. The SST front-end chips send a so-called digital *tick mark*, a normed signal, allowing to adjust the laser gains to achieve a homogeneous *tick mark* at the back-end – such a calibration run is taken about every other month. Altogether the gain of every single APV chip[32] is then a parameter in the offline reconstruction allowing also dE/dx measurements and to determine if a small signal came "*out-of-time*", e.g. from a previous bunch crossing.

In the Pixel system, zero-suppression happens already at the front-end electronics sending only pulse-height information and pixel hit addresses to the back-end ady only pulse height information of a 'hit' pixel is being sent to the chip's periphery and buffered there, waiting for a L1-Trigger signal (rf. Sect. 1.11). The front-end single-pixel-threshold is about 3200 electrons (∼1500 electrons for Phase I). The system needs regular, about monthly, calibration runs due to radiation but also for environment changes, e.g. temperature. Charges are being injected into each pixel, and the efficiency is being determined. Also the FED optical receiver offset needs adjusting to keep the baseline level of the signal in the middle of the ADC range. The hit pixels signals are later correlated to pixel weighted clusters and the cluster shape determines the final hit location and resolution. In first order a centre-of-charge corrected for the Lorentz-Angle gives the location, this is used mainly in the High-Level-Trigger stage. In the full offline reconstruction so-called "*pixel templates*" are being used. These templates "know" a priori the expected pixel cluster shape depending on the spacial sensor placement with respect to the primary interaction zone; e.g. "knowing" the principle incident angle of the particle but also taking into account the Lorentz angle, the operation voltage (thus field configuration and thus drift behaviour) and the possible partial depletion configuration, etc. into account, all following the evolution with radiation.

Some more features of the pixel system: With sparsification in the front-end chips, the limited size of the internal buffer of the readout chips cause a dynamic inefficiency, where hits are lost, increasing with instantaneous luminosity and with trigger rate. At design luminosity the dynamic inefficiencies is around a couple of percent in layer 1

[32] It showed that a gain per individual channel was not necessary.

Table 6.4 Measured Hit Resolution of the CMS SST with analogue readout compared to the binary case. The ranges of resolutions reflect the difference in cluster sizes (how many strips flag a signal), in rϕ but mainly in pseudorapidity η. With larger η thus larger angle thus larger path length in the silicon, the net deposited energy from a traversing charged particle increases. In first order, larger clusters (more strips) allow a better centre-of-charge determination, even more true for thicker sensors. This effect is much better illustrated in the next Fig. 6.40 (Pixel case) where the angle translates directly to cluster size and resolution [295]

Sensor layer	Pitch (μm)	Resolution (μm)	Binary resolution (μm)
TIB 1 – 2	80	10.1 – 19.1	23.1
TIB 3 – 4	120	20.8 – 29.5	34.6
TOB 1 – 4	183	16.9 – 42.3	52.8
TOB 5 – 6	122	11.8 – 26.6	35.2

while being too high for twice the design instantaneous luminosity.[33] This bottleneck has been resolved by the Pixel Phase I upgrade (see Sect. 6.2). Single-event upsets temporarily cause loss of information at a rate of approximately two readout chips per hour – an automatic software recovery mechanism has been established. Finally, readout errors signalled by the FEDs depend on the rate of beam induced background.

The whole process of operation and calibration is more complex than described above and requires regular but not constant effort and presence of experts. All processes are highly automated to allow also inexperienced shifters to monitor data taking and take calibration runs, while the full detector analysis is still in the regime of experts.

The Essentials—Signal/Noise and Point Resolution

During LHC operation, in the SST, the APV25 runs in deconvolution mode, where a weighted average of three samples is formed, effectively shortening the pulse to 25 ns and containing the full signal within a single LHC bunch crossing. This shorter pulse shape comes at the cost of a noise increase of 30%. The resulting Landau distribution, after correcting for the track incidence angle (path length corrected signal-to-noise ratio), exhibits a **M**ost **P**robable **V**alue: MPV (S/N) 18 for the thin sensors at the inner radii and MPV (S/N) 22 for the thick sensors at the outer radii. At first order, the point resolution in an analogue system is $\sigma_x \propto \frac{\text{pitch}}{\text{signal/noise}}$ compared to $\sigma_x = \frac{\text{pitch}}{\sqrt{12}}$ in a binary system. In a real large system, also angle and path length, and corresponding cluster size (signal distribution over several strips), thus location of the module with respect to the primary vertex, play a role. The point resolution, an important factor to measure the sagitta (see definition Appendix A), defining the momentum resolution, of the CMS Strip Tracker are given in Table 6.4.

The correlation of resolution to angle and cluster size is even clearer visible in the Pixel system with a pixel cell length of 150 μm. Figure 6.40 shows the spatial z-resolution versus incident angle in the Pixel Barrel part as derived by the "template" method; the plot also shows the corresponding cluster sizes (number of pixels

[33] It should be mentioned that the LHC is performing much better than the design specification – about twice in 2017.

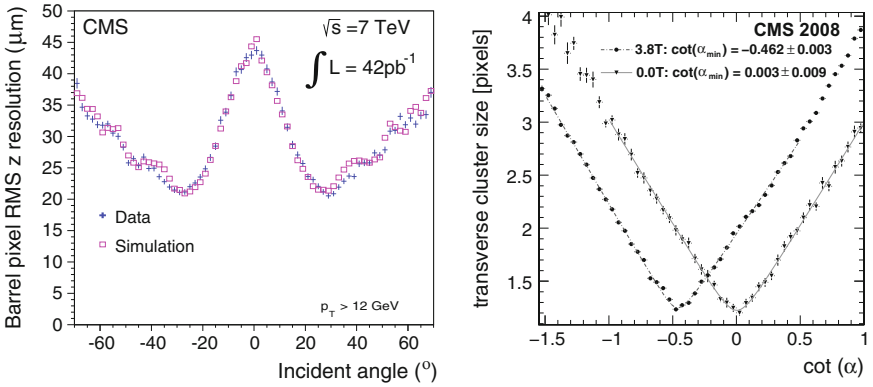


Fig. 6.40 The z-resolution of the Barrel Pixel Detector versus incident angle with ionising particles originating at the interaction point z = 0 [295]. The *right* corresponding plot shows the "cluster sizes versus angle" leading to the resolutions in the *left* plot; optimal at $\cot(30\,^\circ) = 0.5 \mathrel{\hat{=}} 2$ pixel cluster. This plot is also called Happy Face and used to monitor the detector health, meaning when the plot 'smiles', everything is in order. The *right* plot also nicely demonstrates the shift of cluster due to the Lorentz angle / Lorentz shift in the magnetic field [294]

above threshold). With increased angle the path length increases thus the amount of ionisation (charge/signal) increases but also the number of pixel collecting the charge changes. At incident angle single pixels are hit and the resolution is not much better than binary while the best resolution is achieved when "a couple" of pixels participate (two-pixel cluster). At a more shallow angles many pixels see charges below threshold, part of the charge is thus "lost" and the resolution worsens. With even more shallow angle also the multiple scattering increases.

The presented resolution is about optimal thanks to the "template" approach, improvement would theoretically be possible at the detector design state to optimize pixel (sensor cells in general) cells depending more on location. But in all practicalities, chip cells are what they are and the more differences the less practical the realization of the detector. Changes of cell geometries are normally only done for different layers and barrel and forward partitions or in more small "boutique" like detectors like DELPHI, featuring changes even inside layers.

Alignment

As prerequisite for tracking the knowledge of the precise position in 3D space of the sensors down to the individual cells with respect to each other and also with respect to the global CMS coordinate system is of utmost importance – the so-called alignment. The paper [313] gives a very deep insight on the matter while this paragraph only gives a simple description.

Figure 6.41 figuratively illustrates the issue and since it is impossible to construct the tracker mechanically with a ∼10 μm or better precision, which is necessary to

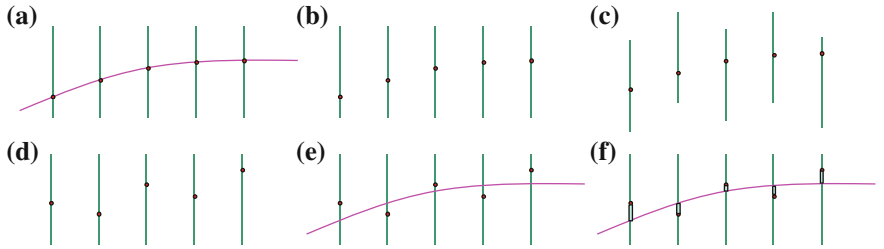


Fig. 6.41 A charged particle traverses the detector (bent by the magnetic field) (**a**) and produces hits in the sensor planes (**b**). The sensors itself might be displaced with respect to the expected ideal geometry (**c**), and without that knowledge or correction the tracking algorithm would see (**d, e**) a zigzag around the real track and thus reconstruct a wrong track. With a good statistics of high purity reconstructed tracks, the deviations from the reality – the residuals – can be determined **(f)** thus understanding the correct geometry and feeding it back into the reconstruction algorithms – telling the database the real position of the sensor cells [10]

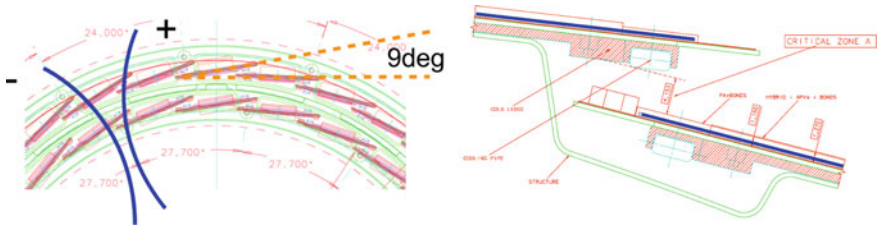


Fig. 6.42 Modules are arranged with a dedicated overlap to ensure hermeticity but also allowing for module to module alignment inside layers, since some tracks passes both modules here. Given the space constraints this is mechanically challenging

achieve the tracking design performance, we have to "align" it later in the geometry description database. The built-geometry already takes the need of module to module alignment into account by implementing an overlap of sensors inside the individual layers – this is illustrated in Fig. 6.42.

For decent track reconstruction, one needs to achieve $\sigma(\text{align}) - \sigma(\text{intrinsic hit}) = 10 - 30\,\mu\text{m}$. Especially in the pixel, a miss-alignment of a few tenths of micrometers would seriously degrade the b-tagging performance. 15,148 SST plus 1440 pixel modules need to be aligned corresponding to 16,588 * (3 positions + 3 orientations + 3 deformations) individual parameters. The large number of individual modules and their arrangement over a large volume with some sensors as far as $\approx$6 m apart takes the alignment challenge to a new stage compared to earlier experiments. All in all, on the order of $\mathcal{O}(200{,}000)$ parameters are being simultaneously determined, together with O(10M) track parameters treated as nuisance parameters in the fit, with the help of dedicated computers with 256 GB of RAM each. To achieve this several millions of high purity tracks of different track topologies are necessary. The module parameters are then updated by minimizing the χ^2 of the residuals (deviation from

correct hit position, see also Fig. 6.41). The final complete set of derived parameters is then simply called "Tracker Geometry".

The "Tracker Geometry" takes the following items into account:

1. internal alignment, mainly adjusting the positions and angles of the modules relative to each other
2. absolute position and orientation of the tracker including "Weak Modes" (see below)
3. bow intrinsic to sensors
4. kinks between sensors of a two-sensor module
5. stability: movements over time, especially at certain operational transients, like thermal cycles or magnet ramps

The high number of collision tracks (isolated muons and minimum bias events[34]) are perfect for the internal module to module alignment and to determine the sensor bow but fail to ascertain the absolute position and global orientation of the tracker. Tracks from the origin only traverse one side of the tracker thus do not connect left/right or up/down. They also have no means to constrain the distance of the endcaps or the distance between endcap and barrel. Any coherent change of alignment parameters for many modules, e.g. of one full sub-detector, cannot be understood by a single kinematic topology. These coherent global changes are called "Weak Modes". Examples of Weak Modes are global twists, skews, barrel detector being off-centre with respect to the endcaps, global bow, z-expansion of endcaps, etc.

A kinematic track diversity is necessary to address all aspects. Especially cosmic ray particles in jargon *cosmics*, having a completely different topology, e.g. breaking the cylindrical symmetry, connecting top/bottom and determine endcap distances, are of great significance. Beam halo tracks can also add important information to constrain the endcaps. Straight muons without magnetic field are useful but with magnetic field the particle momentum is known thus a handle on the multiple scattering is given. Muons from $Z \rightarrow \mu^+\mu^-$ or objects with invariant mass of a resonance give additional constraints.

The intrinsic bow of the sensors introduces a position error up to 10 μm and is compensated in the "Tracker Geometry".

The SST has proven to be mechanically very stable over time; with temperature ramps from RT to −15 °C, only changes of ≈50 μm have been observed relaxing back to the original position at RT – no hysteresis. More information about the measuring method can be found in [297]. Intrinsically also the Pixel detector is very stable but with thermal or magnet cycles the two mechanically unconstrained barrel half-shells can move several tens of micrometres up to a maximum of 100 μm in z. Collisions and cosmics are registered continuously and the full module level alignment is being tuned about 1 – 3 times per year while the pixel global structure position is being checked for every LHC fill and corrected for when necessary. In addition at the start of the year a so-called CRAFT and CRUZET campaigns (**C**osmic

[34] **M**inimum **B**ias MB events with inelastic scattering but minimum detector activity, not the high momentum or high multiplicity part, a special MB trigger exist.

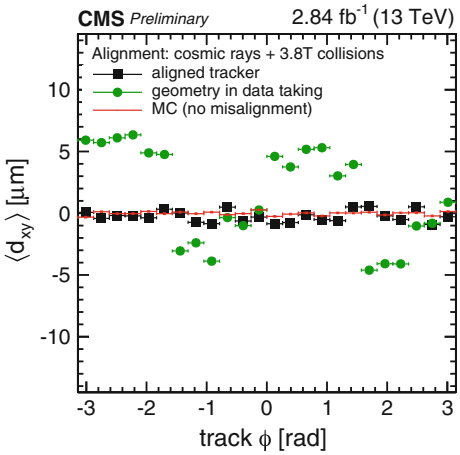


Fig. 6.43 At construction, sensors are placed on modules with a certain precision, modules are placed on bigger structures with another certain precision and structures are assembled themselves with yet another certain precision. The whole structure then experiences magnet-, power-, cooling cycles thus the absolute position in space is, to a certain extent, an unknown. Fitting numerous tracks from cosmic rays and collisions allows a virtual repositioning of all active elements. This mathematical, statistical procedure is called alignment. The plot shows the mean of the track-vertex residuals for the case of ideal positions (*red line*); the geometrical real positions derived by data taking (*green*) and final 'reality' after software alignment process (*black*). The strips of the outer Tracker are aligned to better than 10 μm while the pixel barrel even to about 2 μm. The figure shows the ideal positions; the geometrical knowledge (*green dots*) at the beginning of 2015 and the geometrical alignment at the end of 2015 (*black squares*) [217, 300]

Run at **F**our **T**esla and **C**osmic **R**un at **Ze**ro **T**esla) are being done collecting a statistically significant sample of cosmics. In addition, cosmics are registered during inter-fill periods.

Figure 6.43 demonstrates the final alignment in 2015 being very close to optimal. It is clear that the initial precision, as built, need not be much better then 0.5 – 1 mm and even larger deviations[35] can be aligned during the commissioning period, given enough time and a statistically significant number of tracks.

Tracking at CMS – an Example

Basically, tracking is a clever "road search". The first space-point plus an assumption of its origin (beam spot) gives a first possible track direction (vector). One now 'searches' the next space-point within a 'search window'. This 'search window' is predicted by taking detector geometry, magnetic field, possible momentum and potential combinatorics into account. The first combination of points gives the 'track seed'. With a seed, the position in the next plane is "predicted"; the "measurement"

[35] One extreme example: in 2015, one pixel forward detector was globally located 3 – 5 mm away from its foreseen place, given that it was not properly fixed and the final cable connection pulled it slightly. This was full recognised and mitigated by software alignment.

is then considered and, with a weighted average, merged with the "prediction" – filtered. The prediction/filtering improves with each step and the search window size shrinks. The algorithms are based on the so-called **K**almann **F**ilter KF[36] [39, 40, 110, 206]. The KF also takes potential multiple scattering and energy loss as stochastic deviation from the straight path into account. Material description and inhomogeneities of the B-field information are also included.

In general it is a local process, but still, the reconstruction of trajectories of many charged particles is a computationally challenging task. This is especially true for large combinatorics in high occupancy environments, where more than one hit candidate is statistically found in the "predicted" location/window. Due to the high occupancy thus high combinatorics CMS utilises an *Iterative Tracking* approach, where the initial iterations search for tracks that have less combinatorics (tracks with many pixel hits, with a relatively high momentum to minimise the multiple scattering, with a strong compatibility with the beamspot and a rather small search window).

With design luminosity at the LHC the detector encounters 20 interactions resulting in a total average of above 500 tracks. After each iteration, hits associated with tracks are removed, thereby reducing the remaining combinatorial complexity, and simplifying subsequent iterations in a search for more difficult classes of tracks (e.g., low-p_T, or greatly displaced tracks).
In general each iteration consists of five steps:

1. seed generation with an initial track candidate estimating the initial trajectory and uncertainty
2. track finding utilising KF as described above
3. track fitting with best possible estimate of the parameters of each trajectory
4. track selection sets quality flags, and discards tracks that fail specified criteria
5. remove hits associated with identified tracks

The occupancies in the Pixels, with 0.02%, are one to two orders or magnitude lower than in the strips system with 1 – 3%. With a much higher particle density nearer to the interaction point this is counter-intuitive, but Pixel cell sizes are much smaller: $100\,\mu\text{m} \times 150\,\mu\text{m}$ compared to $80\,\mu\text{m} \times 10$ cm. The track seeds are therefore found and reconstructed in the Pixel system. High channel granularity is the key for good tracking but it comes with high sensor and electronics channel number thus higher power consumption due to larger cable mass and cooling circuits thus larger material budget thus more multiple scattering – a balance is necessary. In dense environments, like within jets, the correct hit-to-track association is even more challenging.

The iterations work inside out (1) pixel triplets then (2) pixel pairs, recovering gaps or non-working modules or dynamic inefficiency, (3) non-pointing (displaced) tracks with pixel triplets, (4) mixed pixel/strip triplets; then inner strips system (5) TIB/inner TEC and outer strips system (6) TOB/ outer TEC. This is a rough basic

[36] Kalman Filters have a broad use-case namely navigation, radar tracking, stock market prediction. Deviations as from wind, money exchange rates, *multiple scattering* can be processed as stochastic noise.

concept since the algorithms are evolving. Generally, iterations beyond the first one improve the acceptance either in p_T or in displacement from primary vertex or take nuclear interactions into account. They also recover the full hermeticity of the detector by considering tracks where the particle passed through a crack or non-working element.

In a nutshell, some contradicting ingredients for high quality tracking and excellent momentum resolution are (rf. also Appendix A on p. 343):

- **key ingredient: good channel granularity** thus low occupancy thus low combinatorics
  - best would be to only have one track in the search window
  - low combinatorics especially important in the seeding layers
- no deviation from track – no multiple scattering – no mass
  - algorithm compensates
- start at lowest possible radius
- have a large lever arm – high radius
- very good point resolution
- any additional constraint helps, e.g. vertex; info from other sub-detectors

The schematic Fig. 6.44 illustrates track paths for different p_T and different η inside the 4 T solenoid field of the CMS Tracker. With this, it becomes clear that tracks below $p_T \sim 1$ GeV 'loop' inside the detector (depending on their initial η direction. All tracks with the same transverse momentum naturally have the same track radius R and sagitta s (rf. formula A.1 in the Appendix). Only the transverse, orthogonal track component with respect to the solenoid B-field contributes to the momentum measurement (bending in ϕ).

Figure 6.45 illustrates the transverse momentum resolution p_T for muons with fixed p_T against pseudorapidity η (left) and against momentum (right). Figures for pions look similar, resolution for electrons is worse, for more details refer to [295]. The degradation of p_T resolution for higher η stems from the decreasing effective lever arm L (see schematic Fig. 6.44). This effect is even more pronounced for high p_T tracks. The step at around $|\eta| = 1$ is due to the gap between Tracker barrel and endcap disks and different resolutions in the disks itself. Note the extra complication for the endcap (TEC and TID), where the accurate measured coordinates are z and ϕ and with wedge shape sensors also the pitches thus point resolutions change within an individual sensor, compared to r and ϕ measurement with fixed pitch for the barrel configuration. The optimum resolution is achieved for particles around 3 GeV. At lower momentum, the resolution is fully dominated by multiple scattering and its value reflects the amount of material traversed by the track (more at higher η). At higher momentum the point resolution and even small miss-alignments play a role – the track is simply straight and a sagitta is difficult to measure precisely. For high p_T muons the larger lever arm taking the muon detector into account helps a lot – see Fig. 6.55.

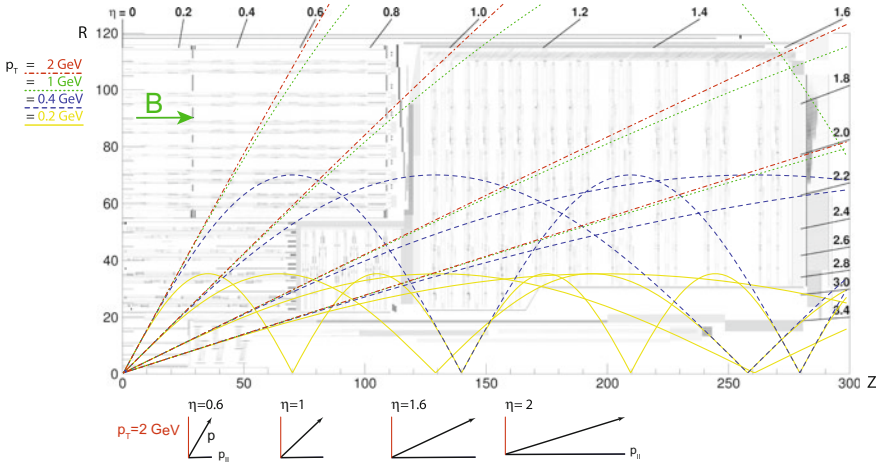


Fig. 6.44 Tracks in CMS – a schematic. For $\eta = 0.6$, 1, 1.6 and 2 the plot features four tracks with the same fixed p_T: 0.2 (*yellow*), 0.4 (*blue*), 1 (*green*) and 2 (*red*). All tracks of the same transverse momentum p_T have the same track radius R and sagitta s in the solenoid B-field; their momentum p though changes with initial direction. Tracks, leaving the tracker at $|\eta| > 1.6$ (in the endcap not in the outermost radius), have a shorter effective lever arm L. For example the *red* 2 GeV track has an $L = 1.1$ m for $\eta = 0.6$, 1 and 1.6 but only about 70 cm for the $\eta = 2$ direction. The 'looper' 0.4 GeV (0.2 GeV) track does not reach the Tracker radius anyhow thus sees only $L = 70$ cm ($L = 35$ cm) or $L = 60$ cm for the $\eta = 2$ case. Not demonstrated here but the low energetic tracks are also affected more by multiple scattering. The cartoons on the *bottom show* the p components (p_T and $p_\parallel$) for the $p_T = 2$ GeV tracks for the different η directions. Figure from [260]

In short

- versus η: path length (multiple scattering), switch from barrel to endcap configuration and effective lever arm L defines the p_T resolution
- versus p_T
  - at low p_T multiple scattering dominantly degrades the p_T resolution
  - at high p_T point resolution and large lever arm are the important factors

Another complication is that particles of different type interact differently with the detector materials and their paths are different. Expressing these differences, tracking efficiencies for muons, pions and electrons are displayed in Fig. 6.46.

The efficiency for muons is basically 100%, with no strong interaction and basically no electromagnetic radiation (Bremsstrahlung), they traverse the full volume and the multiple Coulomb scattering and energy loss is effectively included in the KF. Tracking information from the outer muon detectors give a long lever arm improving momentum resolution significantly, especially at high momenta (straight tracks).

Charged pions, additionally, are subject to elastic and inelastic nuclear interactions, deviating grossly from the original path. Tracks can be interrupted or two tracks with less hits or no tracks at all are being identified. They are also stopped in the calorimeter (giving another crude space point).

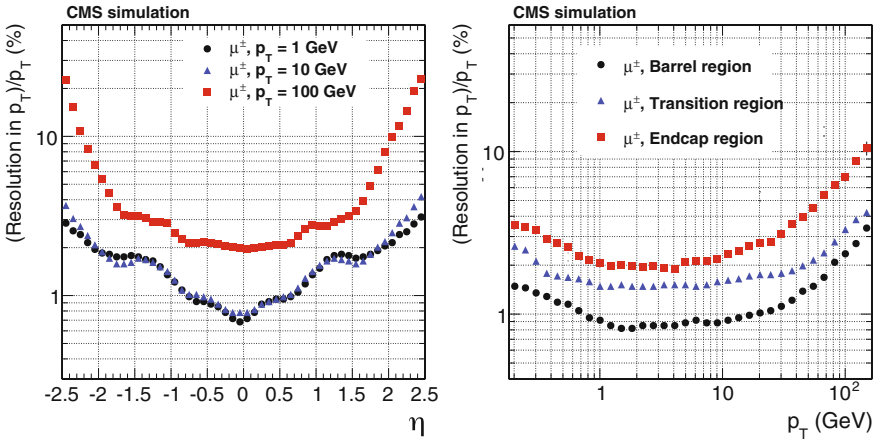


Fig. 6.45 p_T resolution of muons for different energies versus η on the *left* and versus transverse momentum p_T on the *right* [295]

Electrons are much more complex due to their additional energy loss due to Bremsstrahlung, a highly non-Gaussian process and therefore non-standard for the KF. About 35% of electrons radiate more than 70% of their initial energy before reaching the **e**lectromagnetic **cal**orimeter ECAL that surrounds the tracker. Sometimes they do not even reach the outer tracker layers thus produce less hits. In general, their path is less stringent and with momentum loss along the way, changes bending thus changing estimated 'search window' position. To compensate, the energy and position measured in the electromagnetic calorimeter plus the assumption that the electron originated from the beam spot, constrains the trajectory further [305].

Another interesting case are very displaced tracks (identified in later iterations) from photon conversion where the vertex (e^+e^-) can be anywhere with material present. Radiated photons (e.g. from electron Bremsstrahlung) can convert to electron-positron pairs producing new tracks in the middle of the tracking volume, complicating the situation even further. The effect is enhanced in the endcap section due to higher material budget especially in the interface region. For electrons, the highest quality information comes from the pixel region before Bremsstrahlung kicks in.

Figure 6.47 shows the space points where photon conversion happened, clearly reconstructing the modules and mechanical supports as well as cooling and cables. Especially impressive is the correlation of photo conversion and cooling pipe presence.

Finally all tracks are extrapolated towards the primary or secondary vertices. Figure 6.48 shows the resolution of the primary vertex. The more tracks, or better said richness in kinematic diversity and higher integrated momentum, the easier to constrain the vertex and thus the better the resolution. Even with the worst resolution, primary vertex miss-identification is practically excluded.

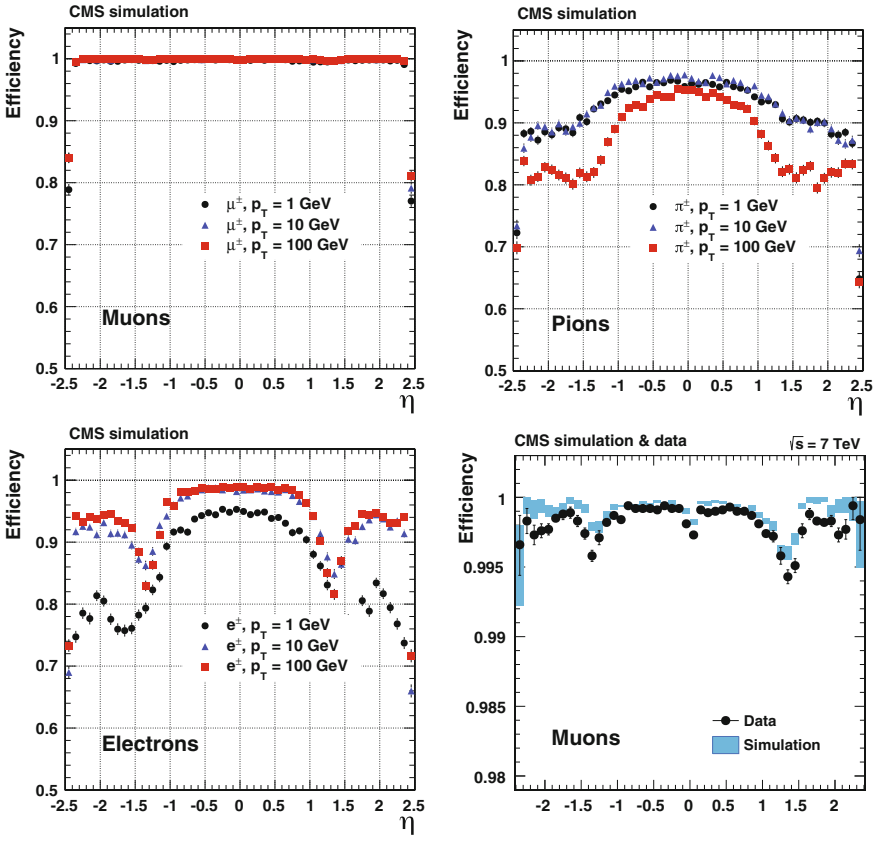


Fig. 6.46 Simulated tracking efficiency for muons, pions, electron and muon efficiency compared with data. The *lower right* plots display a zoom of the first plot and compares data to simulation [295]

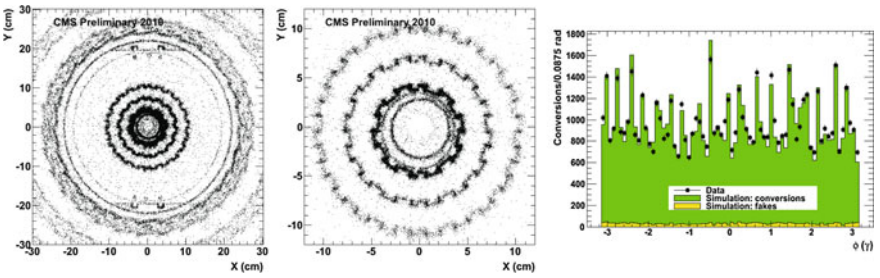


Fig. 6.47 Photo conversion as reconstructed with data. The spacial occurrences clearly depict the support structure module and services. The *middle* figure is a zoom of the *left* one. Also the beam pipe is clearly discernible; it is slightly displaced with respect to the pixel envelope. The *right* figure shows the correlation of photo conversion and cooling pipes locations [307]

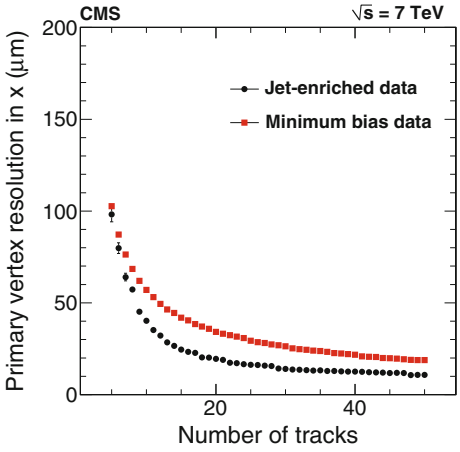


Fig. 6.48 The jet-enriched sample has significantly higher mean p_T, resulting in better resolution in the track impact parameter and consequently better vertex resolution, approaching 10 μm in x for primary vertices using at least 50 tracks. The resolution of the reconstructed vertex position is driven by the pixel detector, since it is the sub-structure that is closest to the interaction point and has the best hit resolution [295]

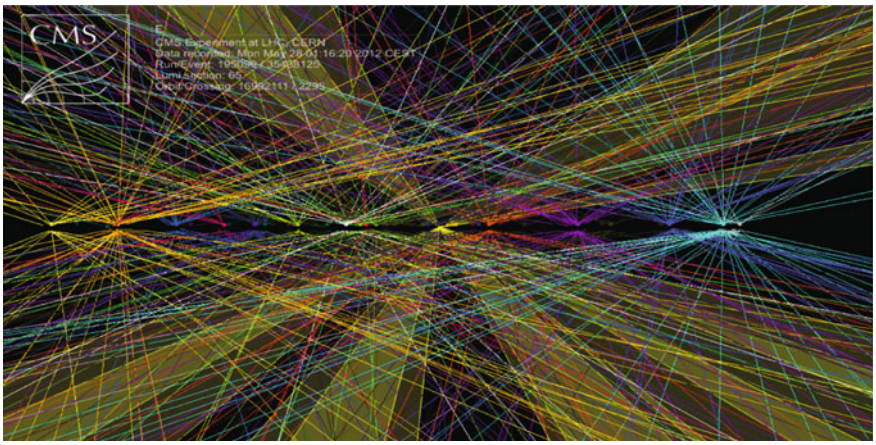


Fig. 6.49 The situation at high pile-up of about PU = 45 with about 1500 tracks cross the CMS tracker every 25 ns. The spacing between individual vertices is about 1 – 2 mm [Courtesy of CMS/CERN]

Figure 6.49 gives an impression of the real environment – a zoom into the primary interaction zone depicting the multiple reconstructed/interpolated tracks at about nominal luminosity. More event displays with lower and higher **p**ile-**u**p PU can be visited in Fig B.6 on p. 355 in the Appendix.

High-Level Trigger

Another key role is the Tracker's extensive usage in the **H**igh **L**evel **T**rigger HLT for all physics event streams [274]. The Level-1 trigger system is based on custom electronics and reduces the input rate of 40 MHz bunch crossing[37] by a factor of 400 down to 100 kHz. The Level-1 decision must be made in 4.8 μs,[38] corresponding to the maximum storage capacity on the front-end chips. The HLT, a software trigger, relies on commercial processors and reduces the stored data by another factor of 1000 down to 100 Hz (since RUN II 2015 more like a factor of 100 down to 1 kHz). While the tracker is not participating in the Level-1 decision, it plays an important role in the HLT. Simplified and much faster versions of the full track reconstruction algorithms, with reduced accuracy, are used in the HLT.

The three basic principles are

- pixel only tracking
- "*regional tracking*": The HLT track reconstruction is restricted to the **R**egion(s) **o**f **I**nterest RoI, which is defined by the information from the Level-1 triggers from the muon and calorimeter systems.
- "*conditional tracking*": Cuts that stop reconstruction during event selection are introduced, like "momentum is definitively above or below threshold", "N good hits found". Often four or six space points are enough for track reconstruction. The transverse momentum p_T can be calculated with only a partial track, since it does not significantly[39] change up to the end of the track – good enough for triggering using strip layers after pixels.

These fast reconstructions subsequently allow general p_T cuts, defined by physics requirements (momentum cuts, keeping events most relevant for physics studies, are for example 0.9 GeV for b-tagging or up to 20 GeV for muons). Low momentum coiled tracks are not reconstructed.

As an example a b jet HLT trigger takes a Level-1 trigger from the calorimeter (jet), the track seed for conditional track reconstruction is created around the Level-1 jet direction. Tracks are solely reconstructed in a cone (RoI) of $\Delta R < 0.15$ around the jet direction. This makes a jet direction refinement possible, to distinguish if it comes from the primary vertex or not. In short, the number of tracks with transverse impact parameter statistically incompatible with a track originating from the beam-line are counted – thus originating from a secondary vertex.

A second example is the τ trigger. A narrow jet in the calorimeter defines the RoI – cone around a jet. The τ hypothesis requires an isolated track associated to the jet. Conditional and regional tracking in the RoI is performed. The τ hypothesis is falsified if no track is found in a narrow cone around the jet (signal cone) or an additional track is found in a second broader cone (isolation cone) – *regional tracking*. To save CPU time, the track is not fully reconstructed and only a loose primary vertex association is requested – *conditional tracking*. Later at the offline reconstruction stage,

[37] With a pile-up of PU = 25 one has about 10^9 interactions every second.

[38] 192 cells times 25 ns.

[39] Except for electrons – Bremsstrahlung.

the full track and decay vertices are reconstructed. To reconstruct a full "good" track, the track seed plus five helix parameters (two positions, two angles plus curvature) have to be determined.

In general not all stages of iterative tracking are executed and good use is being made of pixel hits with intrinsic 3D information. For example seed tracks consists of only pairs[40] of hits (not triples) and in addition the seed track has to originate from within few millimetres of the primary vertex.

Tracking uses only about 20% of HLT resources, due to the fact that track reconstructions is only executed after other requirements have been satisfied, reducing the rate significantly, e.g. jets, or high energy in the calorimeter or tracks in the muon system. Further steps then give the 3D primary vertex and impact parameter to have a handle on *b*-tagging.

Evolution of Sensor Parameters with Radiation at CMS

Are the parametrization models describing the evolution of sensors parameters as presented in Chap. 2 valid for long term operations and do they have projective power for the future? Some early publications about how detector evolution with radiation is monitored in the LHC detectors can be found here [24, 133, 345].

Radiation levels span more than three orders of magnitude within the tracker volume. With mainly charged particle irradiation in the centre and backscattered neutrons from the calorimeter in the outer radius, basically every particle mix and level presents itself – several thousand radiation campaigns in one go. Figure 7.1 on p. 292 gives an impression of the radiation level distribution although the quantitative levels show the **H**igh **L**uminosity LHC HL-LHC case. In addition, temperatures are different during maintenance and operation periods thus also annealing processes can be investigated. There is also a broad spread in sensor temperatures in the SST, due to some strongly degraded cooling contacts and some inoperable cooling lines (see Fig. 6.50a). In CMS, every SST module features an ASIC (DCU) measuring sensor and hybrid temperatures plus sensor leakage currents and hybrid low voltages during operation. This allows a fine granular monitoring for the different cases of radiation and temperature (annealing).

As a reminder and to put the following plots and results in perspective:

- the Tracker is *being operated* cold ($T_{silicon} = -10\,^{\circ}\mathrm{C}$) to lower the leakage current of the moment. Reminder: $I(T) \sim T_{silicon}^2 \cdot e^{-\frac{E_{eff}}{2 \cdot T_{silicon} \cdot k_B}}$
- the Tracker is *being operated* cold to "freeze" diffusion processes thereby preventing reverse annealing of N_{eff} (depletion voltage); $T_{silicon} = 0\,^{\circ}\mathrm{C}$ or even $+5\,^{\circ}\mathrm{C}$ would be sufficient
- the Tracker is *maintained cold* to "freeze" diffusion/annealing thus preventing reverse annealing (depletion voltage)
- the Tracker is being warmed up *during maintenance* for some period to profit from beneficial annealing (depletion voltage) and current annealing

[40] Worth to note: Since 2017 with the phase I pixel detector, the pixel seeding is composed of quadruplets with triplets as backup instead of triplets with pairs as backup.

- annealing of leakage current always decreases the current and the higher the temperature the faster the decrease – thus higher temperatures are beneficial but the requirement is contradicting the requirements above to prevent reverse annealing of N_{eff} (depletion voltage)

Figure 6.50 shows so-called Tracker maps for different parameters (temperature and leakage current increase with radiation normalised to volume or to volume and temperature). Every small square, wedge and triangle (in case of stereo modules – two sensor closely spaced back-to-back) represent a single module and the colour shows the parameter value (°C or $\frac{\mu A}{fb^{-1}cm^3}$). The SST barrel layers are simply unfolded flat and endcap disks are displayed separately. For example "TOB1" shows the innermost TOB barrel layer which has stereo-modules thus triangles inside the squares representing the inner (triangle) or outer (triangle) module of the sandwich (square). All black fields are 'bad components', i.e. modules not participating in the readout or showing false DCU readout (see first map for details).

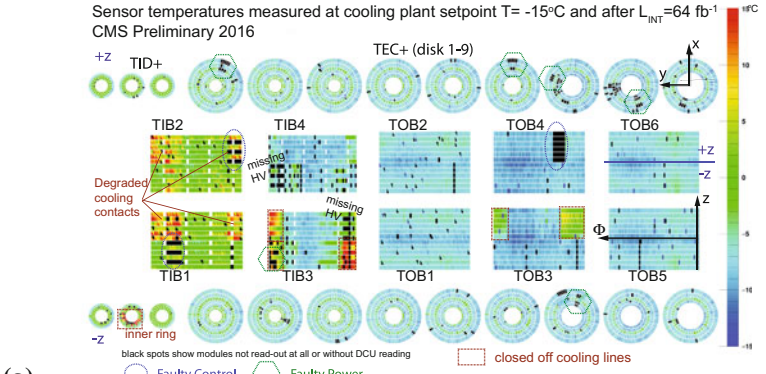
The first plot **(a)** presents the individual sensor temperatures for a cooling plant set temperature of −15 °C illustrating a large temperature spread. The double module layers (TIB1,2 & TOB1,2) are hotter per se, while the individual hot modules or regions within layers represent degraded cooling contacts or non-operational cooling loops. For higher operating temperatures and the same radiation exposure, we expect higher operation current but smaller current increase due to simultaneous annealing.

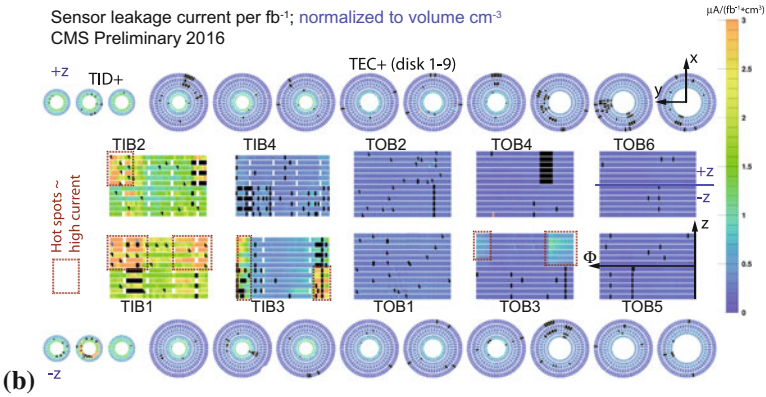
The second map **(b)** shows the corresponding leakage current increase per fb^{-1} normalized to volume[41] after several years of operation thus significant radiation exposure. The plot is not normalized to temperature thus the warmer TIB layers show a higher leakage current and the hot spots within a layer exhibit significantly higher currents, demonstrating the need to *operate* cold; about every 7 K the current doubles/halves. At any given moment, it is possible to "dial" the desired current by changing the temperature within the limits of cooling power and ambient humidities. Due to the exponential current-temperature dependence 'subtleties' as current-radiation dependence, meaning layer-to-layer dependence, are not discernible here.

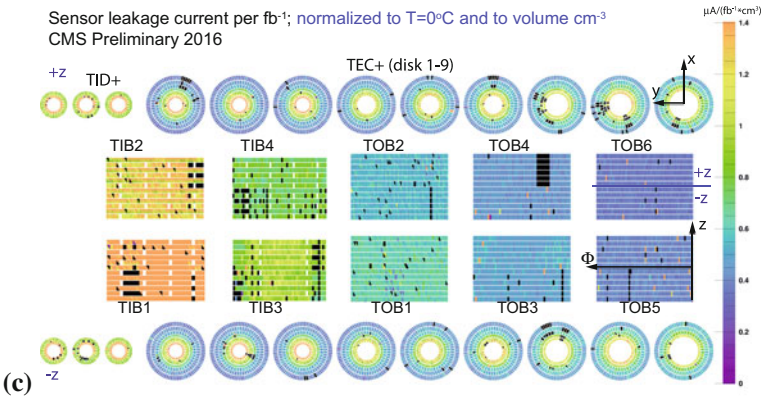
With the next map **(c)**, normalised to volume and temperature (T = 0 °C), a clear layer-to-layer dependence becomes visible corresponding to the different radiation exposure levels. The radial radiation dependence is also nicely visible in the endcap disk spanning over the full radius. Looking closer, one discerns higher currents for disks at higher z position (disk 8&9) meaning closer to the endcap calorimeter thus being effected by neutron backscattering[42]. Still the differences in currents presented do not reflect solely the difference in radiation exposure but the whole thermal history. As discussed in Sect. 2.1.2 the effective α-parameter in $\frac{\Delta I}{V} = \alpha \Phi_{eq}$ is temperature dependent; thus the evolution with radiation differs for different temperatures thus here differs for the different modules. The evolution of currents is a convolution of crystal damage induced by radiation and the parallel annealing process – the higher the temperature, the more annealing (self-repairing), the smaller is α. TIB Layer 1

[41] E.g. TOB has 2 modules and each TOB sensor is 5/3 times thicker than a TIB sensor.

[42] A moderator has been installed instrumented between endap calorimeter and Tracker to limit the neutron exposure.







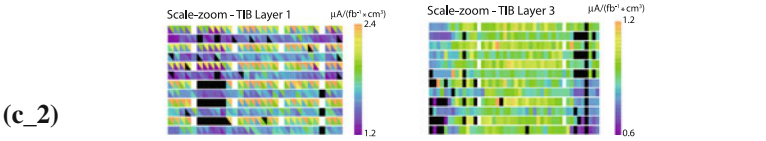


Fig. 6.50 (Continued)

◀ **Fig. 6.50** The figures show so-called Tracker maps where the squares, wedges and triangles represent individual modules and their parameter values are *colour* coded – mind the different scales and units. The larger rectangles represent the barrel layers 1–10 while the large circles show the endcaps spanning the full radius of the tracker. As an example TOB L1 means Tracker Outer Barrel – Layer 1; the black spots show "bad" modules as detailed in plot (**a**). TIB L1+2, TOB L1+2, TID R1+2, TEC R1+2+5 feature double modules and therefore have a higher temperature. The features of the four plots are described in the text above. TEC disks of higher number are located at higher z thus closer to the endcap calorimeter. In order, the maps show (**a**) silicon temperatures for cooling plant set temperatures of $-15\,^{\circ}\mathrm{C}$, (**b**) sensor leakage currents (ΔI) increase per fb^{-1} normalized to volume for a cooling plant set temperature of $-15\,^{\circ}\mathrm{C}$, (**c**) leakage current increase, normalized to volume and temperature ($\mathrm{T}=0\,^{\circ}\mathrm{C}$) and **(c_2)** a 'scale-zoom' into TIB L1 and L3 to overcome the saturation effect of (**c**). All values were measured after an integrated luminosity of $\mathcal{L} = 64\ \mathrm{fb}^{-1}$ and with cooling plant temperature of $\mathrm{T} = -15\,^{\circ}\mathrm{C}$ [Courtesy of CMS]

and Layer 3 of map **(c)** give a faint hint about this feature. The difference of ΔI within a layer (especially TIB1 and TIB3) is purely defined by temperature; the hotter the module the lower the current increase. Tracker map **(c_2)** shows a 'scale-zoom' into TIB L1 and L3 where the hot regions exhibit a significant lower current (more bluish), normalized to $\mathrm{T} = 0°\mathrm{C}$ – compare with hot regions in map **(a)**, mind also the different scale which, again, represents the different levels of radiation exposure.

Figure 6.51a shows the current evolution versus integrated luminosity for the pixel detector layers; the currents are normalised to $0\,^{\circ}\mathrm{C}$ emphasising the behaviour with radiation and annealing. The predicted linear dependence with radiation fluence is given; layers at lower radii see higher fluence for the same luminosity thus exhibit a steeper slope. The distinctive drops represent maintenance periods, **Y**ear-**E**nd-**T**echnical-**S**tops YETS (a couple of months in time) where the temperature was partially at room temperature. The **H**eavy **I**on HI operation period, about 1 month long, before each YETS features a comparably very low radiation fluence thus no rise in current and already some annealing. The big drop in the middle represents the **l**ong **s**hutdown **1** LS1 when the detector experienced room temperature for more extended periods.[43] The increase in slope after LS1 is due to the decrease in operation temperature from cooling set point $0\,^{\circ}\mathrm{C}$ to $-10\,^{\circ}\mathrm{C}$; the simultaneous annealing during operation is largely reduced at the lower temperature ($\sim 0\,^{\circ}\mathrm{C}$ on sensor). Due to initial lower instantaneous luminosity the equivalent radiation damage (first $5\ \mathrm{fb}^{-1}$) has been accumulated over a longer period thus more time was available for annealing thus the initial slope is less steep.

Figure 6.51b presents the leakage current of a representative single module in the SST versus time, neither normalised to volume nor to temperature. The measurements, versus time, are compared to the model prediction. The model works iteratively, each day the increase with fluence and/or the decrease with annealing is being separately calculated and then summed up. The FLUKA simulation [102, 200] provides a conversion factor between luminosity and radiation fluence for the different locations in space thus for each individual module. For the annealing component the actual sensor temperature is taken into account – measured daily. The periods of

[43] The extracted pixel detector was still kept cold most of the LS1 period.

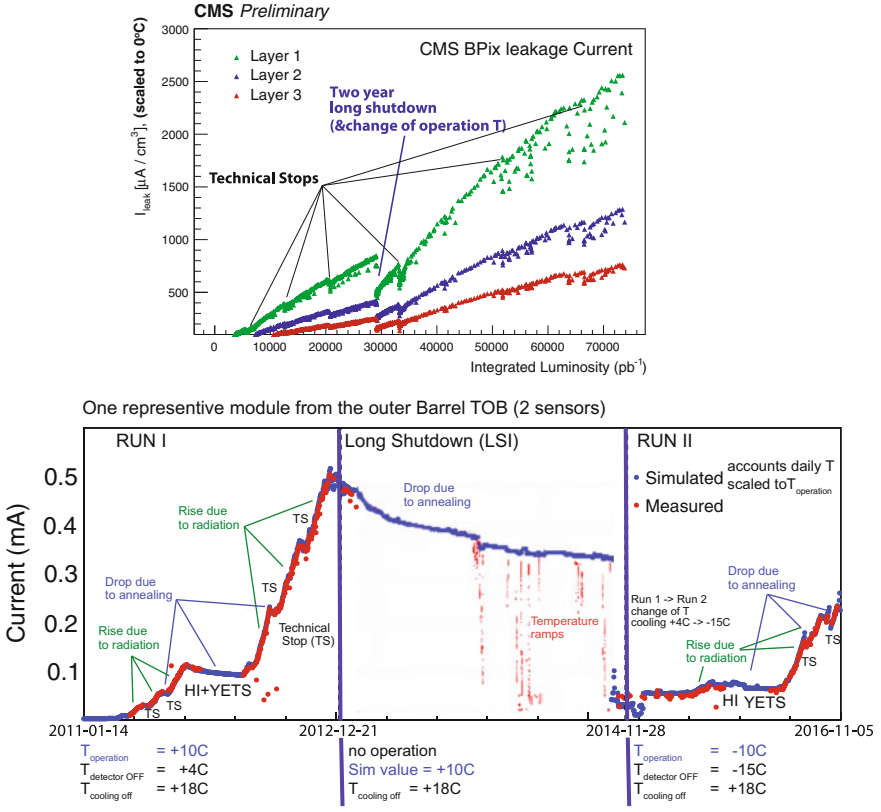


Fig. 6.51 Leakage Current versus luminosity or versus time. (**a**) Shows the mean leakage currents of the three pixel layers normalised to 0 °C exhibiting different slopes corresponding to the fluence at the given layer(radius). The drops are due to enhanced annealing during extended periods at room temperature (Technical Stops TS or YETS). More in the text description. (**b**) Shows the evolution of leakage current of a single SST module with time. It is neither normalised to temperature nor to volume. The *red dots* show the measured current, the *blue dots* the modelled current assuming normal operation temperature. The measurements rely on the full **d**ata **acq**uisition DAQ system thus is not available at all times, e.g. mostly not during LS1. The current drops during maintenance/annealing periods are clearly visible and are modelled correctly as well as increases during the data taking periods where the radiation damage induced current increase dominates but the annealing component is still present. Beginning as of 2015, the operation temperature has been reduced by almost 20 °C reflected in the measured and simulated currents [Courtesy of CMS]

data collection (radiation exposure) and technical stops (annealing only) are clearly visible.

At the beginning of 2014, in the middle of the long shutdown 1 the first temperature ramps had been done. This is visible by the "falling" red dots where the current follows the temperature.[44] From beginning of 2015 onwards the model takes the new temperature into account, see story in Sect. 6.6.1 on p. 257.

[44]For the simulation the temperature is only taken into account 2 times per day thus the simulated currents (blue) do not follow here.

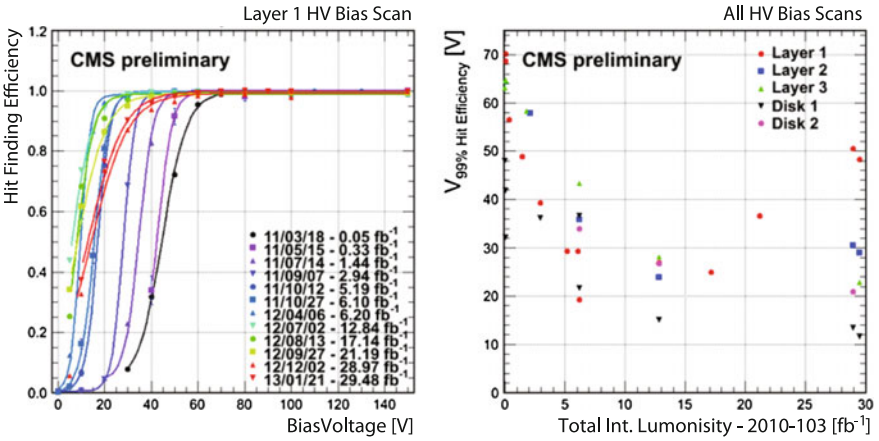


Fig. 6.52 The *left* plot illustrates the Layer 1 CMS pixel HV bias scans, where the start of full efficiency marks the depletion voltage V_{FD}. The *right* figure plots the points of full efficiency for all scans and all pixel layers/disks. This illustrates a clear decrease and later rise after point of SCSI [75]

At regular intervals so-called high voltage HV bias scans are conducted, where the operation (reverse) voltage is being scanned and for different voltages, the corresponding hit finding efficiency is measured for the Pixels and the signal height for the SST. The full detector is normally scanned once or twice per year while a smaller representative part is scanned about monthly[45]. The plateaus in Fig. 6.52 show the point of full efficiency (or full signal) when the detector is fully depleted. Figure 6.52a nicely demonstrates with increasing integrated luminosity the shift of full efficiency to lower voltage values and from a certain value back up again. This point marks the point of **S**pace **C**harge **S**ign **I**nversion SCSI (rf. Sect. 2.1.1) when the initial n-type bulk (n-in-n pixel sensor) changes to p-type. Figure 6.52b shows the points of full hit efficiencies for all Pixel barrel layers and endcap disks. Layer 1 and Layer 2 are already inverted from n- to p-bulk, i.e. overall positive to negative space charge while Layer 3 and all disks are still n-type. 2017, no SST sensor has yet reached the SCSI point but depletion values are decreasing, as expected. The presented situation is a convolution of radiation damage and annealing periods; see Sect. 2.1.

All in all, the Hamburg Model parametrises the situation for n-type Float-Zone sensors extremely well in case all parameters are taken correctly into account. Projections show full functionality for the design integrated fluence with good margin. All initial design parameters prove to be good, e.g. no signal decrease or increase in noise has been observed for the CMS SST (Spring 2017).

In addition to the values described above, also the Lorentz Angle, the laser gain, signal, signal-to-noise, ASIC low voltage currents, thresholds, etc. change.

[45]The CMS data during a full scan is not useful for physics analyses; during a small scan it is.

6.7 Physics with the CMS Tracker and High-Level Trigger

For the nominal luminosity operation, 500 tracks of leptons, heavy quarks, corresponding jets plus jets from additionally created mesons – mostly pions – are expected. A simple event is shown in Fig. 6.53.

Interesting physics signals are many orders of magnitude less frequent than background events! Cross-sections vary over many orders of magnitude where physics processes have to be extracted:

- inelastic: 10^9 Hz
- bb production: 10^6 – 10^7 Hz
- $W \rightarrow l\nu$: 10^2 Hz
- tt production: 10 Hz
- Higgs (100 GeV/c^2): 0.1 Hz
- Higgs (600 GeV/c^2): 0.01 Hz

The Tracker is essential, not only to reconstruct charged particles and measure their momentum, but also for **H**igh **L**evel **T**rigger HLT, lepton identification (μ and e), b- and τ-tagging; also particle flow (e.g. for jet reconstruction) in the full energy range. In combination with the muon detection system and the calorimeters, the CMS Tracker substantially improves momentum and energy resolution. Track isolation cuts can, for example, distinguish between muons coming from jets or from heavy objects. The CMS tracker is also ultimately necessary for all analyses to precisely reconstruct the primary vertex.

The CMS Tracker also plays a role in conjunction with the other sub detectors to identify the particle (particle ID); essential for physics object reconstruction. The calorimeter at higher radius is not able to determine if the energy depositing particle is neutral or has a charge, meaning the electromagnetic calorimeter cannot distinguish between electrons and photons nor can the hadron calorimeter distinguish between neutron and protons. The presence or absence of a pointing track (from a charged particle) arbitrates here. Especially in a dense environment it is important to show that the photon in the calorimeter has no track in the tracker, thus is no electron. As further step the full momentum and energy in all sub detectors are taken into account to form the full and most precise picture – this concept is called **P**article **F**low PF [293, 305]. The PF concept combines the whole information from CMS and afterwards 'forgets' about the detector and use the reconstructed particles as if they were coming from an event generator. The simple story of particle ID with different sub detectors is illustrated in the Appendix in Fig. B.3 on p. 352.

Figure 6.54 illustrates the flexibility of the CMS trigger where data saving on certain energy ranges can be enhanced – trigger selected. The Tracker, dominating the muon momentum resolution at lower energies, participates in selection here during the HLT stage.

Tracker Contributions to Precision Measurements

Several parameter determinations, e.g. electron energy, cannot be attributed to a single CMS sub detector. The following paragraphs introduce the tracker contribution to

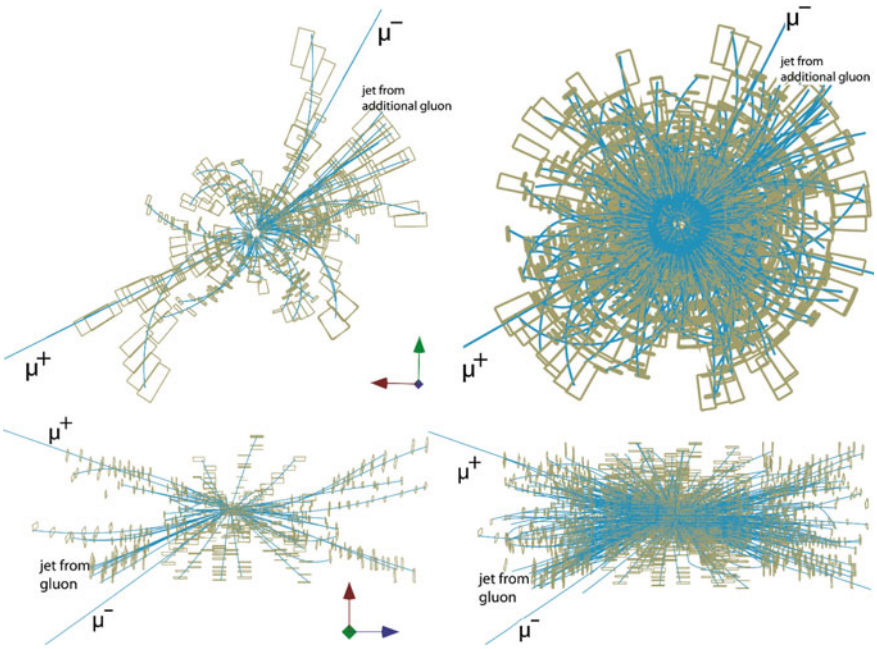


Fig. 6.53 $p\bar{p} \to Z + X; \hookrightarrow Z \to \mu^+\mu^-$ (simulation). The event displayed is a very basic one. As a primary reaction a u quark plus a $\bar{u}$ quark of two protons annihilate to form a Z, subsequently decaying into μ^+ and μ^-. The additional visible tracks are particles (mostly pions) from additional soft processes of gluons and other quarks of the two protons. The *left* picture shows the bare event. The *right* displays the same event with five additional minimum bias events from soft processes of other proton pairs. *Below*, the same event is shown from another angle. At high luminosity operation about 20 minimum bias events are expected. About 500 additional tracks are expected

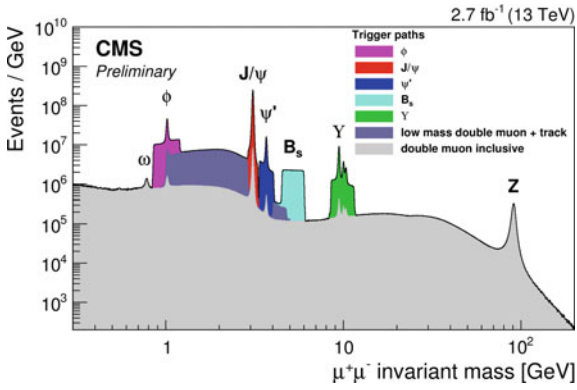


Fig. 6.54 Muon Momentums at CMS and HLT. The *coloured* steps are energy ranges saved and reconstructed with higher priority defined by the trigger menu [296]

muon reconstruction and electron energy determination as examples. A full overview of the CMS detector's physics performance can be found in [298].

Contribution to Muon Reconstruction

A basic strength of the CM(uon)S detector is the reconstruction of muons, where the muon stations together with the tracker and calorimeter are combined in the energy flow concept. In the *global reconstruction* muon trajectories are extrapolated into the tracker volume, where the energy loss in the material plus multiple scattering is taken into account. The tracker measurements improve the transverse momentum resolution at lower energies. Figure 6.55 shows the global muon p_T resolution.

A complementary algorithm to the *global reconstruction* is the *muon identification*. All tracks in the silicon tracker are quantified versus a muon compatibility hypothesis. All available information is taken into account including energy loss in the tracker and associated energy deposited in the calorimeters. This method allows the identification of low p_T muons not reaching the outer muon layers. *Muon identification* adds about 10% to muon detection efficiency with respect to stand-alone *muon reconstruction*. The combination of both reconstruction algorithms enhances statistical samples of events containing multiple muons (e.g. $H \rightarrow ZZ^* \rightarrow \mu^+\mu^-\mu^+\mu^-$).

The tracker together with the calorimeters also significantly contributes to the muon isolation analysis for low p_T muons. Muons originating from b, c, K or π decays are accompanied by particles from the associated jet. Muons from heavy objects like W, Z are more isolated. If the deposited energy in a defined cone around the muon is lower than a certain threshold, the muon is considered isolated – the deposited energy is determined by transverse energy in the calorimeter or sum of transverse momenta of reconstructed particle tracks. The muon isolation (jet/no jet) helps to distinguish between muons of heavy objects or muons produced in jets.

Contribution to Electron and Photon Reconstruction

As for the muons, the tracker improves the ECAL selection of prompt electrons and photons by applying isolation criteria. To distinguish prompt photons and electrons from those produced in a jet, the summed energy deposited in a cone around the identified electron or photon must remain below a defined threshold. For an electron, in addition, the reconstructed energy in the ECAL must match the reconstructed track momentum of the tracker. The isolation algorithm improves background reduction for the channels $H \rightarrow ZZ^* \rightarrow e^+e^-e^+e^-$ and $H \rightarrow \gamma\gamma$, where isolated electrons or photons are expected.

The tracker also improves the precision on the electron energy measurement in a certain range, through the high precision p_T measurement. Figure 6.56 shows the relative energy resolution for electrons measured by ECAL and tracker.

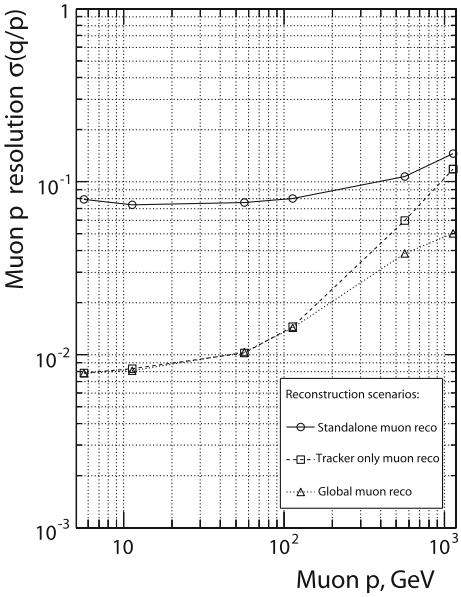


Fig. 6.55 Global muon $p_\perp = p_T$ resolution. Up to energies of 1 TeV, the transverse momentum resolution of a muon is dominated by the tracker, while at higher energies, the multiple scattering term is less significant and the larger lever arm of the muon chambers improves resolution [298]

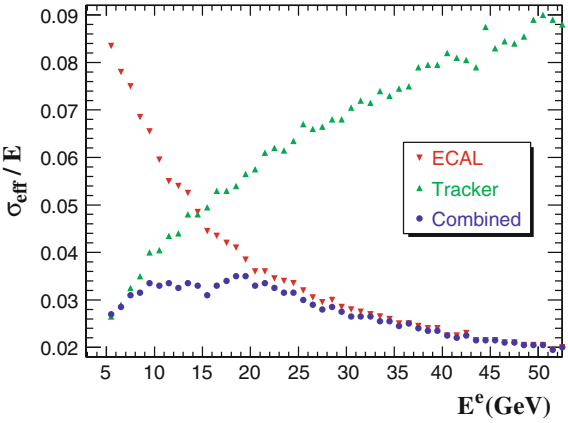


Fig. 6.56 Electron energy measurement. The plot shows the relative electron energy resolution σ_{eff} /E, as measured with the CMS tracker, ECAL and the combined value [298]

Some Examples

With introductory remarks some physics cases with significant CMS tracker impact are listed:

- the key component for the B physics program is the ability of secondary vertex reconstruction also at HLT level with conditional tracking for processes like exclusive B-decays
  - $B_s \rightarrow \mu^-\mu^+$ and $B_s \rightarrow \mu^-\mu^+K^-K^+$
    flavour changing neutral current ($b \rightarrow s$ loop-level processes)
  - $B_s \rightarrow J\Psi\Phi$
    Gold-plated CP violation decay mode sensitive to new physics
- the LHC targeted Higgs discovery in the full mass range up to 1 TeV. Examining the expected decay products, one realizes immediately that lepton-id, b-, τ-tagging and missing energy resolution are crucial. b-tagging is especially crucial to reduce background. The next paragraph with Fig. 6.57 and Fig. 6.58 gives some example event displays of the final discovery at 125 GeV and the tracker role
  - $ttH \rightarrow ttbb$: A very challenging topology where b and lepton tagging is necessary plus a final full reconstruction of e.g. $bbbbqql\nu$
  - $H \rightarrow ZZ \rightarrow l^+l^-l^+l^-(l = e, \mu)$ tracker improves μ resolution significantly with respect to the muon chambers
  - $H \rightarrow \gamma\gamma$ tracker can help identify the vertex or reject π° (jet) but unfortunately for this analysis the high tracker material budget is more a hindrance
  - $H \rightarrow \tau\tau$ a very challenging channel with the τ decaying before reaching a detector element. All decay products need to be reconstructed. It decays in two thirds into hadrons plus a ν_τ.
- **M**inimal **S**uper **S**ymmetric **M**odel MSSM or **SU**per **SY**mmetry SUSY in general. b and τ-tagging play a key role as well as for full reconstruction as also for HLT
  - the neutral MSSM Higgs is predominantly produced via gluon fusion $gg \rightarrow bbH^0$. Besides $AH \rightarrow bb$ with an enormous background significant channels are $AH \rightarrow \tau\tau$, thus τ-tagging is mandatory
  - charged MSSN Higgs production $gg \rightarrow tbH^+$ decays dominantly into tb and $\tau\nu$ where again t, b- and τ-tagging is important
  - SUSY signatures: Squark and gluino production leads to lepton(s) plus missing energy plus jets. Also here the tracker plays a key role

The CMS tracker plays an essential role to address the full range of physics which can be accessed at the LHC. Standard model physics like b physics, t physics, Standard model Higgs searches, MSSM Higgs searches and SUSY searches are planned. At 14 TeV jets are mostly expected but also the standard candles W, Z, tt. The full detector concept and especially its trigger flexibility must cover all eventualities, new physics possibly waits just around the corner.

The Higgs Discovery a Success of the LHC, ATLAS and CMS

At a CERN seminar on 4th July 2012, the ATLAS and CMS experiments at CERN's Large Hadron Collider announced the observation of a new particle in the mass region around 126 GeV consistent with the Higgs boson predicted by the Standard Model. The Higgs boson, as proposed within the Standard Model, is the simplest manifestation of the Brout–Englert–Higgs mechanism. On 8th October 2012, this discovery led to the Nobel prize in Physics for François Englert and Peter Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider." The full story is published in [311], the primary CMS Higgs publication, and in [285] where the first combined ATLAS/CMS result is shown. The two main discovery decay channels are shown exemplarily. The Figures 6.57 and 6.58 show the expected event displays of a simulated Higgs particle compared to a real events of a Higgs candidate followed by the statistics plots derived from simulation compared to the experimentally determined ones. Obviously the simulated ones show a wrong Higgs mass, since it was only measured with the discovery. The lower part of the figures show a 3D event display of a Higgs candidate.

At a first glance one could think that the $H \rightarrow \gamma\gamma$ events are fully reconstructed by the electromagnetic calorimeter or the $H \rightarrow \mu\mu$ channel by the muon system.

The role of the tracker in the $H \rightarrow \mu\mu$ case is an improved momentum and energy resolution. γ identification is only possible with a combination of electromagnetic calorimeter and tracker. Especially in a dense high pile-up environment, a high granular tracker is important to make sure no track points to the γ; meaning the supposed γ is not an electron.

These results and their precision underline the marvellous functionality of the CMS detector and how the different sub-detectors perform together.

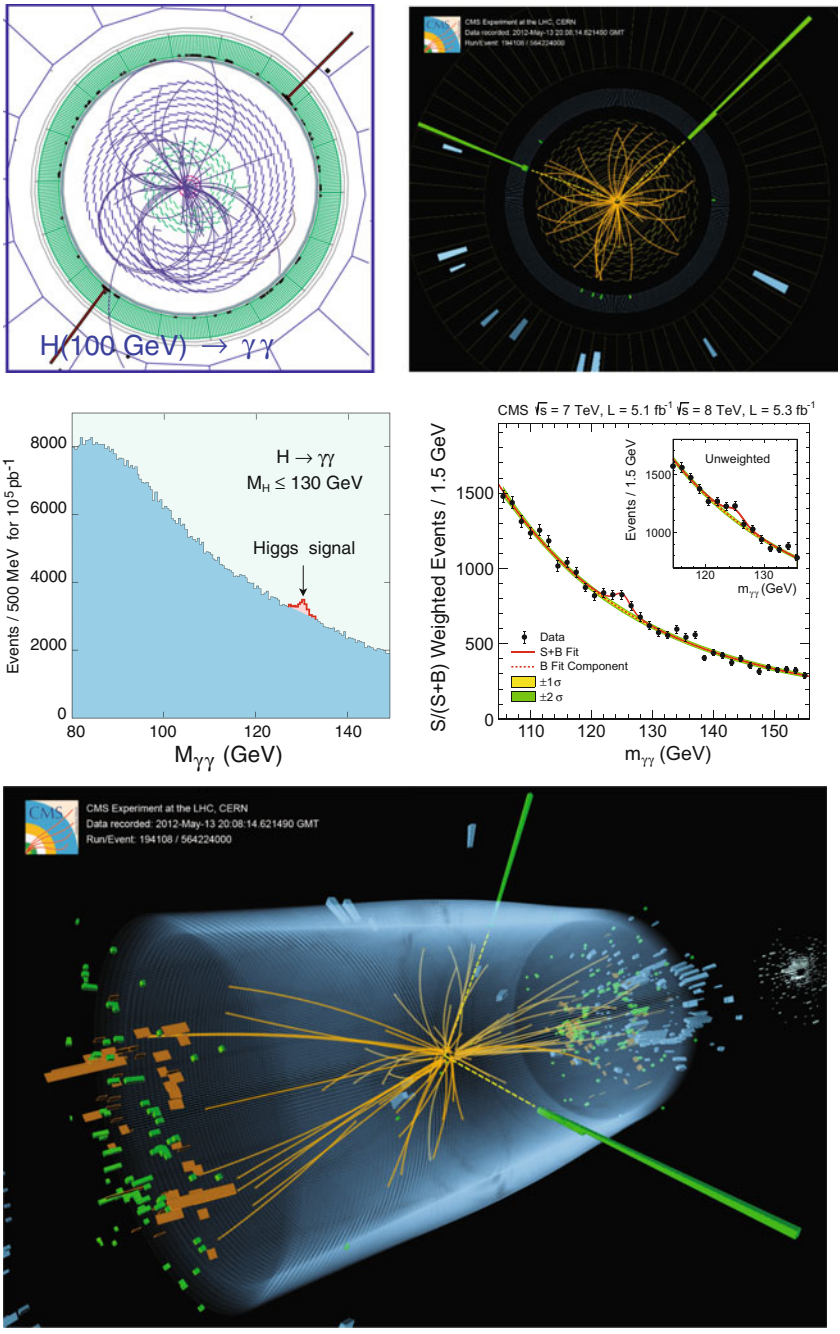


Fig. 6.57 Higgs decaying into $\gamma\gamma$ [Courtesy of CMS/CERN] [291, 311]

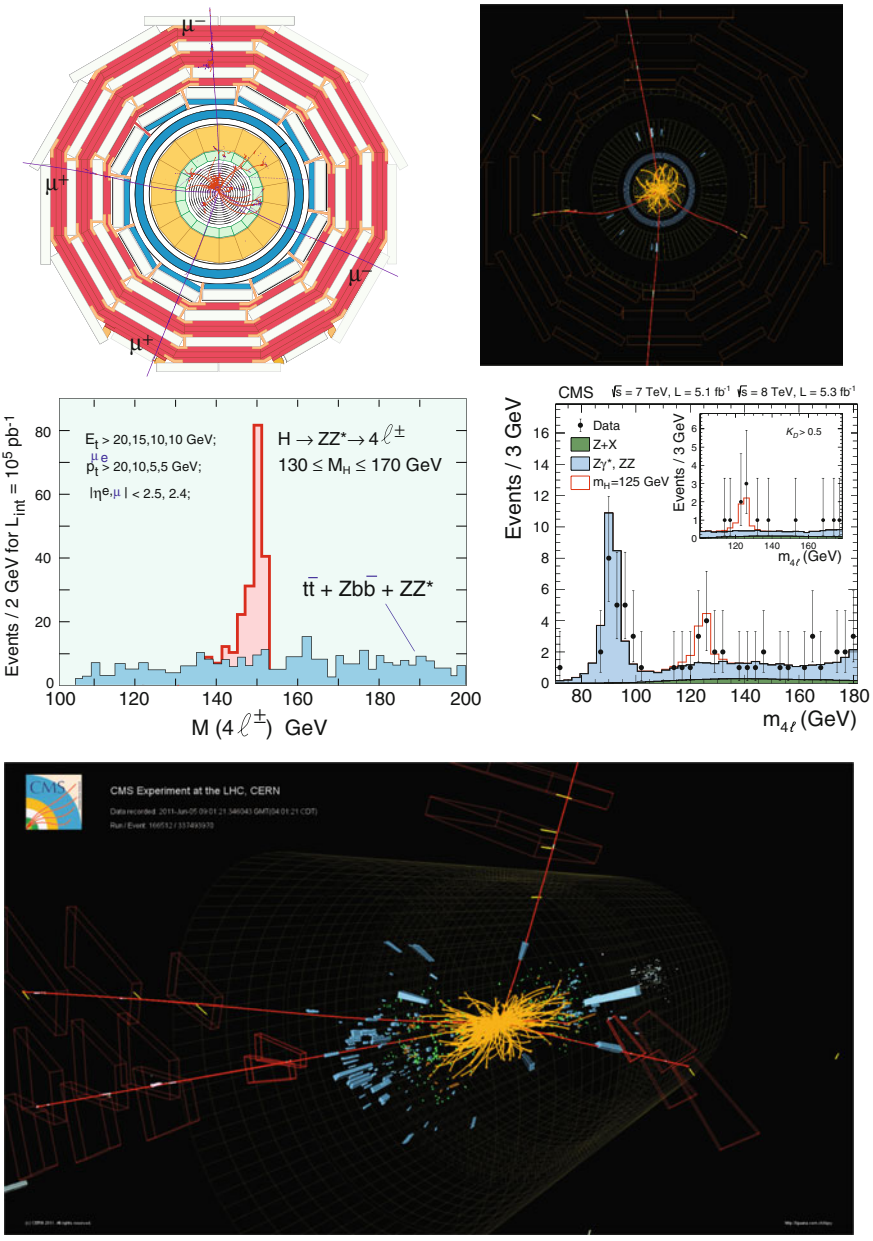


Fig. 6.58 Higgs decaying into $ZZ \to \mu\mu$ [Courtesy of CMS/CERN] [291, 311]

Chapter 7
The Design of the CMS Upgrade Tracker and the CMS High Granularity Forward Calorimeter Equipped with Silicon Sensors for the HL-LHC

The CMS detector has been designed and realized to cope with an integrated luminosity of about $\mathscr{L} = 300\ \text{fb}^{-1}$ with some margin and an instantaneous luminosity of $\mathscr{L} \sim 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. Subsequently, for the inner detectors: Tracker and forward calorimeter this also marks the end of life due to detrimental radiation damage. Both detectors could operate beyond $\mathscr{L} = 500\ \text{fb}^{-1}$ but by far not up to integrated luminosity $\mathscr{L} = 3000\ \text{fb}^{-1}$ – the goal of the **H**igh **L**uminosity LHC HL-LHC operation.

For the next stage – Phase II, planned start 2026 – the HL-LHC, the peak luminosity will be $\sim 0.5 \cdot 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$, 5 – 10 times the design peak luminosity of the LHC, and even higher values are under consideration. This has the following implications for the CMS upgrades: (1) increased radiation and (2) higher **p**ile-**u**p PU[1], therefore increased occupancy. Figure 7.1 presents the expected integrated fluence Φ_{eq} in $n_{1\text{MeV}}/\text{cm}^2$ fluence after an integrated luminosity of $\mathscr{L} = 3\ \text{ab}^{-1}$ at the end of HL-LHC operation. The particle track density scales similarly for different radii.

As seen in Fig. 7.1 the radiation levels at $R = 4$ cm (inner most pixel) will be around $\Phi_{eq} = 2 \cdot 10^{16}\ n_{1\ \text{MeV}}/\text{cm}^2$, decreasing to around $\Phi_{eq} = 10^{15}\ n_{1\ \text{MeV}}/\text{cm}^2$ at $R = 20$ cm (Inner radius of **O**uter **T**racker OT) while at $R = 50$ cm still around $\Phi_{eq} = 3 \cdot 10^{14}\, n_{1\,\text{MeV}}/\text{cm}^2$ has to be endured. The increase in number of events/tracks due to pile-up for the different operation stages of LHC to HL-LHC is illustrated in Fig. 7.2.

This, of course, defines/limits possible technologies versus radii. Average pile-up can be up to PU = 200 events per crossing at the HL-LHC. To achieve the goals defined for the HL-LHC phase, a detector performance at least as good as the one of the current CMS detector is required. Given these requirements CMS opted for an all-silicon tracker plus a tungsten/copper-silicon sampling forward electromagnetic calorimeter, followed by a stainless steel-silicon sampling hadron calorimeter. The **H**igh **G**ranularity **C**alorimeter HGC or HGCAL is based on silicon pad detectors. In the Phase II Tracker and HGC cases, the radiation hardness and technical

[1] Pile-up: number of proton-proton collisions in a single bunch crossing: $\langle PU \rangle = 20$ (LHC design); >50 (LHC reality); 200 (HL-LHC expectation).

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_7

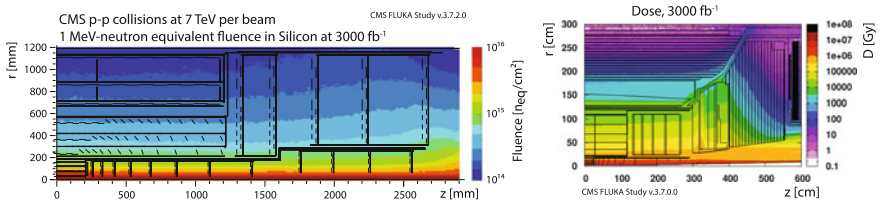


Fig. 7.1 Maps of expected particle fluence corresponding to an integrated luminosity of $\mathscr{L} = 3000\,\text{fb}^{-1}$, expressed in terms of 1 MeV *neutron equivalent fluence* for the Tracker on the *left* or dose for the HGC case on the *right*. *Left* Tracker volume. The expected fluence has a strong dependence on radius, while it is almost independent of the z-coordinate [303]. *Right* High Granularity Calorimeter [308]. Radiation from charged particles increases much faster than $1/r^2$; this behaviour derives from the magnetic field responsible for coiled tracks – this is especially significant at lower radii. At *lower radii*, the radiation is dominated by charged particles while at higher radii neutrons, backscattered contribution from the calorimeters, dominate. The break-even is around $r = 50$ cm. To achieve a reasonable occupancy it can be imagined to instrument 4 cm $< r <$ 20 cm with micro-pixels, 20 cm $< r <$ 50 cm with macro-pixels and $r >$ 50 cm with single strip sensor modules

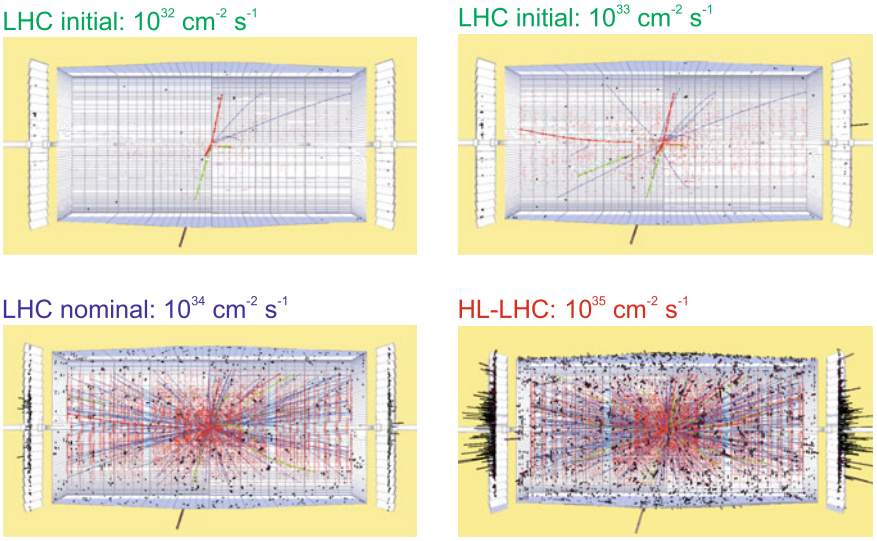


Fig. 7.2 Event pile-up versus luminosity. The current plan for the HL-LHC peak luminosity is $\mathscr{L} \sim 0.5 \cdot 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$, the *last* figure shows about twice this value [346]

structurability of sensors with respect to cell-size is key. These new detectors will largely strengthen the **P**article **F**low PF concept where the particles and their property are being followed/reconstructed along the entire path in all sub-systems – the flow of particles. Initial recipes of ten times more radiation tolerant sensors came from RD50 but extensive R&D campaigns have been conducted within the CMS collaboration as well. Shorter strips and smaller pixels with respect to the current tracker and smaller pads with respect to the current crystal forward calorimeter cells will cope with the much denser particle environment.

In addition, with the ∼10-fold increase of luminosity, a simple increase of existing trigger thresholds (muons and calorimeter) to keep trigger rates below CMS capability is not an option because limits are determined by physics needs and higher thresholds (e.g. on energy) would cut directly into physics performance. A direct implication is the need of a high p_T track trigger at Level-1 for the HL-LHC operation – a complete novelty.

The forward silicon sampling calorimeter concept (HGC) has been inspired by the CALICE [288, 320, 335] idea (**CA**lorimeter for **LI**near **C**ollider **E**xperiment). In addition to strengthening the **p**article **f**low PF concept, the HGC will add timing information for charged and neutral particles.

As of 2017, both designs are in a mature state but changes/improvements are still expected. A snapshot of both, the design of the future Tracker and the HGC, will be described in the next sections. A much more detailed description of the full CMS upgrade plans can be found here [308]. Both detectors will be installed in long shutdown 3 LS3, scheduled for 2024 – 2026.

7.1 The CMS Tracker Upgrade for the HL-LHC – Phase II

2017, after publishing the Technical Proposal TP [308] plus scope document [302] and just before the **T**echnical **D**esign **R**eport TDR [303], the Phase II Tracker design is fairly advanced but details might still change. This section will therefore concentrate on the challenges and the solutions and sketches the earlier evolution to reach the current state. All numbers of e.g. number of modules or layer radius are only good approximations of the future final implementation.

The requirements for the CMS Phase II Tracker at the HL-LHC are:

- 10 times the radiation tolerance of the current detector, namely integrated luminosity $\mathscr{L} = 3000\ \text{fb}^{-1}$ resulting in close to $\Phi_{eq} = 1 \cdot 10^{14}$ to $2 \cdot 10^{16} n_{1\,\text{MeV}}/\text{cm}^2$ in the outer Strip layer and innermost Pixel layers accounting for about 1 GRad of dose.
  - the possibility to replace the innermost pixel layer in a Year-End-Technical-Stop
- increased cell granularity to maintain an occupancy <1% at 140 – 200 PU
- contribution to the Level-1 trigger (L1) – **O**uter **T**racker OT only
  - p_T-modules, which can intrinsically derive if a traversing particle's momentum is above 2 GeV, in the Outer Tracker
- much longer Level-1 trigger latency: 12.8 μs (today 6.4 μs)
- much higher readout rate: 750 kHz – 1 MHz (today 100 kHz)
- extended tracking acceptance: up to about $|\eta| \leq 4$ – concerns mostly the pixel detector (today $|\eta| \leq 2.5$)
- systematic minimization of material budget in the tracking volume

- robust pattern recognition – 'macro-pixels' in Outer Tracker and smaller pixels at low radii
  - improved two-track separation for improved track finding in high energetic jets
- supply much higher power, handle higher data rate and allow for lower temperatures with about the same volume of services (cable channels).

In order to maintain adequate track reconstruction performance at the much higher pile-up levels of the HL-LHC, the granularity of both Outer and Inner Tracker will be increased by roughly a factor four to six. The pitch in the outer strips section will stay roughly the same while the strip length will be reduced (20 to 5 cm and 10 to 2.5 cm and even to long macro-pixels of 1.5 mm length). The inner pixel cell size will be reduced by a factor six from 15000 μm^2 (today – LHC era) to 2500 μm^2. Especially the later reduction will improve two-track separation in high PU environment and impact parameter resolution. Many novel design choices ensure a significant reduction in material budget leading to reduced multiple scattering thus significantly improved p_T-resolution and a lower rate of γ conversion and less Bremsstrahlung[2] for electrons.

With the help of a dedicated tool, "TKlayout"[3] [38, 124, 162], many different layouts have been evaluated and the Tracker layout has been optimized. TKlayout calculates tracking parameters (p_T and impact parameter resolution, number of hits, number of Level-1 stubs, etc.) for offline as well as for L1 tracking. It calculates material budget, radiation lengths and projects leakage current and power. Many of the later design choices and optimizations, e.g. sensor spacing, module locations, usage of square modules in the endcaps, are based on TKlayout modelling and have finally been cross-checked with full simulation.

The perfect Tracker layout would be a sphere, where particles traverse the sensors always perpendicular passing the least amount of material (maybe with a slight angle to hit 2 cells) and would be fully hermetic. The future CMS Tracker is approaching this concept as much as possible in an overall cylindrical envelope. The ingredients, like 2S and PS[4] modules will be introduced in the next section. The full optimization is being done by several people and lots of input from all members of CMS during several years. Figure 7.3 demonstrates the design evolution starting from the current CMS Silicon Strip Tracker design – a small subset of design steps are being described. Outer Barrel configurations from 4 to 16 layers have been studied. Early on studies demonstrated that rectangular instead of wedge-shaped modules can be used in the forward direction with no significant penalty in mass and resolution but large reduction in system complexity (number of module types). An early proposed design featured a barrel-only layout, but as mentioned earlier a sphere would be the perfect design, subsequently the usual barrel-endcap design has been adopted. In a

[2] Often simply called Brems.

[3] 3D tool adopting a non-simulative, parametrization-driven approach to tracker performance evaluation, taking e.g. multiple scattering as measurement error into account. It has been wetted against full simulation but it always provides the best case performance being independent of specific tracking algorithms. TKLayout runs more than 100 times faster than full simulation.

[4] 2S = **S**trips+**S**trips sensors; PS = **P**ixel+**S**trips sensors.

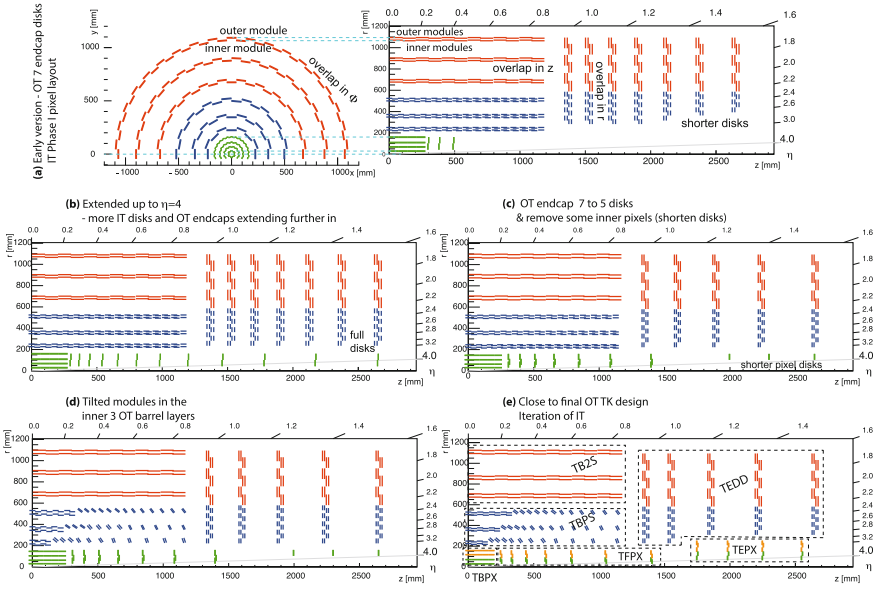


Fig. 7.3 The figures show the evolution of the Phase 2 Tracker layout. Figure **a** shows an early design in $r\Phi$ and rz view where also the module overlaps of inner and outer modules, inside one layer, are discernible. 2S modules are represented by red lines and PS modules by blue ones (each line represents one module) while micro-pixels layers are shown in *green* and *orange*. Initially 3+3 barrel layers and 7 endcaps plus the phase I pixel configuration – 4 barrel plus 3 endcaps have been evaluated. Also 4(2S)+2(PS) (without explicit picture) have been evaluated earlier. As a next step (figure **b**) 9 pixel full-size endcap disks were added (totalling 12 disks) to cover ranges up to $|\eta| = 4$ to have a first glimpse about performance. Consequently also the last three OT endcaps where extended to lower *radius*. In a later step (figure **c**) two OT endcap disks were removed without penalty in performance, as well as some pixel disks were shrunk to adapt to reality, namely a conical beam-pipe otherwise clashing with the detector. As further major step (figure **d**), and after many studies of mechanical feasibility, the larger part of the PS modules have been tilted. This saves a lot of modules thus material and monetary budget. The tilt also improves trigger performance at larger z – more in Sect. 7.1. Figure **e** shows the almost final layout where the IT tracker has been rationalized to allow insertion with beam pipe present – more in Sect. 7.1. The final TDR layout is presented in detail in Fig. 7.4

next step several additional pixel disks have been introduced allowing instrumentation up to $|\eta| = 4$. Several optimization steps later the number of outer endcap disks were reduced from seven to five while it became clear that the OT barrel section needs 3 pixelated plus three strip layers (initially only two PS but four 2S layers had been evaluated). A large TEDD endcap disk has 15 rings (six 2S plus nine PS). The PS module concept, with a very good intrinsic z-resolution due to the short 1.5 mm macro-pixels, also avoids any necessity to have stereo modules, as present in the current tracker. As a side remark, with the p_T module concept, stereo modules are anyhow not applicable since strips/macro-pixels need to be parallel. The sphere concept came back to mind and the modules in the inner three layers have been tilted within mechanical possibilities by 40 to 74°. This concept saves about 1300 PS

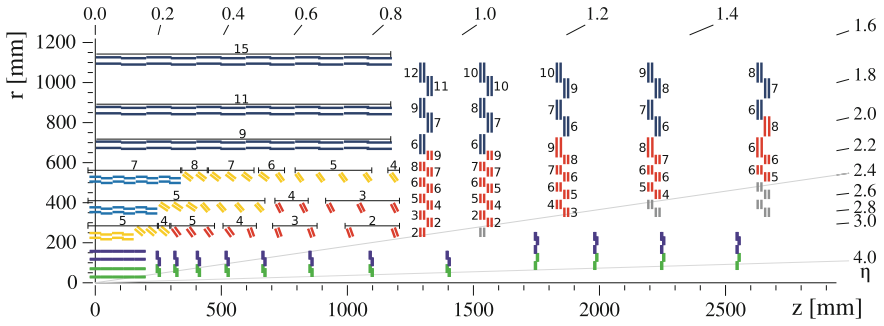


Fig. 7.4 The CMS Tracker TDR layout with emphasis on sensor spacing and track trigger 'acceptance windows'; a quadrant is shown. All OT modules are p_T modules with 2S at $r > 60$ cm and PS modules at $20 < r < 60$ cm. The sensor spacing is colour coded with *dark blue* for 1.8 mm, *light blue* for 1.6, *yellow* for 2.6 and *red* for 4 mm spacing; *grey modules* ($|\eta| > 2.5$) do not participate in the trigger. The numbers represent the 'number of strips in the acceptance windows', these are tunable during running (values here only exemplary or considered as starting point). Comparing the TEDD shown here with Fig. 7.3**e**, some optimization can be spotted resulting in better PS coverage and about 260 modules savings – short and long disks at different z now follow slightly different designs to allow for more space of the Inner Tracker IT forward pixels (*TEPX*). TEDD double-disk 1 moved closer to the barrel – optimise spacing for hermeticity. The IT micro-pixels consist of 4 barrel layers, 8 forward plus 4 extended forward disks covering up to $|\eta| > 4$. The *light green* pixel layers/disks consist of 1×2 chips modules while the violet ones consist of 2×2 chip modules. The tracker, all in all, has seven pixel layers – 4 micro- and 3 macro-pixels [303]

modules thus reduces material budget significantly[5] and a good amount of money. It will be later demonstrated that the change also has a very significant and positive impact on the Level-1 track trigger. There were more subtle changes to optimise the tracking and hermeticity, e.g. moving the first (at low z) TEDD double-disc closer to the barrel section or changing which modules of the TEDD or TB2S are located closer in and which further out (orientation) or overlaps of the modules. Figure 7.4 presents the layout at the time of the Technical Design Report TDR, still with some potential for future tuning. The necessity of having tracking information at the Level-1 Trigger also influenced the design strongly; more details later.

Outer Tracker – Strips and Macro-Pixels

The requirement to contribute to the trigger at Level-1 has a dramatic impact on the future CMS Tracker design, strongly defining the module concept and partially the full layout of tracker. The L1 trigger functionality was initially considered 'THE' challenge, but in the end allowed a very elegant, advantageous and powerful tracker design.

Dedicated p_T-modules in the **O**uter **T**racker OT have been developed, being able to discriminate intrinsically between high ($p_T > 2$ GeV) and low momentum traversing particles. Only the information from high p_T tracks, largely reduced with respect to the full data, is being transmitted for each crossing (40 MHz) and being used for the L1-Trigger – only from the Outer Tracker. The later offline tracking

[5] In the order of 0.2 x/X_0 in the range $1 < |\eta| < 2.5$.

exploits **I**nner **T**racker IT (Pixels) for track seeding and then utilizes mainly the large lever arm (large radius) of the Outer Tracker for momentum resolution; here the number of layers in the OT is less important, as long as it is above the critical number plus one for redundancy. The minimum number of OT layers is therefore defined by the L1 functionality but, on the other hand, the extra novel pixelation in the first three OT layers makes a fifth pixel layer unnecessary.

Being the basic ingredient, the modules and their p_T threshold capability will be introduced first. Every module in the Outer tracker consists of two narrowly spaced sensors with parallel strips(pixel) where the front-end chips receive signals from both sensors thus being able to correlate signals. High p_T tracks, bent in the strong magnetic field of CMS, are straighter and therefore pass in a narrower band of strips (macro-pixels) of the upper sensor – in the 'selection window', while lower momentum particles, bent stronger, miss the window. A pair of hits in the two sensors from one particle passing the correlation criteria is called a 'stub'. Figure 7.5 illustrates the concept. The 3.8 T field allows a relatively close spacing with 1.6, 1.8 and 2.6 mm distance and 4 mm. The centre and width of the 'selection window' is programmable and therefore adaptable/tunable after installation, for different locations in the tracker, adjusting for e.g. different z-position thus longer traversing path length or higher radius r. The 'stub' information is then transmitted for all bunch crossings at 40 MHz; about 100,000 stubs per bunch crossing corresponding to about 3% of all hits. The high magnetic field at 3.8 T of CMS, and consequent strong bending power, is a real advantage allowing p_T discrimination with narrow sensor spacing.

The module concepts are displayed in Fig. 7.6. The two module types are **S**trip-**S**trip 2S modules assembled from two strip sensors and **P**ixel-**S**trip PS modules

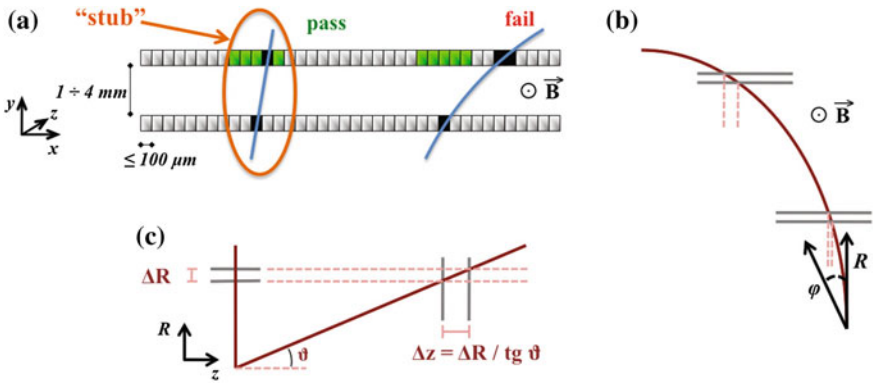


Fig. 7.5 **(a)** Correlation of signals in closely-spaced sensors enables rejection of low-p_T particles; the channels shown in *light green* represent the 'selection window' to define an accepted 'stub'. **(b)** The same transverse momentum corresponds to a larger displacement between the two signals at large radii for a given sensor spacing. **(c)** For the endcap disks, a larger spacing between the sensors is needed to achieve the same discriminating power as in the barrel at the same radius. The 'selection window' can be tuned along with the sensor spacing to achieve the desired p_T filtering in different regions of the detector [308]

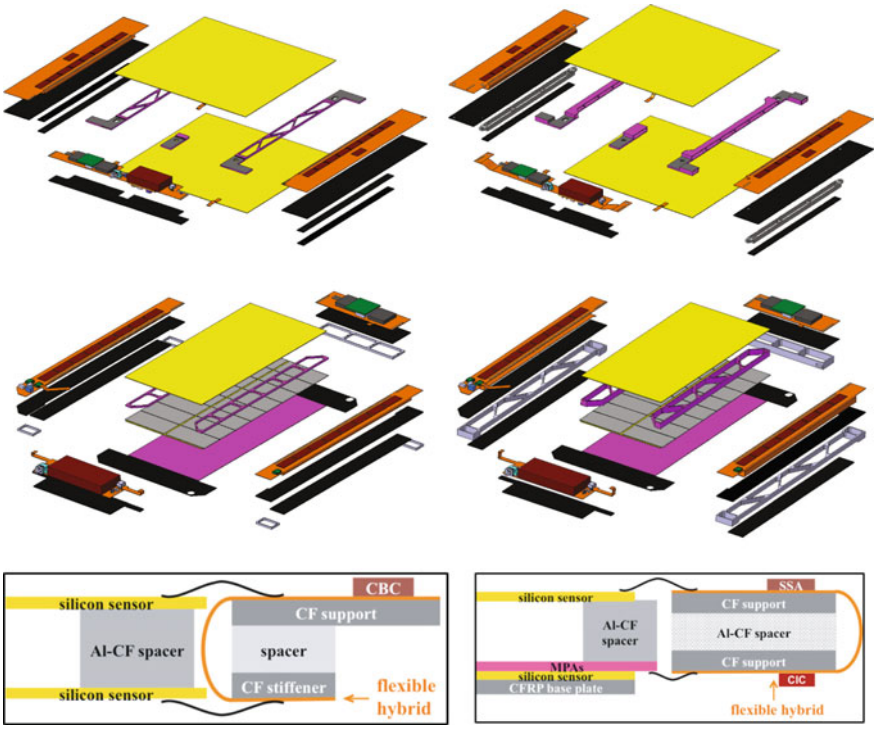


Fig. 7.6 The figures show the component view of modules with 1.8 (1.6) and 4 mm sensor spacings; the *upper* ones 2S and the *lower* ones PS modules (the 2.6 mm PS is not shown) [303]. The *yellow* rectangles are the strip sensors, the *violet* the pixel sensors with their *grey* MPA ASICS on *top*. The spacers are of Al-CF (**A**luminium **C**arbon **F**iber) a new material for stability, machinable and very good isotropic heat transfer – mind the different designs to allow different spacings with lowest material budget (alternatives are still under investigation). The CF (*black*) is for mechanical stability and thermal management, especially necessary as base plate below the pixel-ASIC sandwich. The readout hybrids are located on the *left* and *right* with eight CBC ASICs each plus a **c**oncentrator **c**hip (CIC); the whole ASICS-flex, 'CF – Al-CF-spacer – CF' sandwich including the flex bending will be fully delivered from industry. The service hybrid is situated on the *lower end* for the 2S module and split in two for the PS module. Realized in the technology as the readout hybrids, it contains DC-DC converter for powering and the optical readout. The *lower* figure shows a cut-view illustrating how the flexes are bent, the signals are routed and where the ASICs are located (*left* 2S module and *right* PS module). One can see where and how these modules are placed in Fig. 7.7

assembled of a pixelated (macro-pixel) and a strip sensor allowing for a decent z-resolution in the barrel. The 2S modules come with 1.8 and 4.0 mm sensor spacings and are mounted at $r > 60$ cm. They feature 2 times 8 chips with 254 channels each connected to 2 sensors with 2 times 1016 strips of length 5 cm and pitch 90 μm. The even chip channels are connected to the upper and the odd channels to the lower sensor allowing the above described in-chip correlation logic. The signals are routed through the flex hybrid being bent around the hybrid body. The total power of this

module type after 3000 fb^{-1} will be around 5 W. The readout chip is called **C**MS **B**inary **C**hip CBC, a subsequent continued development from the current APV-ASIC.

The PS modules are located at 20 cm$< r <$ 60 cm and come with sensor spacings of 1.6, 2.6 and 4 mm. The strip sensor features 2 $\times$ 960 strips with 2.5 cm length and a pitch of 100 μm connected via wire-bonding at the end to the **S**hort **S**trip **A**SIC SSA[6]. 16 **M**acro-**P**ixel-**A**SICs MPA with 16 $\times$ 120 cells are bump bonded to the macro-pixel sensor with 2 times 8 times 16 $\times$ 120 (30720) pixels of length $\sim$1.5 mm and a matching pitch of 100 μm. Processed signals from the SSA are routed via the bent flex hybrid to the MPA where the p_T correlation (stub finding) is handled. Being a 'pixel' module the power dissipation is higher but still remains at 8 W. The pixelation of the PS modules allows z-vertex resolution of the order of $\sim$1 mm at L1 in the central region. Due to the higher power dissipation, cooling pipes run along a base plate under the pixel sensor while the 2S modules have simply 5 individual cooling contacts.

In the end, 2S modules consist of two sensors, two readout hybrids plus one service hybrid plus 3 spacers and is planned to be mounted manually with dedicated jigs, different to the current tracker where modules have been assembled by robotic gantries. Similar for PS modules where industry delivers two spacers, four hybrids, one sensor and one bump bonded sensor-ASICs sandwich.

Sensor technology will be discussed in Sect. 7.1.1.

It is worth to note for future discussion that the module is a complete electrical entity with all necessary components; there are NO additional boards with additional data treatment in the tracker; all cables and optical fibres are directly routed outwards to power and readout. This reduces the complexity and material budget significantly. It also facilitates testing during assembly. Also data-out (trigger and full readout after L1-accept) and control-in signals share the same fibre pair – one fibre IN one fibre OUT per individual module. In the 2S modules binary strip data is fully synchronously readout to the **C**oncentrator **A**SIC (CIC) which sorts, sparsifies, formats and buffers the data before optical transmission. The **S**hort **S**trip **A**SIC SSA transmits the strip data to the **M**acro-**P**ixel-**A**SIC MPA responsible for the p_T selection L1-correlation. Trigger & full readout data, after a L1-accept signal, is then transmitted to the CIC to be sorted and shipped to the optical transmitters. All channels store events individually up to a latency of 12.8 μs, waiting for a trigger Level-1 accept and then transmit sparsified data with up to 750 kHz L1 rate.

The modules design was originally driven by the trigger requirements but has also very significant other benefits:

- it is an effective way of collecting two space-points Two sensors share mechanics, cooling and electronics resulting in significantly reduced mass.
- each layer (module) acts directly as track seed with **vector** (4D = 3D position plus bend) information instead of **space points** (3D).

[6]Interesting fact: The 'binary' SSA chip features two comparator thresholds, namely 0.4 and 1.5 MIPs, enabling to distinguish MIPs and highly ionising particles, a possible signature for a hypothetical long lived stable particle.

- in addition to the p_T discrimination at Level-1; the stub as well as full track information from Level-1 can enter the High Level Trigger stations. This should even speed-up later offline reconstruction.
- the self-contained design is very appealing. Only one data link connection, one power and one high voltage line. No other auxiliary electronics in the tracker volume, simplifying validation and construction.
- PS macro-pixel modules provide unambiguous 3D information facilitating pattern recognition in a high PU environment. This allows 3D track seeding.
- PS modules are more cost-effective than micro-pixel hybrid detectors due the larger cell size thus larger pitchs allowing for more industrial cheaper bump bonding.

The mechanics for the outermost three OT barrel layers, consisting of 2S modules and called TB2S (**T**racker **B**arrel **2S**), will be very similar to the current tracker implementation (see previous chapter). In contrast, Fig. 7.7 depicts the novel mechanical structure of the inner three OT layers populated with PS modules. The inner (at small z) straight sections are built of so-called planks holding the PS modules. The PS modules in the tilted section (at higher z) are mounted on rings also holding the cooling loops. This part is called TBPS volume **T**racker **B**arrel **PS**.

The innovative decision to populate the endcaps with rectangular (square for 2S and rectangular for PS) instead of wedge shape sensor makes the current petal concept obsolete and bigger structures like full or half-disks are more adequate. To

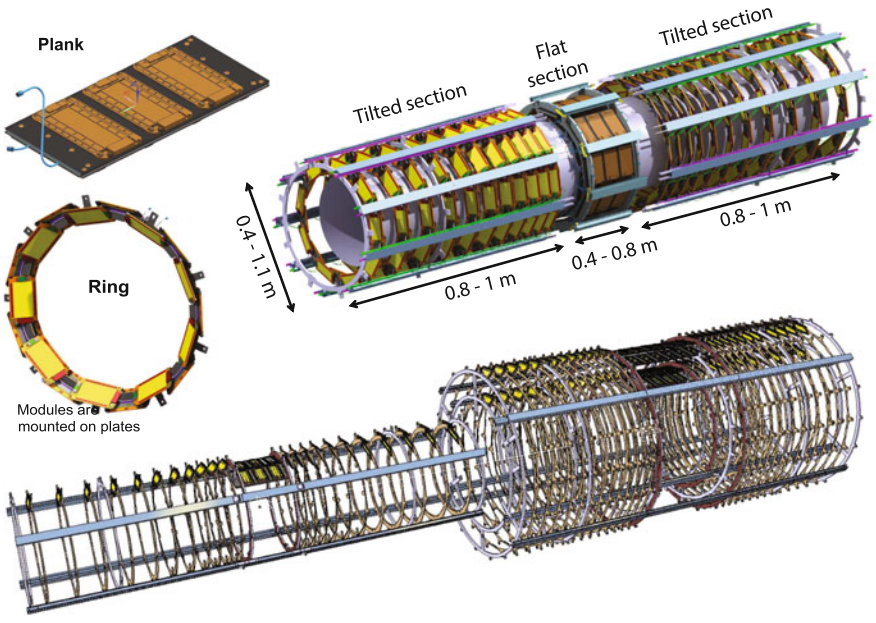


Fig. 7.7 The three TBPS layers are populated with PS modules on 'planks' for the inner flat section and on rings for higher z positions to achieve the tilt shown earlier. The *lower* part also illustrates how the final pieces will be put together [303]

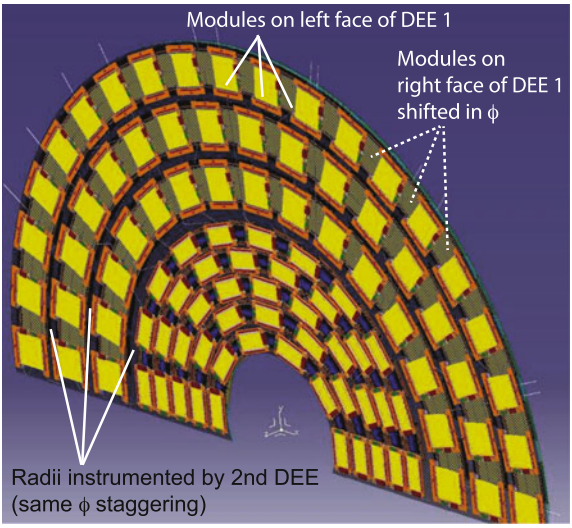


Fig. 7.8 The figures shows a single dee, to achieve full hermeticity without modules clashing, the other side of the dee needs to be populated with modules shifted in ϕ and the second dee needs to equip the locations in-between radii of the other dee [308]

facilitate handling half-disks called 'DEEs' are chosen and to guarantee hermeticity dees are assembled to double-Disks (four dees form one double-disk). Consequently the endcap volume is called TEDD **T**racker **E**ndcap **D**ouble **D**isk. One dee populated with modules is shown in Fig. 7.8. The dees support both 2S and PS modules and integrate the cooling loops. Even without additional electronics components the system is quite complex and routing of cables and fibres is a challenge.

Inner Tracker IT—Micro-Pixels

The system will consist out of 4 barrel layers (r = 3 to 16 cm) plus 8 regular sized end-disks plus 4 extended end-disks and will cover up to $|\eta| = 4$. The Outer/Inner Tracker boundary is located at r = 20 cm (r = 30 cm) $|z| < 160$ cm ($|z| > 160$ cm). The CMS central beam pipe deviates from cylindrical to conical shape at larger $|z|$. A full flat cylindrical pixel mechanics could not pass this 'obstacle' with straight insertion and therefore needs to be inserted from an angle. To achieve this the pixel detector is radially displaced further out at $|z| = 160$ cm. This is to allow pixel insertion/removal with beam pipe removal. This comes natural the current tracker, spanning only $|\eta| < 2.5$, and where the OT endcaps disks at higher z are naturally shorter. For an $|\eta| < 4$ coverage system, this volume has to be instrumented with pixel disks but at higher radius for higher z (see Fig. 7.9). CMS will keep the possibility to remove/repair the pixel detector or to replace the inner layer in a standard **Y**ear-**E**nd-**T**echnical-**S**top YETS. The extended disks are also perfectly suitable to accurately measure the luminosity (with background reduction taking tracks into account instead of simple hit counting).

It is fair to say that the layout, today 2017, is not final, e.g. the numbers of disks might change and even tilted designs are still under discussion. Even the pixel cell size might increase slightly depending on the ASICs development, its radiation hardness and number of features – more details on the RD53 ROC in Sect. 1.11.

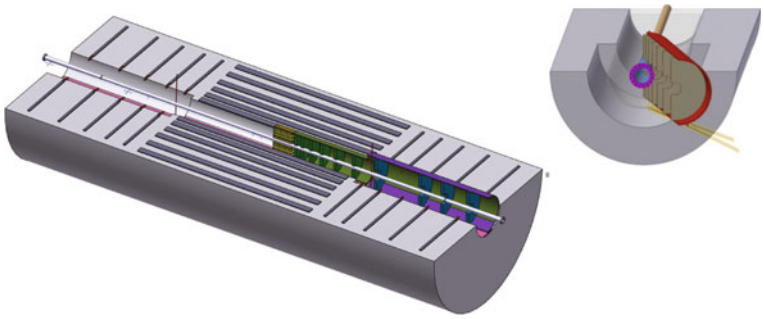


Fig. 7.9 CMS Pixel Phase II TDR layout and insertion studies

With respect to the current detector, the pixel cell size will be reduced by a factor six from 15000 to 2500 μm^2, while the total coverage will be much increased. The readout chip (ROC), to be bump bonded, will have cell sizes of $50 \times 50\,\mu m^2$ allowing easy mapping to pixel sensor cells of $50 \times 50\,\mu m^2$ or $25 \times 100\,\mu m^2$ for different locations, e.g. barrel and forward. Sensor technology will be discussed in Sect. 7.1.1. Also multiples like $100 \times 100\,\mu m^2$ or $50 \times 200\,\mu m^2$ are feasible if readout pattern in the chips can be appropriately configured to save power. The ASIC is a common ATLAS/CMS development driven by the RD53 collaboration (rf. also Sect. 1.11). Bump bonding of these small pitches will be a challenge, but seems doable and affordable.

Square ($50 \times 50\,\mu m^2$) pixels are better for z-resolution in the central region but would require lower detection threshold at higher z in the barrel region since, with the inclined track, charges are spread over many pixels; and this might get worse with radiation damage. Rectangular ($25 \times 100\,\mu m^2$) pixels are better in ϕ-resolution due to the smaller pitch (25 versus 50 μm), and clearly momentum resolution will benefit from this. The magnetic field and tracks are more or less aligned for disks at very high z (high η), thus the effective **B** is low thus the momentum resolution is pretty bad. The instrumentation at high η (forward region – disks) serves more the purpose of particle flow to reconstruct the full event and especially link all tracks to the individual vertex, especially in a high pile-up environment – also called pile-up mitigation. This might favour $50 \times 50\,\mu m^2$ pixels in the forward direction to emphasize z- over ϕ-resolution here.

As for the OT and for the current CMS Phase I pixel detector only rectangular modules are foreseen limiting the number of module types; here limited to two types with 2 or 4 ASICs.

Due to the high radiation, the very limited space and the very stringent requirement of low mass, services are moved out to higher radius.[7] Due to limited radiation tolerance, the optical links are located at higher radius and signals have to be driven

[7]Unfortunately, the trick, used in Phase I to move them out of the tracking volume does not work any more with tracking up to $|\eta| = 4$ – the whole volume is dedicated to tracking.

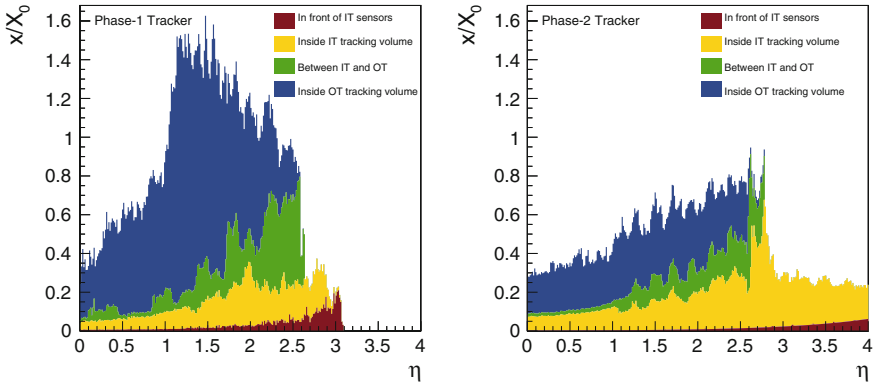


Fig. 7.10 Material budget distribution versus η for the current and future Tracker (TDR geometry). As an example, the move from flat to tilted geometry saves in the order of 0.2 x/X_0 in the range $1 < |\eta| < 2.5$ [303]

electrically over 1 to 2 m. The RD53 ROC will be the only active element on the modules! Direct powering and even DC-DC converters would result in many kilograms of cables thus serial-powering is being adopted.

CMS expects about 1 GRad of integrated ionising radiation and a hit rate of up to 3 GHz/cm^2 ($\sim$100 kHz/pixel) in the innermost layer. The chip is under development within the RD53 collaboration (joint venture of ATLAS and CMS). The chip will cope with a L1-accept data rate of about 750 kHz and features a trigger latency of 12.8 μs. Different to the Outer Tracker, no pixel information is available at Level-1. The high rate does not allow to ship data to the chip periphery for buffering but several cells store events locally. Instead of having a large analogue (pre-amplifier) section with digital logic 'just at the periphery', the new chips will have analogue 'islands' in a 'sea' of digital logic. Chip size will be 22 mm $\times$ 16.4 mm active area plus 22 mm $\times$ 2 mm periphery. Depending on locations modules will consist of 1 $\times$ 2 (rectangular-inner radii) or 2 $\times$ 2 (square – outer radius) chips with up to 3 low-mass data links per chip with about 1.3 Gbits/s (see also Fig. 7.4).

CMS Tracker Phase II System

A very important aspect is the reduction of mass inside the Tracker volume and in front of the electromagnetic calorimeter. The tracker material is a limiting factor for ECAL performance (energy corrections), increase Bremsstrahlung and photon conversion and the main source of track reconstruction inefficiency (nuclear interactions).

Figure 7.10 demonstrates the substantial reduction with respect to today.[8]

[8] Both plots do not account for the material of services, e.g. connectors or manifolds outside the TK volume.

The reasons for the massive reduction are manifold, the main saving lies in the reduced cable cross-section possible due to the DC-DC conversion at module level:

- less layers
- DC-DC converters will reduce the amount of copper (diameter of the power cables) drastically
- CO_2 cooling comes with thinner/smaller pipes and the liquid CO_2 per se is also lighter than C_6F_{14}
- reduce silicon thickness if affordable (although the Fig. 7.10 shows the standard sensor thickness of 300 μm)
- less additional boards and connectors
- avoid meander cabling
- optimise module design – save at every 'corner'
  - no extra glass pitch adapter – integrated in flex hybrid, pushing the hybrid technology to its limit – down to 42 μm track width and 42 μm vias
  - two sensors in a single frame
  - balance thermal management – the margin is in the cooling plant capacity, not in massive thermal contacts
- The whole system is optimised/balanced for minimal power consumption. This sets limits on some parameters, e.g. longer latency (more buffers), faster readout, more channels would increase power thus higher current (more copper) and temperature (cooling infrastructure).
- serial powering in the Inner Tracker system.

Table 7.1 compares the current CMS Phase I and the future Phase II Tracker

Track Trigger Technique

The Track Trigger at Level-1 is based on the capability of the intrinsic low p_T discrimination power of the p_T-modules in the Outer Tracker and their capability to sent out 'stub' data for every event at 40 MHz. The concept works down to 20 – 25 cm in radius with a realistic 100 μm strip/pixel pitch and reasonable sensor to sensor spacing of 1.6 – 4 mm, thanks to the strong CMS magnetic field of 3.8 T. Around 3% of all tracks are above 2 GeV accounting for roughly 100,000 stubs at PU = 140. These 3% of tracks at 40 MHz use about 80% of the readout bandwidth leaving about 20% for the full data with a L1-accept rate of 750 kHz. The data path is illustrated in Fig. 7.11 where the 512 deep buffers guarantee a latency of 12.8 μs before event info is overwritten. The stub identification is handled by the CBC and MPA ASICs. It is a pure self-seeding detector-intrinsic measurement at a true 40 MHz level not depending on any pre-trigger info from the calorimeters or the muon detectors at L1.

The Track Trigger relies on six double-sensor layers giving six stubs (5 + 1 redundancy) with at least two (plus one for redundancy) measurement points with good z resolution from the PS modules (2 + 1 for redundancy). Hermeticity is a key requirement.

Table 7.1 The table compares the current and future CMS Tracker (TDR version, numbers to be fine-tuned in a future engineering design review). According to TDR plans the Phase II Tracker active surface will be slightly smaller than the first CMS Tracker - larger IT but smaller OT

	CMS Tracker (Phase I)	CMS Phase II Tracker
Outer tracker (OT) – Strips and Macro-pixels		
Silicon Surface	∼206 m^2	∼191 m^2
Strips	9.3M	42M
Macro-Pixels	0	172.5M
Modules	15148	∼13300
Readout rate	100 kHz	750 kHz (on L1 accept) / 40 MHz (@L1)
Cooling	mono-phase C_6F_{14}	bi-phase CO_2
Cooling pipes	Al, stainless steal, Titanium	Titanium
Powering	direct	DC-DC
Inner Tracker (IT) – Micro-Pixels		
Pixel surface	∼1.75 m^2	∼4.9 m^2
Pixels	∼127M	∼2000M
Modules	1632	∼4300
Readout rate	100 kHz	750 kHz
Cooling	bi-phase CO_2	bi-phase CO_2
Cooling pipes	stainless steal	Titanium
Powering	DC-DC	serial powering

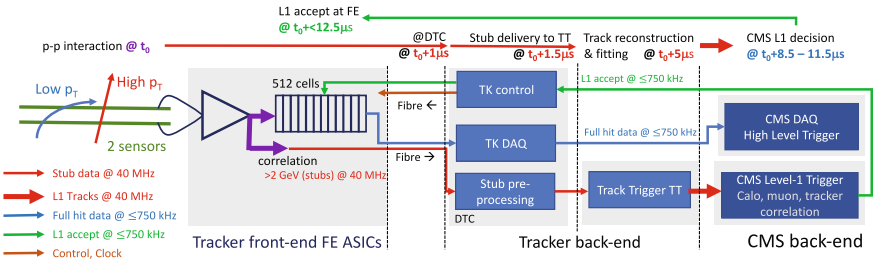


Fig. 7.11 The stub data (*red path*), identified in the front-end ASICs is sent with full 40 MHz to the back-end, where tracks are pre-processed and in dedicated boards identified, fitted and sent further to the CMS central trigger. The main data handling is done by the **D**ata, **T**rigger and **C**ontrol DTC boards. Evaluating the full information of Level-1-tracks, Calorimeter information and Muons in dedicated FPGAs farms, a L1-accept is being sent back to the front-ends where all data is buffered for the latency of 12.8 µs. The *upper* figure text 'repeats' the Track Trigger data flow with timing information. After the L1-accept, the full, although sparsified, data is being sent to the Tracker/CMS DAQ (data acquisition) with a rate of ≤750 kHz

As can be seen in Fig. 7.6, readout ASICS are situated on both ends of the sensors of PS and 2S modules increasing granularity/halving strips lengths. Due to this fact p_T correlation only works for the two halves of the modules individually with no

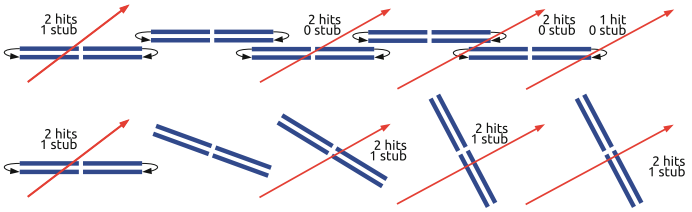


Fig. 7.12 The *upper* cartoon shows the flat geometry where tracks at L1 are lost when passing in the *middle* and at an angle where the *upper* sensor is being 'hit on the other end'. Readout and thus p_T correlation processing is only individually per end possible. The lower cartoon demonstrates how the tilt recovers the stub finding [301]

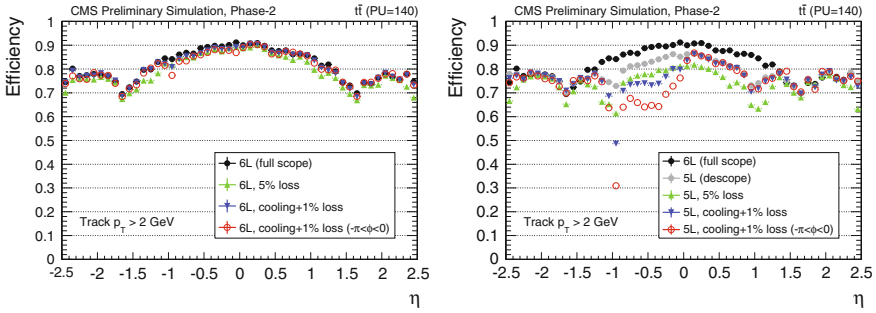


Fig. 7.13 Simulation of Track Trigger perforance. The *left* plot demonstrates the robustness and redundancy of a six-layer detector compared to a five-layer detector (*right*) where failures are largely detrimental [302]

communication between left to right end[9], thus tracks passing close to the middle and at an angle cannot be correlated by the front-ends thus no stub is being produced.

Tilting the sensors (modules) with respect to a beam-parallel orientation; turning them to face the track orthogonal remedies this and all stubs are found. Track Trigger efficiencies would be reduced at larger z by about 30% without tilting the modules. Figure 7.12 illustrates how the stubs are being recovered by the tilt. With the lower number of modules, the tilted layout also reduces the stub rate by about a factor of two compared to the flat geometry, decreasing data transmission rate and facilitating Level-1 track finding.

The simulation results shown in Fig. 7.13 demonstrate the necessity for a minimum of six barrel p_T-layers. It should be mentioned that with a high granular four-layer pixel detector, missing 1 to 2 outer layers would not be much detrimental for offline reconstruction but the Track Trigger is based solely on the Outer Tracker. For lower radii ($r < 20$ cm), tracks are all 'straight' within any mechanically reasonable sensor to sensor spacing (Δr) in an already very densely populated volume. A p_T based L1-

[9]Establishing a communication line between the two module/sensor ends would be an alternative but the corresponding R&D to establish this with **T**hrough **S**ilicon **V**ias TSV proved to be a not yet mature technology.

trigger is therefore not possible with the inner micro-pixel detector (Inner Tracker). The left plot, featuring six OT layers, demonstrates its robustness versus random and globally localised failures, e.g. like a cooling loop spanning a full hemisphere (180°). The five layer detector starts with a lower efficiency but suffers significantly more from failures, especially locally. Six layers provide the necessary redundancy.

The dedicated back-end electronics has about 3 to 4 μs for track identification and track fitting to deliver tracks to the CMS global trigger with further trigger recipes to allow final L1-accept signal arriving back at the front-end within the latency of 12.8 μs before the data is overwritten. Tracks above a defined momentum have a maximum bending thus are always confined in a defined detector segment/volume. Virtually segmenting the detector in such slices ensures that a high p_T track stays within such a segment or within the segment plus a neighbour segment. Taking the front-end to back-end cabling into account, time-multiplexing[10] can now deliver all data of such segments plus their neighbours for each individual bunch crossing to a single track-trigger processing unit. The track trigger system starts with pattern recognition of the already reduced set of stubs (compared with the full set of space points with full readout). The identified tracks are then being fitted, doublets are being removed and final information is being sent to the CMS trigger.

Three pattern recognition methods have been developed and demonstrated to work. More info can be found at [303] and [240].

- **A**ssociative **M**emories AM (custom ASICs) basically storing all possible a priori calculated patterns and then compare them with the found stubs
- the 'Tracklet' approach where stubs of one layer are combined with stubs in the second layer forming a so-called 'Tracklet' which is then followed to the next stubs/tracklet – all realized in FPGAs
- Patterns/Tracks are identified by Hough Transforming the stub information – all realized in FPGAs

The track fitting has also been fully realized in FPGAs ranging from χ^2 fit to a full Kalman filter. All methods have been fully demonstrated in hardware; all residing within the required 3 to 4 μs.

7.1.1 Sensors for the HL-LHC CMS Tracker

This section describes steps during the sensor development and evaluation and not the very final implementation.

As already discussed in Sect. 2.2.4, with the onset of trapping at high fluences, the concept of full depletion voltage becomes more and more abstract and less relevant. In principle we only need to guarantee high efficiency, high resolution and low power

[10]Data from multiple sources are routed through a multiplexing network which directs all the data from an individual bunch crossing to a single processor.

consumption, but we are also interested in current, CV[11] scans to evaluate N_{eff} and its profile, trapping, CCE, **s**ignal-to-**n**oise S/N and strip parameters (inter-strip resistance, inter-strip capacitance, etc.).

The CMS strip-, macro-pixel- and micro-pixel-sensor R&D is far advanced but some parameters, e.g. thickness, punch-through structures, lengths of polysilicon resistors, p^+-stop structures, doping concentrations, active edges, alignment markers, test structures are still under final refinement. Global parameters like bulk material (n-in-p), outer dimensions and, pitches are chosen.

To reach this point, CMS conducted a long and in depth R&D with the goal to evaluate (a) base material with respect to radiation tolerance and annealing, (b) geometry and (c) special designs, e.g. routing structures. Figure 7.14 shows the basic wafer layout containing diodes, several sizes of mini-sensors, dedicated test structures (improved version of the former CMS ones), large areas of 2S and PS sensors with different pitches and strip width-to-pitch versions and two different lengths of macro-pixels. Wafers have been procured in large variety of base materials and thicknesses from a single vendor to allow absolute comparison of results. CMS procured Epitaxial material **Epi**, standard floatzone **FZ**, deep diffused FZ **dd-FZ**, magnetic Czochralski **mCz** all in p-in-n & n-in-p in different active thicknesses ranging from 50 µm to 300 µm (thinning rf. also Sect. 1.9.3, p.93). Some of the p-in-n & n-in-p came with a double-metal layer to study routing options, mainly implementation of the **p**itch **a**dapter PA functionality in the sensor itself saving the glass PA, and thus its mass[12] (routing rf. also Sect. 1.4). For the macro-pixels (DC-coupled) punch-through as well as polysilicon biasing has been explored (biasing rf. also Fig. 1.27 on p.34). It is worth to mention that the wafer-backplane is an Al-grid (mm-scale) to allow laser-light injection from the back for CCE and TCT studies. The full granularity is listed in Table 7.2. The main results have been published here [20, 73, 82, 89, 146, 196, 222, 223, 243, 244, 308, 314, 328].

Learning from RD50 (rf. Sect. 2.2) that NIEL is violated and only the real mix/percentages of charged and neutral particle radiation at the given location fully validates the material, CMS irradiated the wafers to the more realistic radiation mix for several radii. The main important values are **s**ignal-to-**n**oise S/N or **C**harge **C**ollection **E**fficiency CCE but also IV, CV and all strip parameters have been investigated; e.g. to determine depletion voltage and power consumption. It is worth to mention that, in all cases, strip sensors plus diodes have been investigated and LHC electronics with correct sampling times and frequency have been used. Using strip sensors allowed full characterization of strip parameters, like inter-strip resistance, inter-strip capacitances, coupling capacitances, bias resistors, etc. Many structures, especially diodes, have also been evaluated with TCT, DLTS and TSC (rf. Sect. 1.8)

[11] At the relevant fluences, depletion voltage (V_{FD}) is a more abstract concept.

[12] The R&D was successful but it is even more economic to implement the PA routing into the hybrids.

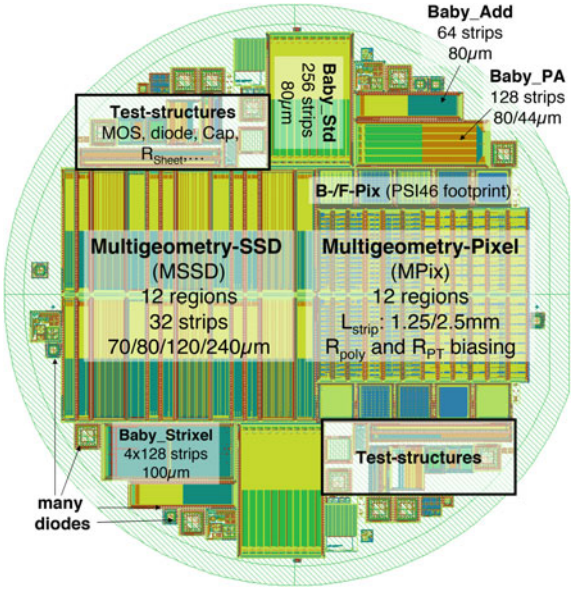


Fig. 7.14 R&D wafer with different structures, described in the text to evaluate base material, design rules and geometry with respect to radiation and subsequent annealing. The main work-horse for the radiation/annealing campaign are diodes, Baby_Std and Baby_Add but also the multi-geometry structures have partially been irradiated; plus some of the test-structures. Multi-geometry-strip detectors and multi-geometry macro-pixel have been evaluated for geometry and design optimization. Baby_Add was used to study Lorentz angle wrt. to radiation, temperature and voltages. The test-structures are being used for process evaluation [308]

Table 7.2 The table shows the variety of wafers composing the several-year long testing campaign of CMS. Some of the wafers have been processed with a second metal layer for routing studies. The structures on the wafers can be visited in Fig. 7.14

Material	Bulk doping	Isolation for n-in-p	Thickness [μm]
EPI	p-in-n, n-in-p	p^+-stop, p-spray	50, 70, 100
FZ physical thickness	p-in-n, n-in-p	p^+-stop, p-spray	200, 300
FZ physical on carrier wafer	p-in-n, n-in-p	p^+-stop, p-spray	120, 200
dd-FZ	p-in-n, n-in-p	p^+-stop, p-spray	120, 200
mCz[a]	p-in-n, n-in-p	p^+-stop, p-spray	200

[a] Deep diffusion is not available for mCz, due to the thermal donor creation during the high temperature process thus drastically changing the base material. Some discussion happened recently with a wafer provider, which believes it could be done with a "donor-kill" temperature cycle – very high temperature to break the oxygen clusters and then cool down fast

for different radiation levels. The measurements have been cross-checked with simulation and often led to a better tuning of the simulation itself (rf. also Sect. 1.8.5, p.80). Finally, after the mixed irradiations, the structures have been subjected to a longterm annealing campaign.

Table 7.3 Integrated Radiation fluences expected after 3000 fb^{-1} according to FLUKA simulations at the time and the corresponding radius in the detector. The small radii 5 and 10 cm * are only relevant for *thin* micro-pixel structures, thus only diodes and a small number of Baby_Stds have been irradiated to that level. Structures have always been measured after individual neutron and proton fluence and with the mix in the end

Radius (cm)	Protons	Neutrons	Total	Ratio	Thickness (μm)	Sensor type
	Fluence in ($10^{14} n_{1MeV}/cm^2$)					
60	3	4	7	0.75	≥200	2S sensors
20	10	5	15	2.0	≥200	PS sensors
15	15	6	21	2.5	≥200	Micropixels
10*	30	7	37	4.29	≤200	Micropixels
5*	130	10	140	13	≤200	Micropixels

The main work-horse structures for the radiation tolerance studies have been implemented twice on the wafer (upper and lower half) allowing to irradiate one structure with neutrons and one with protons, measure these and then subject the neutron irradiated one to proton irradiation and vice versa resulting in two pieces irradiated to the correct fluence-mix. Table 7.3 shows the main fluences Baby_Std (rf. Fig. 7.14) and diodes have been submitted to.

All measurement results have been saved in a global database to allow full comparison and evaluation; all participating institutes went through a measurement cross-calibration early on.

In a nut-shell the campaign, with several tens of man-years of work, validated the routing concepts and identified the Phase II sensor baseline for the CMS Outer Tracker, namely n-in-p with an active thickness of 200 μm. The following radiation results of the R&D campaign allow also clear recommendation for the IT. As a result, thin (n-in-p) planar sensors are the baseline for most parts of the CMS micro-pixel volume, while 3D sensors (rf. also Sect. 1.12.7) are also candidates for the innermost layer. Final layout and even final cell geometries are still being discussed but are close to be frozen. Also the active thickness needs final evaluation.

Outer Tracker Sensors for the HL-LHC Tracker

The several year long CMS R&D campaign produced a very exhaustive set of data for different materials, thicknesses, geometries etc.. The next section gives an excerpt only, while a more detailed discussion can be found in [328]. The campaign concentrated on measurements of small strip sensors plus diodes. All strip parameters, also inter-strip resistance and capacitance stayed within limits for all levels of radiation independent of strip isolation strategies.

Figures 7.15 and 7.16 demonstrate the main results leading to the baseline sensor choice – n-in-p with an active thickness of 200 μm (or at least below 300 μm).

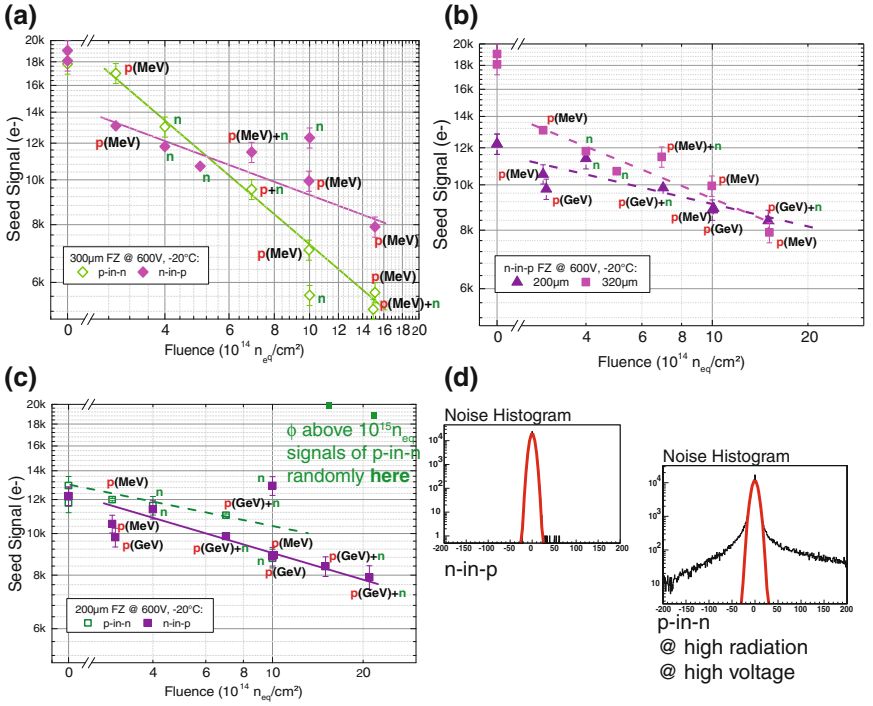


Fig. 7.15 Overview of the results of the CMS Phase II sensor campaign. 'Seed' signals – single strip signals are plotted and not the summed signal of strip clusters. Figure **(a) - (c)** show the 'seed signal' evolution versus radiation fluences for *n*- and *p*-bulk for thicknesses of 200 and 300 µm. The *small labels* indicate the type of radiation namely neutron (∼1 MeV) or proton of different energies (23 MeV or 24 GeV) and the mixed cases (proton+ neutrons – labled as p+n) as described in the main text. Figure **(d)** show the noise histogram of *n*-in-*p* sensors (*left*) and for *p*-in-*n* (*right*) for higher radiation levels. The *p*-in-*n* show anomalous high non-Gaussian noise which is often higher then the signal itself; we see micro-discharges at high fields (or locations of charge amplification) [78, 308, 328]

The future CMS readout is binary[13] thus registers only above-threshold signals. Therefore, only 'seed' strip signals are evaluated and shown.

Figure 7.15a compares $D = 300\,\mu\text{m}$ thick *p*-in-*n* (hole readout) with *n*-in-*p* strip sensors (electron readout). As expected, at a fluence level of $6 \cdot 10^{14}\ n_{1\text{MeV}}/\text{cm}^2$, where trapping is becoming the dominant detrimental effect, charge collection for holes is decreasing below that of electrons (slower drift of holes thus more affected by trapping, better electric and weighting field combination and also higher electric field at *n*-electrodes after irradiation; rf. also Sect. 2.2.4). For the 300 µm thick sensors, the hole signal clearly drops below the electron signal and also below the readout threshold.

[13]The on-chip comparator has no cluster algorithm but works individually for each channel. No centre-of-charge can be applied since the smaller signals are simply not recorded; resulting in binary position resolution $\sigma_x = \frac{pitch}{\sqrt{12}}$.

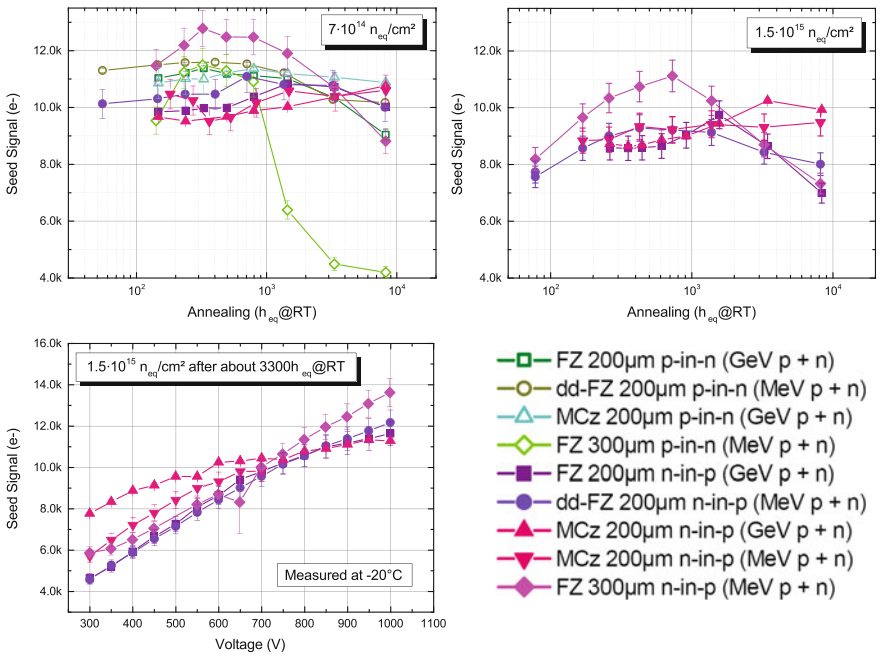


Fig. 7.16 The annealing for highly irradiated sensors are shown (*n*-type bulk only for *lower* fluences); all times normalised to hours at room temperature. Lines are drawn to guide the eye. The legend names the material, followed by sensor thickness, polarity and finally type of irradiations, e.g. "dd-FZ 200µm *n*-in-*p* (MeV/p+n)". FZ stands for floatzone, mCz for magnetic Czochralski and dd-FZ for **d**eep **d**iffused-FZ. The *upper* plots demonstrate clearly that signals of 200 µm thick sensors do not decrease with annealing time while 300 µm, especially *p*-in-*n* do. The *lower left* plot shows the signals versus bias voltage demonstrating that higher voltage improves the margin. The superior behaviour of mCz is explained in the main text and demonstrated further in Fig. 7.17 [308]

The advantage of higher primary charge generation for larger thicknesses D becomes marginal at high fluences when trapping dominates. Thinner sensors have less current ($I \sim D$) and much smaller depletion voltages ($V_{FD} \sim D^2$), meaning the electric fields across the sensor and at the electrodes are much higher! Thinner sensors also draw less power ($P \sim D^3$). We will revisit this effect when we discuss the annealing behaviour with Fig. 7.16. Figure 7.15b, comparing 200 and 300 µm thick *n*-in-*p* strip sensors, show the above described effect.

With $V_{\text{bias}} = 600$ V at $\Phi_{eq} = 10^{15}\ n_{1\text{MeV}}/\text{cm}^2$, 300 µm thick silicon does not provide more signal than a 200 µm one; a marginal improvement at higher voltage (e.g. 900 V) exists.

Figure 7.15c compares *p*-in-*n* with *n*-in-*p*, both 200 µm thick. The significant detrimental effect as seen for the D = 300 µm thick comparison is clearly gone and at first look, *p*-in-*n* seems even superior to *n*-in-*p*. As a matter of fact, field strengths at the electrodes for thin sensors are high (enough) for both types. To note, the 'higher' signal for *n*-in-*p* is present only for strip sensor and not for diodes thus is a consequence of the segmentation-design and not of the base material. Taking

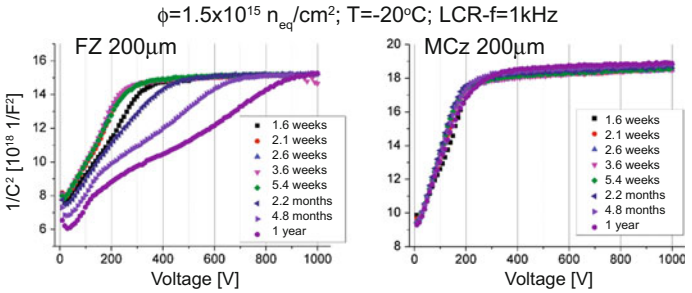


Fig. 7.17 The CV scans of mCz, compared to FZ, show basically no change in depletion voltage nor electric field strength [143]

not only signal but also noise into account the picture changes drastically. At higher fluences and/or higher voltages, p-in-n strip sensors exhibit large non-Gaussian[14] tails in the distribution expressed in Fig. 7.15d. These would result in fake hits. De facto, p-in-n sensors see charge amplification (rf. also Sect. 2.2.5). At high fluences (where the plot shows no green data points), non-physical signals up to 44k electrons and down to 5k electrons have been measured. Dedicated simulations[15] confirm very high fields at the p^+-electrodes for p-in-n strips sensors, while the fields are more 'distributed' for n-in-p sensors with additional p^+-spray or p^+-stops.

Figure 7.16 shows the annealing behaviour of 200 and 300 µm thick n- and p-type sensors. Signals of the 300 µm thick sensors (especially p-in-n) drop with annealing time, while signals from thin sensors stay basically constant. As shown earlier, N_{eff} changes with annealing thus field strengths changes and this affects thinner sensors simply much less than thicker ones, since electric fields always stay high enough.

This gives a fantastic outlook for future operation without the need to keep the detector cold during maintenance periods or even better to warm it up in a controlled fashion, benefiting from the leakage current annealing. This is already true for V_{bias}=600 V, and with even more margin towards higher bias voltages increasing field strengths. The full Tracker system is designed to cope with 800 V – ample margin.

Looking closely, mCz[16] is strikingly superior to FZ with even less changes during the annealing process. Figure 7.17 shows virtually no change in CV scans for all

[14]This non-Gaussian, also called anomalous, noise exists for the p-in-n sensors in this campaign. It is probably that this could be overcome with an adapted strips design but it still looks like, that n-in-p sensors with p^+-stops have more margin than p-in-n.

[15]Device simulations [316] have shown that irradiated p-in-n strip sensors develop high electric fields at the strip edges, intensifying with increasing accumulation of oxide charge. The electric fields around the n-strips in n-in-p sensors are instead reduced by higher oxide charge, which makes them more robust with respect to effects, such as breakdown, noise or micro-discharge after heavy irradiation with charged particles.

[16]The oxygen content of dd-FZ is already very high due to the high temperature treatment. Still, the oxygen concentration in mCz is one order of magnitude higher and due to the crystal growth process in the magnetic field more homogeneous.

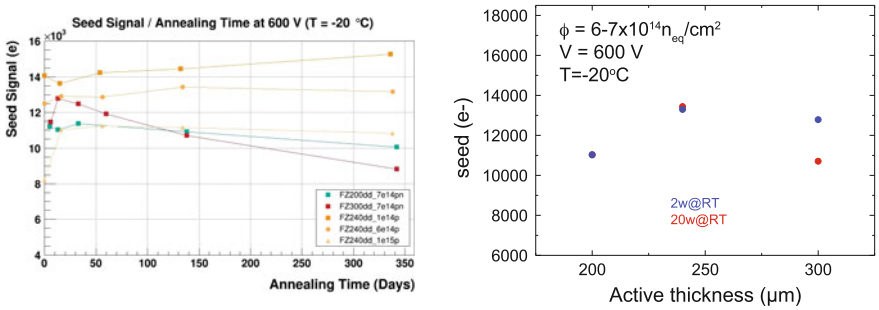


Fig. 7.18 The plots demonstrate the signal robustness of thin sensors against annealing. It seems that 240 is as robust as 200 μm with globally a larger signal. The *left* plot shows 'seed signal' versus annealing normalised to hours at room temperature. The *right* plot compares signals after irradiation for three thicknesses before and after annealing [78]

annealing times thus a low and constant depletion voltage, thus high electric field strength. The changes of electric field strength in FZ are inside working parameters, the E-fields of mCz do not even change.

Adding all observations up, the optimal choice would be mCz n-in-p with a thickness of $d = 200$ μm implemented in a system with enough margin with respect to high voltage robustness. Unfortunately, mCz might not be readily available in quantity in the high resistivity ρ regime[17] thus the next best choice is FZ or dd-FZ available with higher ρ. Now, the signal in 200 μm is a bit marginal, with respect to the noise in the front-end readout electronics, thus maybe the best choice lies in-between $d = 300$ and 200 μm. Full size prototypes (along with baby sensors on the wafer half-moons) have been procured and evaluated. The results can be visited in Fig. 7.18, demonstrating that an active thickness of 240 μm is as robust as 200 μm with globally a larger signal. Thickness can be further tuned; it can be chosen precisely by physical thinning or deep diffusion.

Due to different field strengths, leakage currents at same voltage are not scaling exactly with thickness; they are always a bit higher than expected but are still lower for thin sensors (200 μm thick sensor currents are about 85% of 320 μm).

Table 7.4 presents the different sensor dimensions of the Outer Tracker (TB2S and TBPS). The 2S and PS_strip sensors are common single-sided, AC-coupled, polysilicon biased n-in-p with atoll-p^+-stop isolation. The PS_pixel sensor is a single-sided, DC-coupled, punch-through biased[18] n-in-p with atoll-p^+-stop isolation. Dimensions make use of the maximal area within the vendor design rules of a 6 in. wafer. Edges, metal overhang, p^+-stop concentrations ([243]), etc. have been optimized with simulation and cross-checked on different process runs. The sensors

[17] As reminder, low resistivity ρ would result in a high starting depletion voltage which for n-in-p sensors only increase with radiation.

[18] Also polysilicon is an option.

Table 7.4 The recipe used for the current CMS sensor strip layouts (rf. Sect. 6.4.2) also holds for thinner *n*-in-*p* sensors: All sensors feature a width-to-pitch ratio of w/p=0.25 and a metal overhang of 5 μm for strips and 5 μm for the macro-pixels

Sensor type	Width (mm)	Length (mm)	Strip/Pixel		Number of strips/pixels
			Pitch	Length	
Strip-Strip (2S)	94.183	102.7	90 μm	5 cm	2 × 1016 = 2032
Pixel-Strip (PS_s)	98.74	49.16	100 μm	2.5 cm	2 × 960 = 1920
Pixel-Strip (PS_p)	98.74	49.16	100 μm	1446 μm	32 × 960 = 30720

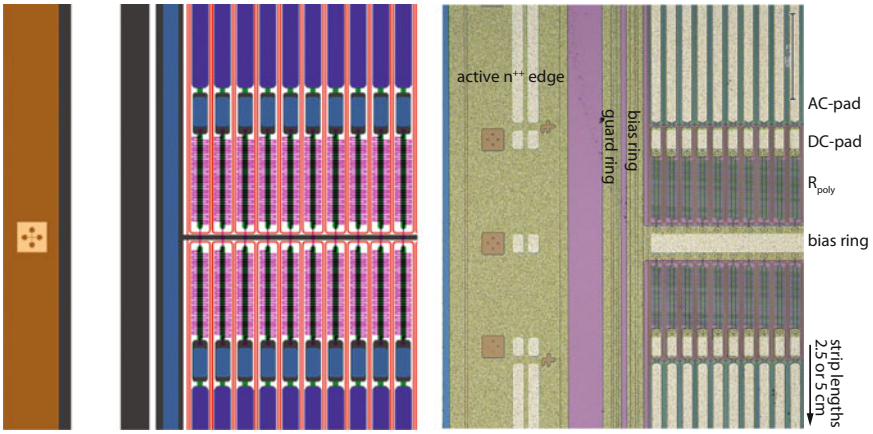


Fig. 7.19 Centre edge of a 2S sensor. The *left* picture shows the design while the *right* a real photo. The polysilicon bias resistors R_{poly}, DC and AC pads are all located in the sensor center to facilitate testing in on row instead of at the sensor edges. The picture also nicely shows the atoll p^+-stops guaranteeing the strips isolations in an *n*-in-*p* sensor. Looking closely, one can see the vias from DC pad to R_{poly} and back to the bias ring. Lighter parts on the pads and rings are openings in the passivation

will be procured thinned or utilising the deep diffusion[19] process which might save cost at some vendor. dd-FZ has the advantage of high mechanical stability and no need for thinning at the company but obviously introduces more mass to the Tracker with respect to a physically thinned sensor.

Figure 7.19 shows a potential sensor design which has also been prototyped in some varieties. The details of the centre of the strip sensors (2S and PS_strip) where the DC-pads and polysilicon resistors are located to allow testing all strips (both ends) in a single row. In addition, in the absence of bias resistors at the sensor ends, the AC pads could be moved further out decreasing wire-bond length.

[19] Only dd-FZ exists, the deep diffusion process is not available for mCz. Although some companies might explore dd for mCz.

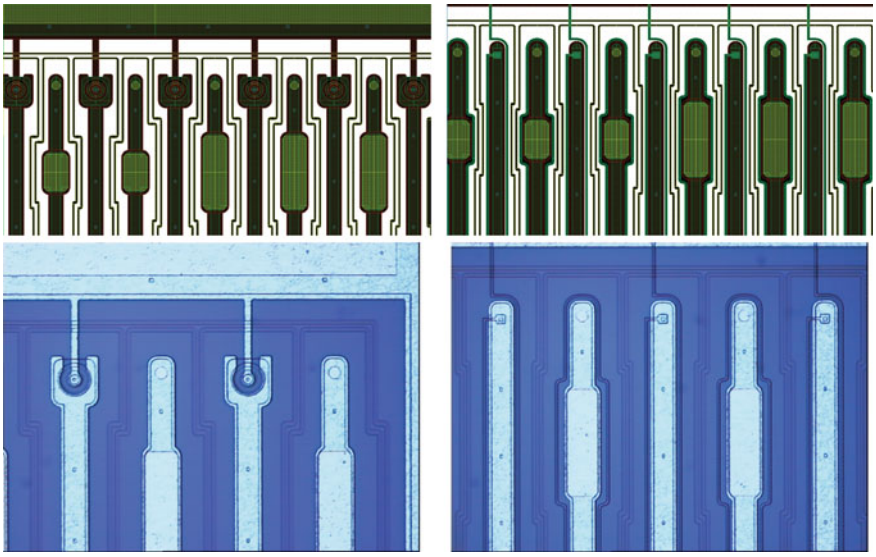


Fig. 7.20 The CMS PS_p sensor; layout and photo. The *left* shows the **p**unch-**t**hrough PT implementation and the *right* the polysilicon one, were the bias resistor runs around the full macro-pixel. All strips are surrounded by individual p^+-stops

Figure 7.20 shows the layouts and photos of early prototypes of the PS_pixel sensors. **P**unch-**T**hrough PT and polysilion biasing have been tried. Both are working and PT requires less processing steps. The p^+-stop design is also roughly exemplarily for 2S and the PS_strip sensor with an atoll for every single strip/macro-pixel plus a common one surrounding the combined strip area, isolating against the bias and guard rings.

The CMS Tracker community was also exploring, mainly for potential cost savings, the usage of 8 in. wafers despite the fact that everything was designed for sensor sizes optimised for 6 in. ones. For the macro-pixel ($5 \times 10\,\text{cm}^2$; DC-coupled) wafers, three sensors would fit on a wafer instead of two. One 2S ($10 \times 10\,\text{cm}^2$; AC-coupled) plus one PS_strip ($5 \times 10\,\text{cm}^2$; AC-coupled) sensor would fit but a smaller number of PS_strip sensors than 2S sensors are needed. One gains when the price increase from 6 to 8 in. is less then a factor 1.5. In earlier design studies, CMS explored and produced a 2S sensor with a special size of $16 \times 10\,\text{cm}^2$ with $D = 200\,\mu\text{m}$ thickness to make proper use of the 8 in. area, but dismissed it finally. The photo of this special 2S-long sensor, the first 8 in. wafer ever processed in HEP is shown in Fig. B.7 on p.356 in the Appendix, for more information see [36].

Inner Tracker IT Sensors for the HL-LHC Tracker

Being subjected to even higher radiation levels than the OT, the micro-pixel sensors will also be processed in n-in-p technology collecting electrons. The innermost layer will have to withstand up to $\Phi \approx 2 \cdot 10^{16}\ n_{1\text{MeV}}/\text{cm}^2$ of fluence or a dose of

0.5 – 1 GRad integrated after $\mathscr{L} = 3000\ \text{fb}^{-1}$. CMS will keep the global detector design, allowing for access without beam pipe removal.

The sensor thickness will be reduced even more to achieve high enough fields after high radiation and annealing. Values like 100 – 150 μm of active thickness are being evaluated – physical thinning, active on handle wafer or deep diffused material (dd-FZ). With much smaller capacitances compared to strips thus much lower 'noise' smaller signals can be accommodated for. The specifications for the RD53 pixel readout chip target a threshold below 1000 e^- down to 600 e^- thus signals about 5000 to 6000 e^- are ok. Indications are good, but due to the lack of radiation tolerance ASICs with small cell sizes (pitches), the usage of planar thin sensors have not been fully proven for the innermost layer. Therefore, radiation tolerant 3D sensors (rf. also Sect. 1.12.7) are also probable candidates for the innermost layer. But, the small pitch, thus 3D-column small **width to depth**, will be a technology challenge to establish in 3D technology.

With direct ASIC to pixel channel **b**ump **b**onding BB the sensor size has to be a multiple of the chip size. Modules will be only rectangular, no wedge shape is planned. With chip sizes about $2 \times 2\ \text{cm}^2$, final BB yield will most probably define the maximum size module, while 2×1 and 2×2 versions are being evaluated. The electronic cell will be $50 \times 50\ \mu\text{m}^2$ while sensor cells will be $25 \times 100\ \mu\text{m}^2$, $50 \times 50\ \mu\text{m}^2$ or potentially a multiple of this (50×200, 100×100) where only some electronics channels are being connected to save power in the outer layers. The BB pitch is defined by the electronics cell and the BB ball size by the smallest dimension of the sensor pixel – BB-pad on ASIC defines placement/routing on sensor. The BB pad spacings of sensor can be seen in the prototype layouts shown Fig. 7.21. Industry is setting-up for this feature size. With very small pitches like $25 \times 100\ \mu\text{m}^2$ charge is basically always shared between two pixels, improving resolution.[20] As mentioned earlier $25 \times 100\ \mu\text{m}^2$ sensor cells preliminary seems to be the best choice despite slightly degraded z-resolution. Another aspect of the thin sensor is the corresponding mechanical bow or warping resulting in extra complication for BB. Often edges lose BB contacts. Possibly deep diffused material with the good balance of active thickness[21] and mechanical stability could be the best choice; e.g. physical thickness of 200 μm and active of 130 μm. In general mechanical stability degrades with decreasing physical thickness and increasing surface.

Another challenge posed by the small pixel size is the pixel isolation and biasing strategy. The p^+-stops barely fit in between the cells, especially in the atoll-geometry, but it is doable. Common p^+-stops would give more margin but also more risk since a local defect could influence much larger areas.

In the final configuration, pixels are directly and individually connected to the pixel ASIC thus a biasing grid is per se not necessary. It serves for initial testing and could provide protection against large currents in the final system. The protection feature, though, could be implemented in the ASIC. With a very high processing yield, testing

[20] Time-over-threshold readout will be used and centre-of-charge calculation is possible; different to the binary readout in the Outer Tracker.

[21] Final active thickness still to be optimized – 130 μm is just an example.

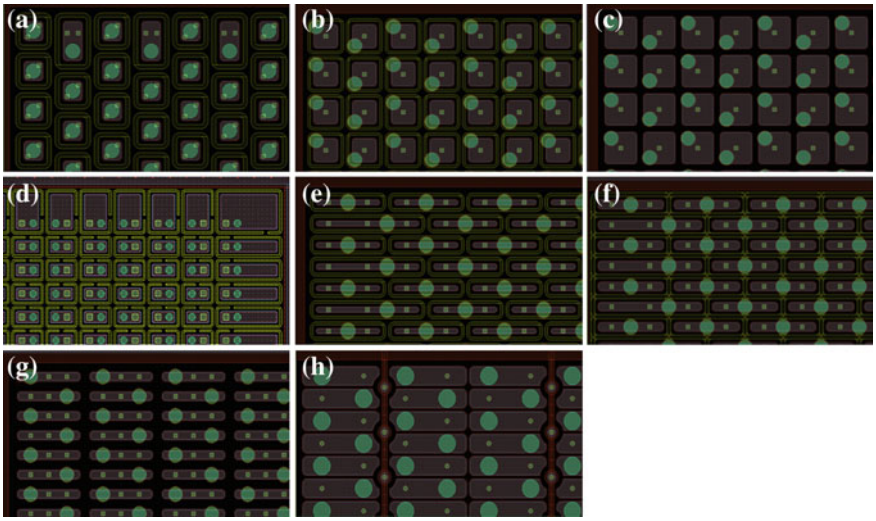


Fig. 7.21 Several potential small pixel cell layouts are shown. The two baseline cell dimensions $50\times50\ \mu m^2$ (*upper row*) and $25\times100\ \mu m^2$ (*lower rows*) are being studied. The *small squares* show the via connections between metal and implant, the *circles* the bump bond pad. Their positions are always staggered to maximise the bond pitches. Layouts **(a), (b), (d), (e)** feature individual p^+-stops, layout **(d)** has an opening in p^+-stop similar to the CMS forward pixel layout shown in Fig. 6.24. Layout **(f)** shows a common p^+-stop configuration. p^+-spray configurations are shown in **(c)**, **(g)**, **(h)**. With a pixel cell area of $50\times50\ \mu m^2$, space is a premium; p^+-spray implants can be larger due to the missing p^+-stops. It is worth to note that only **(d)** and **(h)** can be power tested at sensor level while all others connect to ground ONLY after bump bonding through the electronic chips. **P**unch-**T**hroughs PT connections for each individual cell are basically excluded and a common PT scheme, where one dot connects to four cells is presented in **(h)**

could be reduced to the minimum and a bias grid might not be necessary. This would also largely simplify the layout thus processing and thus probably yield. Another aspect is that every **p**unch-**t**hrough PT contact and the bias rails themselves are areas of slightly degraded efficiency.

Figure 7.21 shows potential layouts for the pixel upgrade. Cell dimensions are either $50 \times 50\ \mu m^2$ or $25 \times 100\ \mu m^2$. These are very challenging; implementations of p^+-stops and on sensor biasing schemes are demanding. Punch-through PT biasing per pixel would use up comparatively more space of the pixel cell and would render substantial amount of area potentially less effective for charge collection; a common PT connection for four pixels seems feasible. A probable scheme without any sensor bias scheme (no PT, polybias or even a bias grid) is envisaged where final connection only happens after bump bonding. In principle an extra metal layer, short-circuiting all pixel cells could be applied during production to make an IV-scan and then be removed before assembly. The figure is educational, showing many different layout possibilities. The only fix point in all configurations is to match the bump bonding pads of sensors and electronic chips. The cell geometries are largely independent

and even longer routings than shown here would be possible. It is important though to maximise space between bond pads. The figure shows a staggered **a, e, f** (also called bricked design) and non-staggered **b, c, d, g, h** cell placement.

Another challenge on system level is the danger of sparking since n-in-p sensors have the high voltage at the sensor cut edges and the bumb-bonded chip on GND-potential nearby. Several options are being explored from dedicated extra passivation[22] on sensor level to under-filling between sensor and chips. This is a real challenge.

CMS will minimise the non-active edges as much as possible with planar technology but is not exploring slim edges with DRIE processes. In case, 3D technology will be used for the innermost layer, DRIE comes with the process thus can be envisaged as active edge processing.

7.2 The CMS Endcap Calorimeter Upgrade for the HL-LHC

The CMS **H**igh **G**ranularity **C**alorimeter HGC, replacing the Run I endcap electromagnetic and hadronic calorimeter, will be the largest silicon based detector, with a larger silicon surface than all existing silicon based tracking detectors altogether. Radiation levels are as challenging as in the innermost trackers and potentially even beyond.[23] According to simulations (see Fig. 7.1 right plot on p.292), radiation levels will reach $\Phi_{\text{eq}} = 1.6 \cdot 10^{16}\, n_{1\text{MeV}}/\text{cm}^2$ (at $|\eta| = 3$) very similar to the innermost layer of the Inner Tracker (pixel detector).

A tracker detects the individual charged particles themselves (even in dense jets) while the calorimeter detects electromagnetic[24] or hadronic showers resulting from the original particle and its interaction with the calorimeter absorber. The massive calorimeter stops the particle fully, absorbing and measuring its full energy while the light-weight Tracker only registers the small amount of ionisation energy of the traversing particle—a **m**inimum **i**onizing **p**article MIP (changing particle energy and momentum as little as possible). The shower energy is still distributed over many cells but each HGC cell gets a much higher signal than sensor cells in a Tracker *allowing quite novel and different design choices*. The HGC has only some sensitivity to MIPs for calibration reasons, but has not been designed to detect individual MIPs.

The details of the following chapter are mainly based on [308] to show the concept in a self-consistent way. In reality, several design choices are still in the flux these days and some improvements and probable (or decided) design changes will be listed in the last paragraph later.

---

[22] E.g. BCB **B**enzo**c**yclo**b**utene deposition and parylene coating is being investigated.

[23] In case the layout would be enlarged to cover up to $|\eta| = 4$ the expected levels would increase to even $\Phi_{\text{eq}} = 10^{17}\, n_{1\text{MeV}}/\text{cm}^2$.

[24] Electrons, positrons from pair production and γs from Bremsstrahlung.

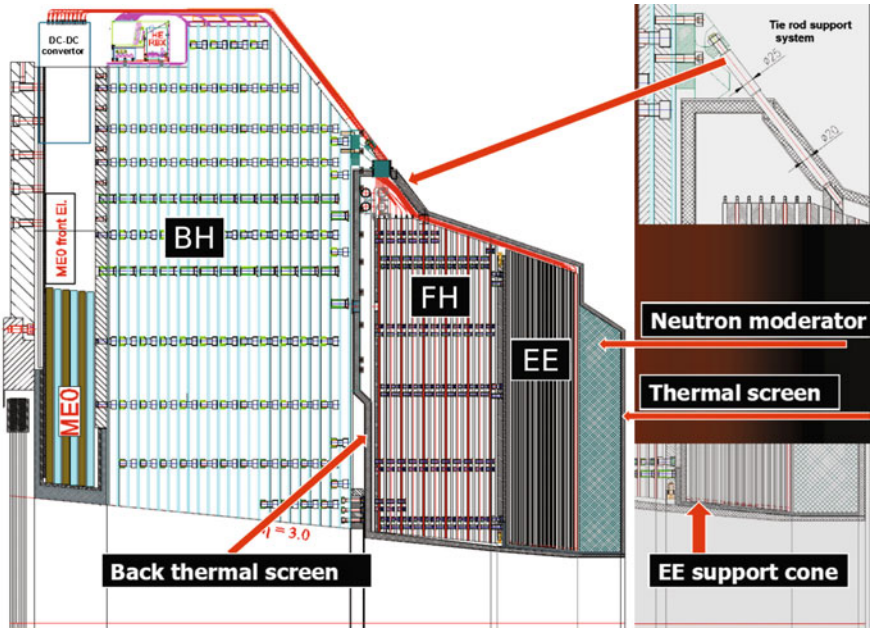


Fig. 7.22 The design layout of the HGC – CMS Technical Proposal [308]

The HGC is a new high-granularity sampling calorimeter, covering $1.5 < |\eta| < 3.0$ with a silicon/tungsten electromagnetic section followed by two hadronic sections, both using brass as the primary absorber material. In the front layers, the active material is composed of silicon pad sensors while the back layers will be instrumented with plastic scintillators. This chapter describes primarily the silicon instrumented sections and discusses the different and/or similar challenges compared to the tracker and the proposed optimized solutions.

The proposed layout is shown in Fig. 7.22 and more details can be found in [308]. The system is divided in three parts, the **E**lectromagnetic **E**ndcap Calorimeter EE, the **F**orward **H**adron calorimeter FH and the **B**acking **H**adron calorimeter BH. The global design parameters are listed in Table 7.5. Details about the silicon sensors can be found in Table 7.7 on p.325.

Inspired by CALICE [320, 335], the HGC approach majorily champions the **p**article **f**low PF concept, but the (a) high radiation tolerance of silicon sensors and (b) the possibility of mass production of highly, dedicatedly structured sensors, was the main reason for the choice. With the relatively high channel granularity, the HGC will be able to continue "*tracking*" inside the calorimeter. This will be especially important in the high pile-up environment, allowing particle separation and energy-particle linking (PF). The example of a simulated event is displayed in Fig. 7.23.

Table 7.5 The table shows the global design parameters as described here (Technical Proposal version, final numbers and design choices might change) [308]

	EE	FH	BH	Total
Layers	28	12	11	41
Active material	Silicon	Silicon	Scintillator	
Absorber material	tungsten/copper	brass	brass	
Channels	4.3M	1.8M		6.1M
Area of Silicon (m^2)	380	209		589
Hadronic interaction length[a] λ	$1.3 \approx 25X_0$	3.5	5.5	10.3
Temperature	$-30°C$	$-30°C$	RT	

[a] λ Hadronic interaction length. Muon chambers are normally placed behind at least 10λ corresponding to $P = e^{-(x)/\lambda} = 4.5 \cdot 10^{-5}$ for $x = 10\lambda$, 'preventing' all particles but muons from reaching them

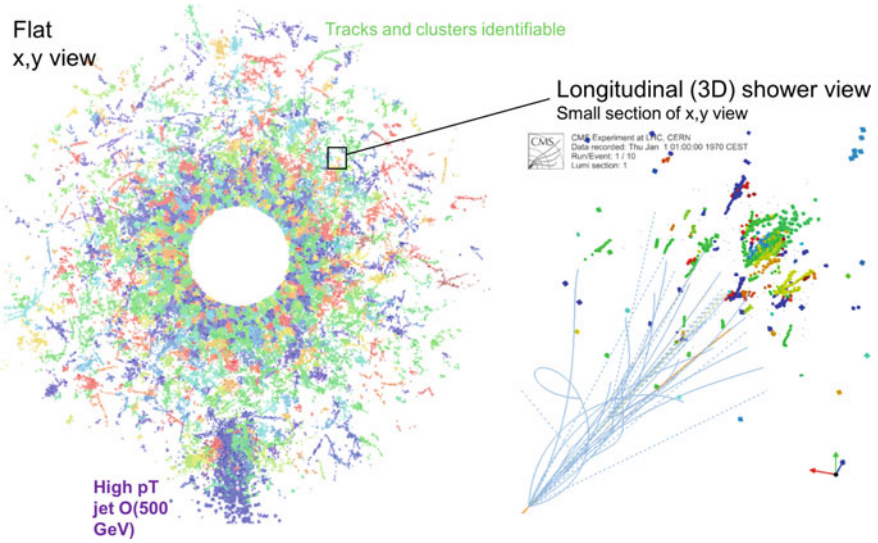


Fig. 7.23 A simulated 200 pile-up event of the CMS HGC. The high granularity allows, even in a high pile-up situation, to associate energy deposition to individual particles. This detailed topical information is ideally suited for the particle flow concept. The different colours aide, to see the energy deposition of different particles. The HGC "*tracks*" particles in the calorimeter [Courtesy CMS]

In EE, tungsten absorbers alternate with active layers. The active layers are composed of a complicated multi-sandwich. A 6 mm thick centre plate consists of Cu with embedded cooling pipes with modules placed on both faces. The situation is illustrated in Fig. 7.24. The FH structure is similar but the absorber will be made of brass with single plane structures in-between, instead of the double-module concept with Cu-absorber cooling plate.

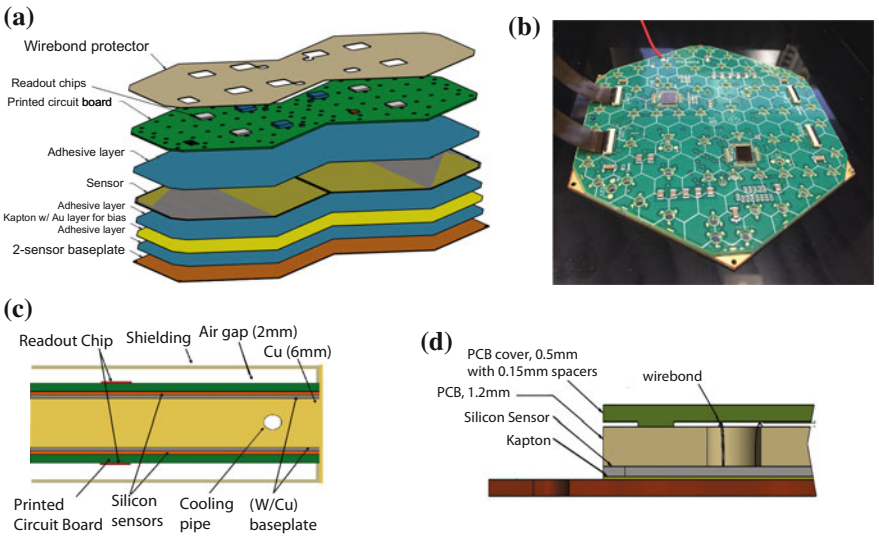


Fig. 7.24 HGC modules on the cooling plate **(a)** the module schematics showing the different layers (printed circuit board, silicon sensors, and W/Cu (75/25%)-baseplate); **(b)** photo of an early half-module prototype; **(c)** Multi-sandwich or double-module (HGC module with W/Cu-base plate – Cu with embedded CO_2 cooling pipe – HGC module with W/Cu-base plate), the copper serves as additional absorber; **(d)** illustrates the detail how the PCB – sensor contact is realized via wire-bonding through a hole in the PCB [Courtesy CMS][308]

The relatively low channel number per wafer, with relatively large pads of $A = 0.5$ and 1 cm^2, is manageable and allows easy wire-bonding from the upper PCB to the pads via holes. All services, a real challenge, have to be routed in the very small gap of 2 mm (see Fig. 7.24c): low voltage and high voltage power in and data out, while the cooling is comfortably located inside the absorber.

To limit the current and power in the silicon sensors after an integrated luminosity of $\mathscr{L} = 3000$ fb^{-1}, the sensor temperature needs to be at $-30\,°C$ or lower during operation. Similar to the HL-Tracker bi-phase CO_2 will be used. Still the total sensor power is estimated to be ∼25 kW. In contrast to the Tracker, material budget is not a problem. While cooling contacts in the Tracker have to be minimized, in the HGC they will be implemented as 6 mm thick copper plates with embedded CO_2 cooling pipes using the full sensor surface as thermal contact. The electronics adds about another 100 kW of power thus a total of 125 kW has to be handled inside the cold EE+FH volume.[25]

As in the CMS Tracker, the levels of radiation varies by three orders of magnitude along radius but also along the z-coordinate, illustrated in Fig. 7.1 on p.292. Accordingly sensors with different active thicknesses of 300, 200 and 100 μm are being considered – the thinner the active volume the higher the radiation tolerance,

[25] Potentially the BH part will be in the cold volume too, to ease thermal interfaces.

rf. Sect. 2.2.4. Electromagnetic and hadron showers deposit much more charge than a MIP, the fact that the signal also scales with thickness is less of an issue for a calorimeter. As a matter of fact, the front-end electronics must have a very large dynamic range as about 5000 MIPS are expected for a 1 TeV electromagnetic shower compared to 50 MIPs at 10 GeV. A charge of 10 pC would correspond to 3000/4500/9000 MIPs in a 300/200/100 μm thick silicon sensor – for physics signals, signal charge thus signal/noise is therefore less of an issue, neither is the total noise. MIP sensitivity is a desirable feature for the HGC, but not a must to detect individually; it will be used mainly to calibrate the detector. Simulations show that MIP calibration can be achieved by combining information of several layers along a MIP 'track' or by implementing dedicated smaller calibration cells in the detector to follow evolution of signal with radiation. This calibration is especially important since, different to a tracker, the calorimeter must be more sensitive to charge/energy deposition measurement than 'simple' 3D local position. The combined multi-layer information will also be useable for muon-tagging.

Large pads together with a thin active volume equals to a large cell capacitance of 40 or 60 pF, compare to about 20 pF for the largest CMS double sensor module or about 5 pF for the largest Phase II Tracker sensor. Together with a planned fast shaping time of 10 to 20 ns, noise will be significant (rf. Sect. 1.5) especially for a standard charge amplifier. The HGC will use a charge amplifier/shaper plus a **T**ime-**o**ver-**T**hreshold ToT circuit. This allows to span the necessary huge dynamic range. After the DC-coupled preamplifier and shaper, a 10-bit ADC is used for measurement of small pulses, and saturates for pulses of ≥100 fC. For signals ≥80 fC the ToT comparator, starts and stops a TDC, thus providing a digitization measuring the magnitude of large signals. Very simplified, by counting the time-over-threshold of the signal it doesn't matter if the signal saturates in between. S/N values for MIPs (small pulse) are expected to be between 13.7 down to 1.7 depending on sensor thickness and radiation level. Here it becomes clear again, that the individual detector cells are not suited to detect MIPs. Sparsified/zero-suppressed data will be transmitted to the back-end electronics.

The large signals (large gain/signal - fast rise time - low "jitter", the main contributor to the time resolution σ_t) opens up to the possibility to achieve very fast timing information; a precision of about 20 to 50 ps for individual sensor cells is envisaged. The publication [71] gives a first impression of what is achievable. Tests show timing resolution of about $\sigma_t \sim 20$ ps for ~ 5 MIPs in 300 μm thick silicon or 20 MIPs with 120 μm thick sensors. $\sigma_t \sim$700 ps/(S/N) seems a reasonable approximation. Thus precision timing of 50 ps for $Q_{deposited} > 60$ fC seems possible given adequate chip functionality (Fig. 7.25).

In this context, next to the ToT circuit, the FE-ASIC will implement an ADC and **T**ime-**o**f-**A**rrival ToA mode. Timing information in the forward region is especially useful to disentangle primary vertex information derived from arrival time – important in the future high-pile-up environment. In principle the HGC will provide 4D coordinates plus energy information. Different to the tracker, the calorimeter also measures neutral particles (e.g. photons, neutrons, π°, etc.); the timing information here will allow to link these particles to the primary vertex.

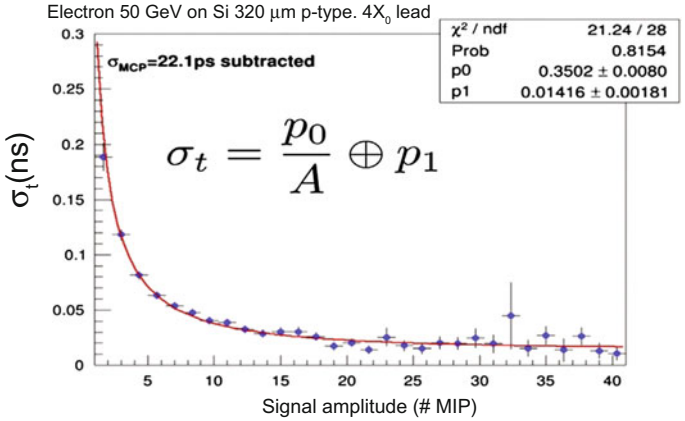


Fig. 7.25 Fast timing with silicon at large charge deposition. 300 μm thick silicon diodes have been tested in a test beam at CERN where 50 GeV electrons have been directed on a $4{\cdot}X_o$ lead absorber. $\sigma_t \sim 700$ ps/(S/N) seems a reasonable approximation according to [71]

Table 7.6 The table shows the similar and different challenges of the CMS-HGC and CMS -HL-Tracking system

	HGC	Tracker
Radiaton tolerance	$1 \cdot 10^{16}\, n_{1\mathrm{MeV}}/\mathrm{cm}^2$ – more neutrons	$1.6 \cdot 10^{16}\, n_{1\mathrm{MeV}}/\mathrm{cm}^2$ – more charged particles
Occupancy	mid-size pads	thin pitch strips and/or pixels
Channel capacitance	up to 60 pF	up to 5 pF
Dynamic range	1–5000 MIPs (shower)	1 MIP
Calibration	MIP calibration is difficult	no problem
Material budget	high mass density necessary	∼ largest challenge to keep mass low
Services	only 2 cm gap for routing	space and stringent mass constraints
Alignment	hundreds of micrometer	couple of micrometers
Data rate	similar challenge	
L1-trigger	similar challenge – large BE electronics system (combinatorics)	
Cooling	similar challenge CO_2	

A comparison of similarities and difference between HGC and the CMS HL-Tracker is given in Table 7.6.

As for the Tracker, due to bandwidth limitations, not all information from the HGC can be transmitted from the detector to the back-end electronics in the service cavern. Since the calorimeters are essential for the Level-1 trigger, part of the information must be summed up and congregated. 2×2 sensor cells from EE and FH will be summed up for every other sensitive plane and checked against a threshold; the

Table 7.7 Different Silicon sensors for different locations depending on the radiation environment (Technical Proposal version [308], final numbers might change)

Thickness (μm)	300	200	100
Maximum dose (Mrad)	3	20	100
Max fluence ($1\text{MeV}_{eq}/\text{cm}^2$)	$6 \cdot 10^{14}$	$2.5 \cdot 10^{15}$	$1 \cdot 10^{16}$
EE region	R > 120 cm	120 > R > 75 cm	R < 75cm
FH region	R > 100 cm	100 > R > 60 cm	R < 60cm
Si wafer area (m^2)	290	203	96
Cell size (cm^2)	1.05	1.05	0.53
Cell capacitance (pF)	40	60	60
Initial SN for MIP	13.7	7.0	3.5
S/N after $\mathscr{L}$=3000 fb^{-1}	6.5	2.7	1.7

congregated (also sparsified) info will be sent at the full rate of 40 MHz by the front-end electronics to the service cavern, where trigger primitives will be generated. The back-end electronics will then check for shower signatures taking the projection from the collision region into account. The complexity will be similar to the one of the CMS HL-Track Trigger.

HGC Sensors

The different sensor configurations and their expected signal/noise values before and after irradiation are listed in Table 7.7. Given the very large area, a large numbers of wafers is necessary and cost must be minimised. About 14000 EE plus 7500 FH modules will be installed where a module contains 2 sensors (rf. Figure 7.24) thus either 43000 6 in. or 21500 8 in. wafers are needed. The sensors are hexagonal shaped (see Figs. 7.26 and 7.27), which allows tiling as a honeycomb structure. This layout maximizes the surface-use of a circular wafer. On the other hand final offline reconstruction and especially fast trigger processing must be able to cope with this geometry and pattern matching, thus square cells (and square sensor) geometries had initially been evaluated as well. The studies proved that hexagonal cells or ok.

The sensors will be n-in-p as baseline (with p-in-n still an option) DC-coupled wafers with hexagonal pads of about A=1 or 0.5 cm^2. n-in-p would be even more cost effective since no extra processing for p^+-stops is necessary to guarantee pad isolation.

Active wafer thickness will be 300, 200 and 120 μm (rf. Table 7.7). While the first two feature 128 cells of size 1 cm^2, the 120 μm thick sensor has 256 cells of size 0.5 cm^2. The cell capacitance is large due to the big pad area plus the small thickness D of the sensor. To keep it under control, the pad size is halved for the 120 μm thin sensor. This limits the maximal capacitance to 60 pF. The individual cells are then directly wire-bonded to pads on a PCB and then routed to the readout chip located on the PCB.

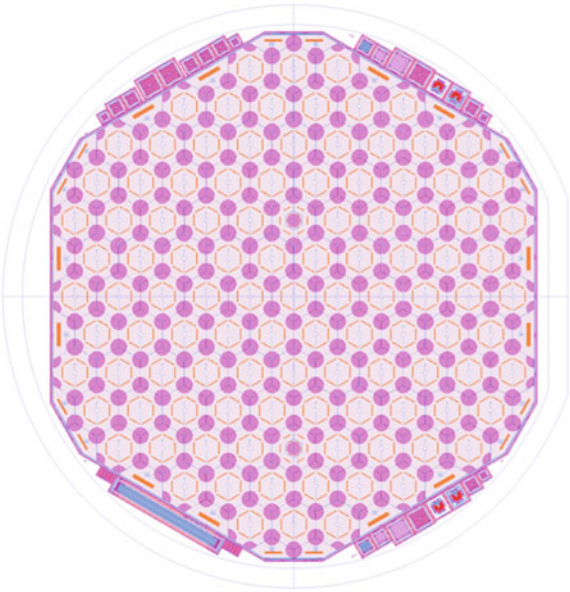


Fig. 7.26 The HGC hexagonal sensor design. The hexagonal shape allows to utilize 137 cm^2 of the wafer and the sensors can be tiled later to create a homogeneous area. The pads/cells are hexagonal too

On a 6 in. wafer, the sensor covers an area of 137 cm^2 compared to the nominal $10 \times 10\,cm^2$ of a square sensor thus a surface gain factor of 1.3 is achieved, reducing cost. The wafer layout and a photo of the first 6 in. prototype is shown in Figs. 7.26 and 7.27.

The active thickness will either be achieved by direct processing on physically thin sensors or by using deep-diffused wafers (rf. Sect. 1.9.3 on p.93 and Sect. 7.1.1). The use of high-volume commercial lines is a must and explorations are ongoing if 8 in. wafer production will be possible, while the baseline is 6 in. wafer processing. Deep-diffused material is only available on 6 in. and not for 8 in. for the moment, thus 8 in. processing would have to happen on thin wafers. One will see what the future holds.

Different to the tracker volume, the HGC radiation environment is strongly dominated by neutrons instead of charged particles, thus all defect engineering concepts based on compensation do not really hold (rf. for example Fig. 2.16 on p.155). The suppression of built up of negative space charge (acceptors) after charged hadron radiation in oxygen enriched silicon does not help (rf. Sect. 2.2.2 on p.154).

Still, radiation tolerance studies have shown adequate performance after the full expected fluence, but active thickness has to be adapted depending on the integrated fluence level thus location. The advantage of thin sensors for very high radiation levels has been discussed in Sect. 2.2.4. As long as the electric field strengths are high enough, charges are drifting fast enough to induce a sizeable signal in the readout electronics (remember, no charge amplification). Another advantage of pad detectors is, that both electrons and holes are contributing equally to the signal formation (rf. also Sect. 1.3.2). The sensors and modules must feature a very high voltage robust-

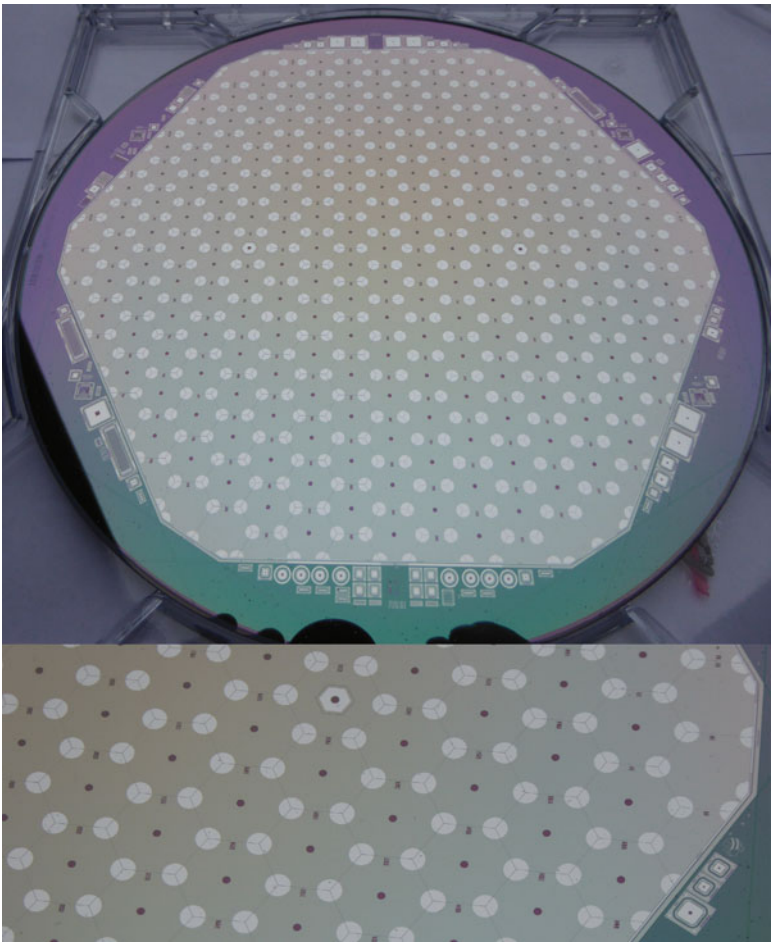


Fig. 7.27 Photo of an 8 in. HGC prototype wafer scaled down by a factor 2. The hexagonal structure optimizes the use of the wafer surface and can finally be tiled to cover a flat surface with minimal inactive areas. Below, a zoomed view of the right wafer corner is shown. All readout cells are hexagonal. The *circles* at the cell corners are openings in the passivation to allow wire-bonding. Two wafer cells have an extra smaller pad in the centre with smaller capacitance (thus lower noise) allowing for more precise calibration of MIP signals

ness, since sensor bias voltages up to 900 V after radiation are foreseen – a solvable challenge. Given early results of a radiation study, 100 μm might be just too thin and the optimum lies at a thickness of 120 μm. The Tracker is undergoing the same kind of optimization, especially since the deep-diffused material or thinned sensors allows to choose the active thickness quite precisely. In case the active/instrumented volume increases from $|\eta| \leq 3$ to $|\eta| \leq 4$, the highest integrated fluences would even reach

$\Phi_{eq} = 10^{17}\, n_{1\mathrm{MeV}}/\mathrm{cm}^2$. Then, even thinner sensors like 50 μm active thickness or 3D sensors would be taken into account.

The HGC is probably the most courageous endeavour due to its sheer size and complexity, but it will also lift the possibilities of particle detection to a whole new level.

Some Further Evolution

Usually, on the path from a technical proposal towards a Technical Design Report, several formerly open design choices are decided and new opportunities are exploited. Some HGC examples, expressing this kind of evolution, are briefly given here. Also the naming convention has been rationalized, FH+BH have been combined into EC-H or ECH **E**ndcap **C**alorimeter-**H**adron and EE is now named EC-E or ECE **E**ndcap **C**alorimeter-**E**lectromagnetic.

Interestingly, the hadron calorimeter absorber material has been changed from brass to stainless steel. Steel is much cheaper than brass and offers more engineering possibilities.

Initially mechanical alveolar structures, where wedge-shaped structures would have been slided-in, competed with more standard disk-structures, and finally the disk-concept was selected (stack of 30° or larger segmented cassettes and tungsten plates with spacers). Finally, the whole endcap calorimeter (EE, FH, BH) will reside in a single cold volume with a single thermal and gas/humidity barrier, while initially the BH could have stayed in a warm volume. This allows to have all feed-throughs at the back with more space available. The cold operation opens up several possibilities for the BH part, namely the instrumentation with silicon sensors at lower radius with higher radiation levels and restricting scintillators to lower irradiated volumes. The scintillators, in turn, can then use SiPMs directly on tile. Cold operation lowers the radiation induced dark noise of the SiPMs. The new considered layout is displayed in Fig. 7.28.

As mentioned earlier, the single module sensitivity is not really good enough for MIPs (with S/N down to ∼1.7) but combining information from several layers opens the door for muon-tagging (MIPS), besides the mentioned calibration. This is especially true for layers at larger z where the showers of other particles, for example pions, have already decayed. The new additional high precision silicon layers in the BH (at lower radius), therefore facilitate the muon-tagging significantly.

Most probably, 8 in. wafers will be used (now baseline), reducing the number of wafers significantly and increasing individual sensor surface largely. Given the increased sensor surface, single-sensor-modules are becoming attractive, being easier to manufacture and to handle with still a ‘reasonable’ total number of modules—ballpark: the use of 6 in. wafers would result in 40,000 *single*-sensor-modules or 20,000 *double*-sensor modules while the use of 8 in. wafers would reduce this to a reasonable number of 20,000 *single*-sensor-modules. Along the way, rectangular sensor cells instead of hexagonal ones had been discussed, ending up with

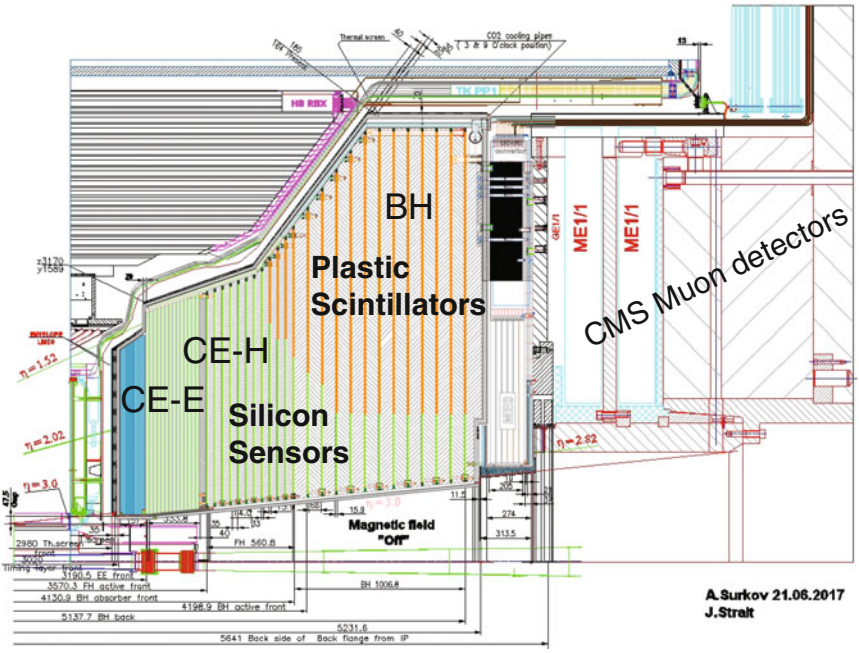


Fig. 7.28 The design layout of the HGC—version considered for the TDR, shown at the HGC Comprehensive Review June 2017. This layout, compared to Fig. 7.22 gives a nice impression of how such detectors evolve. The whole detector will reside in a cold volume with balanced instrumentation of plastic scintillator with SiPMs on tile (*orange*) at higher z and higher radius and silicon sensors (*green*) now also in the BH section at lower radius [Courtesy CMS]

the original hexagonal concept. Today 2017, no deep-diffused wafers exist in 8 in. The idea is to use physically thin FZ material for the D= 300 μm and 200 μm case and use EPI material for the D = 100 μm thin case.

Chapter 8
Continuing the Story: Detectors for a Future Linear Collider ILC or a Future Circular Collider FCC

Beyond the **H**igh **L**uminosity **L**arge **H**adron **C**ollider HL-LHC, described in the previous chapter, the two main future colliders, currently under discussion are the next potential e^+e^- machines, the **I**nternational **L**inear **C**ollider ILC [320], briefly the **C**ompact **LI**near **C**ollider CLIC and the **F**uture **C**ircular **C**ollider FCC. The requirements of both colliders and possible detectors are significantly different but also some synergies exist. Figure 8.1 informs about the increase in size of detector systems over the past 40 – 60 years, plus the envisioned ones for the next two decades.

The logarithmic plot behaves Moore-like[1] so far but is flattening out in the near future. The field is very active. R&D, assembly efforts, quality control plus the related logistics increased a lot in the last years and will even further in the future. To construct a detector with a couple of people over the summer-break is no longer possible. Also R&D efforts today are quite diverse as a result of the many new possibilities and needs for the next generation of tracking detectors. Whole R&D-*only* collaborations exist, e.g. RD50 [336]. In Fig. 8.1, it can be noticed that the Tracker detectors are not getting larger neither for the High Luminosity LHC nor for the future planned linear collider. The future high granularity silicon calorimeters will take over, although surface areas of around 400 m^2 for the FCC tracker are being discussed these days.

The future will tell us where we will go next. In the next two sections some ideas, plans and current R&D activities will be discussed.

Compact Linear Collider – CLIC

In competition to the ILC, another e^+e^- machine up to 3 TeV with a different acceleration concept is proposed, namely the **C**ompact **LI**near **C**ollider CLIC. The CLIC detector concept is, in many ways, similar to the one of the ILC, described in the next section. The main difference or better additional challenge is high beam induced

[1] According to Gordon E. Moore co-founder of Intel, the number of transistors placed on an integrated circuit increases exponentially, doubling approximately every 2 years. Published in 1965, Moore's law predicts the future of integrated circuits.

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_8

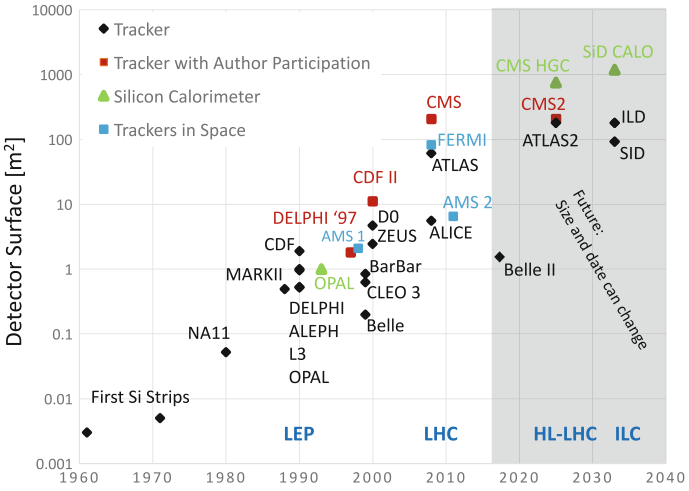


Fig. 8.1 Evolution of silicon detectors – area. In logarithmic scale the plot displays the increase in area for the last 40 – 60 years, the detectors approved and considered for the next decades. Plotting number of channels and power consumption would result in similar figures until the LHC era but will increase further for the future detectors

backgrounds (Beamstrahlung[2] due to the planned ultra-small bunch length/width (45 nm/1 nm/44 μm (σ_x / σ_y / σ_z)). These nanometer sized bunches are required to achieve high luminosity despite the relatively low repetition rate of 50 Hz (bunch trains). Typically only 1 hard interaction per bunch train is expected. The means to reduce this background is by time-tagging the physics events per strips/pixels better than 10 ns. The front-end readout therefore features the capability of **T**ime **o**f **A**rrival ToA and **T**ime **o**ver **T**hreshold ToT.

Sensor technologies for the vertex and tracking detector under discussion are

- Full monolithic HV-/HR-CMOS (rf. also Sect. 1.12.4, page 112).
- **C**apacitively **C**oupled **P**ixel **D**etector CCPD configuration (rf. also Sect. 1.12.4, page 112) glued to an ASIC
- **S**ilicon on **I**nsulator **S**oI technology with in-pixel signal processing (rf. also Sect. 1.12.3, page 109)
- Very thin planar sensors plus thin ASIC (50 μm sensor + 50 μm)

The collaboration is also working closely with the CALICE collaboration on a high granularity calorimeter. The reader is referred to [1, 152, 192, 275, 339] for more details.

[2]From beam + bremsstrahlung – electromagnetic beam-beam interaction.

8.1 A Silicon Tracker for the International Linear Collider – ILC

A linear e^+e^- collider in the TeV centre-of-mass range[3] is considered to be the next ultimate challenge – the **I**nternational **L**inear **C**ollider ILC. The Technical Design Reports, with all details, can be found here [320]. At the current time, a lot of technical aspects about possible detector concepts is under ongoing discussion, meaning despite the TDR, choices can change depending on the ongoing developments and the final installation date. The one clear fact is as the instrument will be built for precision physics, all detector parameters are driven by physics channels need and resolution. The vertex detector must be able to tag b, c, τ decays and some people think about s quarks. The detector will be operated trigger-less, i.e. all collisions will be recorded. Detailed but very early proceedings about the vertex detector ideas are presented in [50, 218, 355], a first discussion about strip implementation in [86]. By comparing these with [271] and [320], evolution of design and detector solutions become clear while the original constraints and concepts are mostly still valid. The concept of "**p**article **f**low PF algorithm" is followed throughout the detector design, **every** particle must be reconstructed for optimum jet resolution. Therefore the tracker must be able to reconstruct **all** charged particles with high momentum resolution.

There were still four concepts in 2008, namely GLD (**G**lobal **L**arge **D**etector); LDC (**L**arge **D**etector **C**oncept); SiD (**Si**licon **D**etector) and the **4th Concept**. In 2009, the GLD and LDC merged into a single detector concept: the **I**nternational **L**arge **D**etector ILD [321].

Today (2017), there are two validated concepts and which have been selected by the International Detector Advisory Group IDAG in 2009: the ILD (vertex+Si strips+TPC[4]) and SiD (vertex+silicon strips).

SiD "**Si**licon **D**etector" will be an all-silicon detector similar to CMS consisting of an inner high precision vertex detector plus a tracker at higher radius. The vertex detector with 5 pixel barrel layers ranging from $R = 14$ to 60 mm plus 4 large (small stereo angle) forward disks plus pixelated 3 forward disks. The outer tracker then equips 5 barrel layers ranging from $R = 22$ to 122 cm plus 4 slightly tilted disks in forward direction.

The ILD will consist of an inner vertex detector (**S**ilicon **I**nner **T**racker SIT) with three high precision double-ladder layers $R = 16$ to 60 cm, a large TPC followed by an envelope Tracker at $R = 180$ cm as precision link towards the calorimeter (**S**ilicon **E**xternal **T**racker SET) in the central barrel as well as in the forward direction, plus a silicon forward detector (**f**orward **t**racker **d**etector FTD). The high-radius envelop[5] Tracker will consist of small pitch (50 μm) silicon strip sensors implementing the **P**itch **A**dapter PA as first or second metal routing in the sensor itself. Also the TPC

[3] The accelerator is planned to have a centre-of-mass energy of 500 GeV with the option to upgrade later to 1 TeV.

[4] **T**ime **P**rojection **C**hamber TPC.

[5] Also called SET or Intermediate Silicon Layers.

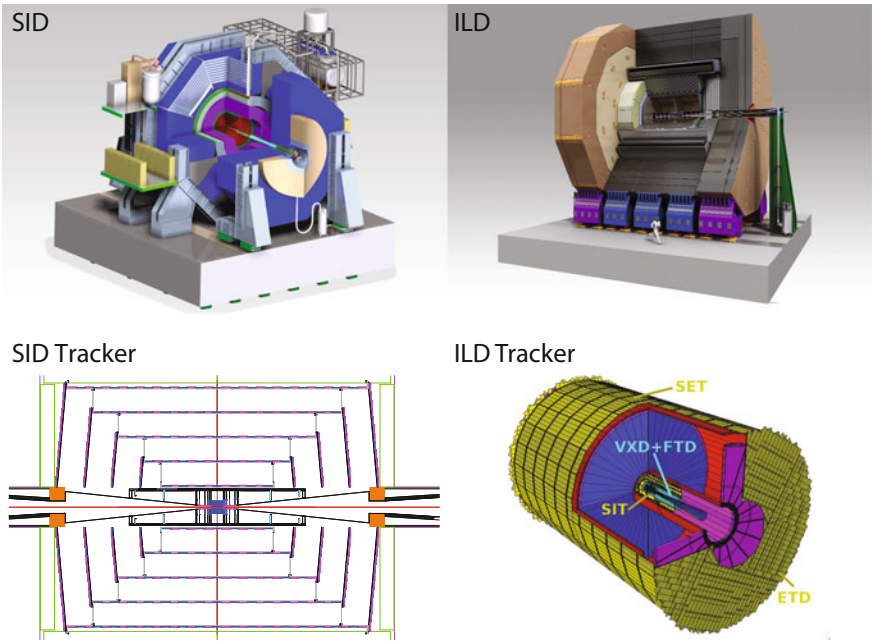


Fig. 8.2 The schematics of the two ILC detector concepts and their Trackers from [320]. The usual onion shell concept is being followed for both SiD and ILD: vertex – tracking – ECAL – HCAL – solenoid – muon chambers, for the barrel part as well as for the endcaps. The main difference between the detector concepts is the large TPC of the ILD detector. Interestingly, due to the large radius of the SET, the total silicon surface is even larger for the ILD

readout at the volume end will possibly be realized with silicon sensors or high-resolution gaseous chambers.

The solenoid due to its high mass will reside outside the Tracker and also outside the calorimeters for both concepts (5 T for SiD and 3.5 T for ILD), similar as for CMS.

Design studies and small prototypes exist for a silicon tungsten Si – W calorimeter[6] with $5 \cdot 5\,\text{mm}^2$ cells and even a digital "tera-pixel" ECAL with $50 \cdot 50\,\mu\text{m}^2$ MAPS readout.

The schematics of SID and ILD are displayed in Fig. 8.2.

In this section the current vertex detector requirements are discussed plus some promising sensor technology candidates. The future will tell which technology will be chosen.

[6] Adapted concept will be used in CMS rf. Sect. 7.2.

The requirements are listed below:

- space point resolution of 3 – 4 μm or better for
  - 20·20 μm^2 (or 25·25 μm^2) pixels
- impact parameter resolution in Rz and $R\phi$:
  - $\sigma_{d_0} = (5\,\mu m)^2 + (10\,\mu m/(p_T \sin^{3/2}\Theta))^2$
- a two-track resolution of 40 μm or better
- transverse momentum resolution: $\Delta(1/p_T) = 5 \cdot 10^{-5}/\text{GeV}$
- ultra-low material budget 0.1% X_0 radiation length per layer
  - low power consumption
  - gas flow cooling, crucial to allow low mass construction
  - very light structural elements, e.g. silicon carbon foams
- hermeticity with special focus on forward geometry
  - full coverage $|\cos\Theta| < 0.98$
- operation in magnetic field up to 5 T
- triggerless operation – particles come in bunch trains with a large gap in between
  - fast electronics
  - in situ storage
  - high granularity
  - fast sensor
- moderate radiation tolerance

Unlike a circular electron–positron collider with a constant time interval of beam crossings the ILC will operate with particle bunches organized in trains. In the planned configuration as of today [30] there will be five trains with 1312 bunches per second. The bunches are separated by 554 ns (baseline) summing to a total duration of 1 and 199 ms quiet time. Therefore fast electronics operating in power pulsed mode are necessary, especially with the need to consume very low power to run without active fluid cooling. The triggerless ILC operation together with the dense particle bunches and long trains poses the following requirements on the vertex detector: Detector response must be fast, occupancy must be low to avoid high pile-up and detectors must be either read out very fast (e.g. 20 times per train) or signals must be storable in the detector/readout cells. Without a trigger, all signals are read out, therefore sparsification is ultimately needed as well as a pixelated geometry for the vertex detector.

The possible candidates having been studied were numerous and included (1) **C**harge-**C**oupled **D**evices CCDs, CPCCD **C**olumn **P**arallel **CCD**s, (2) **M**onolithic **A**ctive **P**ixels **S**ensor MAPS based on CMOS technology, (3) **DE**pleted **P** channel **F**ield **E**ffect **T**ransistor DEPFETs, (4) **S**ilicon **o**n **I**nsulator SoI, (5) **I**mage **S**ensor with **I**n situ **S**torage (or **I**n-situ **S**torage **I**mage **S**ensor) ISIS,

(6) **H**ybrid **A**ctive **P**ixel **S**ensors HAPS and advanced 3D sensor-electronics integration concepts. The basic technologies are explained in Sect. 1.12 while the devices for the ILC come in several flavours with some specific implementations and also some technology combinations. Standard CCDs as used in digital cameras are not fast enough, proposed column parallel readout CPCCD helps or **S**hort **C**olumn **C**harge-**C**oupled **D**evice SCCCD, where a CCD layer and a CMOS readout layer is bump bonded together. Chronopixels are CMOS sensors, with the capability to store the bunch ID (time). ISIS sensors combine CCD and active pixel technology, a CCD-like storage cell together with CMOS readout implemented. Also **F**lexible **A**ctive **P**ixels FAPs integrate storage cells in the traditional MAP cells. **F**ine **P**ixel **CCD**s FPCCDs are under discussion to decrease occupancy. To summarize, the different varieties of CCDs, DEPFET, MAPS and SOI are designed to be read out every 50 μs, while ISIS and FAPS store signal in cell memory and will be read out in the 199 ms between trains. Also FPCCDs and chronopixels are designed for in-between train readout.

Given the recent success of **H**igh **V**oltage HV-CMOS and **H**igh **R**esistivity HR-CMOS, CMOS devices will certainly deploy a certain 'low' high voltage to deplete (or at least partially deplete) the thin sensors. In my personal humble opinion, the full monolithic CMOS concept with some 'low' high voltage would be the ideal candidate (rf. also Sect. 1.12.4).

In addition to all the above-discussed technology choices, the requirement to achieve a ultra-low mass tracker demands thin sensors. Methods are under investigation to either have thin epitaxial sensors like CCD and to thin the other devices. Edgeless processing is under investigation to allow near-adjacent sensor placement instead of traditional staggering; the particle flow algorithm would strongly suffer from non-instrumented regions.

While the above mentioned HV-CMOS technology can be considered baseline for the more outer tracking layers, the inner vertex layers are targeting an even more ambitious technology. New 3D integration technologies are under investigation where two or more thinned layers of semiconductor devices are interconnected to form a monolithic "chip" (see also Sect. 1.11.1, page 106). In early design concepts, the real sensor together with analogue, digital and even optical electronics are vertically integrated in a monolithic circuit. This concept has many obvious advantages and early prototypes are very promising.

The above-mentioned link layers (FTD) towards the TPC and especially the layer (SET) surrounding it and also the full silicon tracker (SiD tracker) might still be realized with the strip technology of today, but also here thinning of sensors is under discussion. Also long ladders had been investigated with "long" strips to stay within a reasonable number of readout channels. Avoiding standard hybrids at the end of modules, second metal routings were discussed with a bump bonding field even on strip sensors and chips with 512 or even 1024 channels were in planning in 2008 [86] and are now proven to work [320]. With the current advances in HV-/HR-CMOS, this might change.

The rich number of valid technological choices proves the liveliness of the silicon sensor development in the last years.

8.2 The Next Big Future Circular Collider – FCC

Preliminary studies have been started to discuss a 80 – 100 km ring accelerator with about 100 TeV centre-of-mass energy. 16 Tesla dipole magnets would be needed and technology seems in reach. The International **F**uture **C**ircular **C**ollider FCC collaboration consider CERN as hostlab for first studies but also China is studying/evaluating a similar endeavour (CepC/SppC study CAS-IHEP with 50 or 100 km circumference). The main emphasis defining the infrastructure requirements is a 100 TeV hadron-hadron collider (FCC-hh or HE-LHC), an e^+e^- collider (FCC-ee) in the same tunnel is studied as a potential intermediate step, a proton electron collider option (FCC-he) is studied as well. Also the instantaneous and integrated luminosities for the hadron collider are supposed to be very high (hypothesis) – peak baseline $\mathcal{L} = 5 \cdot 10^{34}$ and ultimate $30 \cdot 10^{34}$ as a later second step. This would lead to an integrated luminosity of $\mathcal{L} = 250$ or 1000 fb^{-1} per year summing in 10 years to $\mathcal{L} = 2.5$ ab^{-1} and 15 ab^{-1} in the next 15 years; thus totalling to about 20 ab^{-1}. In this case, detectors should be designed to last about $\mathcal{L} = 30$ ab^{-1}. The ultimate peak luminosities correspond to about a pile-up of PU = 1000 with about 10 vertices per millimetre at 25 ns bunch crossing or PU = 200 for 5 ns bunch crossings.

This section is not meant to describe a specific detector design (FCC-hh case) but to collect the basic ingredients to do tracking of particles with transverse momentums in the tens of TeV range. Most numbers come from [256, 334].

To exploit such a machine, the detectors should feature

- a tracking resolution of about 15% at 10 TeV
- a precision tracking and calorimetry coverage up to $|\eta| = 4$
- tracking and calorimetry for jets up to $|\eta| = 6$
- b-tagging
- timing for pile-up mitigation

The need for high precision tracking in the forward region demands either a very long solenoid or a dipole magnet in the forward region. Since the magnet coils will be very large, an iron return yoke will be very costly and heavy, so geometries with a shielding coil or an unshielded solenoid are investigated. With such high integrated luminosities the requirements for the radiation tolerance is incredibly high. For $\mathcal{L} = 3000$ fb^{-1} = 3 ab^{-1} at the HL-LHC we expect about $2 \cdot 10^{16}$ $n_{1\text{MeV}}/\text{cm}^2$ for the innermost layer at about r = 3 cm with the increased inelastic cross-section at 100 TeV this translates to roughly about $4 \cdot 10^{16}$ $n_{1\text{MeV}}/\text{cm}^2$ for 3 ab^{-1} at the FCC or $4 \cdot 10^{17}$ $n_{1\text{MeV}}/\text{cm}^2$ for the full lifetime corresponding to 30 ab^{-1}. Decreasing radius a bit further and/or adding a bit of error margin one arrives at the $1 \cdot 10^{18}$ $n_{1\text{MeV}}/\text{cm}^2$ or 100 GRad level. We do not 'yet' have a radiation tolerant technology for these levels. This probably means we cannot instrument as low in radius as we are doing for the LHC but still, even the outermost Tracker radii will suffer about $1 \cdot 10^{16}$ $n_{1\text{MeV}}/\text{cm}^2$.

Now, how can a 15% p_T resolution at 10 TeV be achieved? The FCC community is using TKLayout for first design studies (rf. Sect. 7.1 on page 293). The Glückstern

formula (rf. [118]) plus the multiple scattering term gives the basic answer – see formula 8.1.

$$\frac{\sigma p_T}{p_T} = \frac{\sigma p_T}{p_T}_{|res} \otimes \frac{\sigma p_T}{p_T}_{|MS} = \sigma_x \frac{p_T}{0.3B \cdot L^2}\sqrt{\frac{720}{N+4}} \otimes \frac{0.0136}{0.3B \cdot L}\sqrt{\frac{x}{X_0}} \quad (8.1)$$

with L lever arm $\sim$ full detector radius; B magnetic field in Tesla; N number of track layers,[7] σ_x point resolution, x/X_0 fractions of radiation lengths.

Clearly large $L \cdot B$ and small σ_x seems key, since it improves resolution. Having more layers helps but impacts adversely the multiple scattering. Unfortunately a high B at large radius is very difficult (high field energy). One of the layouts being discussed would even feature 2 solenoids (one around Tracker and calorimeters and a shielding one around the muon detectors) plus dipoles in the forward direction.

For example a Tracker with r = 2.5 m, a B-field of 6 T, 16 layers and 3% of a radiation length per layer would need a point resolution σ_x of about 20 μm. Decreasing the radius by a factor of two (CMS design) would mean decreasing σ_x and x/X_0 per layer by a factor of four. Of course increasing the B-field by a factor four would also do the trick, but this is probably technically not feasible. A σ_x of about 5 μm or even lower is however possible and given the very high pile-up, a full pixel detector is anyhow the only viable solution. x/X_0 of 0.75% seems doable as well, especially if monolithic technology would be used. For such a precision also the alignment needs to be perfect to even exploit the point resolution. For muons, the precision will be defined by the muon chambers with much larger L.

Tracking in very forward direction requires disks very very close to the beam pipe and for b-tagging also as close as possible to the interaction region, where barrel layers already occupy the space. LHCb, as forward spectrometer tracks up to $|\eta| = 5$. The CMS Upgrade will manage up to $|\eta| = 4$ but without b-tagging capability and without good momentum resolution due to weak 'effective' B-field. Can one realise a tilted layout to cover both needs?

To cope with the high pile-up probably also the calorimeter will be done the CALICE/HGC way, as fine granular imaging calorimeter with silicon-pad or even silicon-pixel detectors in-between absorber plates.

The amount of silicon surface of tracker plus calorimeter will be huge!

The super-stringent requirements require probably monolithic pixelated sensors having a very low material budget, a very fine granularity and excellent point resolution. $20 \cdot 20$ μm^2 could be used, a factor 2.5 smaller than the envisaged CMS and ATLAS High Luminosity upgrade and due to bump bonding size constraints probably not possible in HAPS technology. HV-CMOS (rf. Sect. 1.12.4) are probably the best candidates but they have to become faster and significantly more radiation hard. All in all, silicon sensor technology will be again the key to success, but still has to be developed; most probably with very good timing resolution.

Needless to say that challenges on other detectors and especially the accelerator are of equal or maybe even higher difficulty.

[7] A factor we neglected in earlier chapters.

Final detector concepts will be defined by particle physics needs but it will be very challenging.

I like to cite a speaker from a recent (2017) conference talking about 500 million strips, ∼10 billion macro-pixel, ∼5.5 billion micro-pixels and the incredible high radiation levels:

Yes, we know this is crazy.

This will be fun!

Chapter 9
Conclusion and Outlook

Reviewing history, collecting experience, looking into the future; an experimentalist in the field of particle and detector physics has to live on the waves of new technologies. As a matter of fact there will always be a new detector with new R&D challenges. The last 30 years of semiconductor devices brought significant progress together with significant science done. Devices have grown 4 orders of magnitude in size, orders of magnitude in readout speed and today, a compensation for 4 orders of magnitude in radiation damage is possible. For larger and larger projects the experts in the field have to always remember what they have learned while utilizing their natural curiosity to invent new lanes. The sensor activity has obvious spin-offs like medical applications and material analysis but we have to sell our science for its content and provide high energy particle physics with its ever-growing needs. This field has a clear natural synergy between particle physics, astroparticle physics, applied physics and condensed matter physics. Silicon detectors with their c-, b- and τ-tagging capability opened a new gate to physics. No other device is able to label jets with their initial content. Candidates for takeover: diamond detectors are not yet mature, neither in technological realization nor in cost and will always suffer from the advanced developments of processing in industry. Silicon exists in abundance and is currently the best-known material in the world.

Starting from NA11 with 6 sensors, DELPHI set the next benchmark with 888 detectors followed by CDF with 1752 large area sensors and today CMS leads the record with 24,328 sensors from 6 in. wafers, where will we go next? A rich number of technology choices for the next two large HEP instruments – the HL-LHC and ILC – exists.

Standard passive sensors plus custom readout ASICs will remain the working horse for the HL-LHC, while the baton for the ILC and FCC might well be given to monolithic HV-CMOS sensors.

Independent of which technology will win the race, the field will remain interesting and versatile. And, with ideas for a possible **F**uture **C**ircular **C**ollider FCC,

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3_9

completely new challenges are on the horizon – Ultra-radiation tolerances, granularity, size and speed.

I like to emphasize the past, present and future of particle tracking detectors with the words of two very dear colleagues of mine:
Joe Incandela:

> I have been very fortunate to be involved in many great silicon detector projects over the years. The CDF projects were all very successful, and provided lots of great physics but they were not as easy as they looked, even when they didn't look that easy. There is no routine way of doing things with large scale silicon detectors. Problems arise at every stage of every project. This has also been true of CMS. The silicon groups I've led and collaborated with over the past nearly 2 decades have included the most dedicated, talented, and hard working people I have ever known and this is the real reason why we were successful. It's also why the experience has been so memorable.

Manfred Krammer:

> The success and advantage in particle physics relies on the rich diversity on detector technologies and their continuous development.

Silicon sensors were, are and will be my favourite field of interest and work.
Frank Hartmann

Appendix A
Glossary

Some Physics Constants and Definitions

Colliders and Detectors

LEP: **L**arge **E**lectron **P**ositron Collider at CERN, operating at CERN from 1988 to 2000. It hosted

- **ALEPH A**pparatus for **LE**P **Ph**ysics
- **DELPHI DE**tector with **L**epton, **P**hoton and **H**adron **I**dentification, one the four multipurpose detectors at LEP
- **L3 L**EP **3**
- **OPAL O**mni **P**urpose **A**pparatus at **L**EP experiments

TEVATRON: The proton – antiproton Collider at Fermilab near Chicago, operating since 1983, hosting the CDF and D0 detector

- **CDF**: **C**ollider **D**etector at **F**ermilab, one of the two multipurpose detectors at the TEVATRON
- **D0**: Multipurpose detector situated at the interaction D0 at the TEVATRON

LHC: **L**arge **H**adron **C**ollider, proton–proton collider at CERN in Geneva, started 2008 and its planned successor
HL-LHC: **H**igh **L**uminosity - **L**arge **H**adron **C**ollider, proton–proton collider at CERN in Geneva, starting 2026

- **ALICE**: **A L**arge **I**on **C**ollider **E**xperiment at the LHC at CERN in Geneva, specializing in heavy ion physics [330]
- **ATLAS**: **A T**oroidal **L**HC **A**pparatu**S** [286] a multipurpose detector at the LHC
- **CMS**: **C**ompact **M**uon **S**olenoid [292] a multipurpose detector detector at the LHC
- **LHCb**: One of the four LHC detectors, specialized on *b* physics and *CP* violation [323].

ILC: **I**nternational **L**inear **C**ollider, a high energy linear collider in design and planning phase with the two detectors:

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3

- **SiD Si**licon **D**etector at the future ILC
- **ILD I**nternational **L**arge **D**etector at the future ILC

CLIC: **C**ompact **LI**near **C**ollider
FCC: **F**uture **C**ircular **C**ollider

Sub Detectors

ISL: **I**ntermediate **S**ilicon **L**ayers; the link silicon detector in-between main vertex detector SVX and outer drift chamber COT at the CDF detector
MVD: **M**icro **V**ertex **D**etector; innermost silicon strip and pixel vertex detector of DELPHI
SVT: **S**ilicon **V**ertex **T**racker, level two hardware trigger on displaced vertices in the CDF silicon detector
SVX: **S**ilicon **V**erte**x** Detector of the CDF detector at Fermilab. Historically, starting with four single-sided layers upgraded to five layer double-sided, then named **SVX II**: Also the name of the ASIC readout chip used (SVX1 to SVX4)
SST CMS **S**ilicon **S**trip **T**racker SST with CMS contains the following sub detectors: **TEC T**racker **E**nd**c**ap; **TIB T**racker **I**nner **B**arrel; **TID T**racker **I**nner **D**isk; and **TOB T**racker **O**uter **B**arrel
CMS Pixel detector with **B**arrel **Pix**el BPIX and the **F**orward **Pix**el FPIX, all in two half shells names inner (near LHC center) or outer (far form LHC center) and divided in plus minus.
CMS Tracker Phase II will contain the following sub detectors: *Outer Tracker*: **TEDD T**racker **E**ndap **D**ouble **D**isk; **TB2S T**racker **B**arrel with **2S**-modules; **TBPS T**racker **B**arrel with **PS**-modules
Inner Tracker: **TBPX T**racker **B**arrel **Pix**el Detector; **TFPX T**racker **F**orward **Pix**el Detector; **TEPX T**racker **E**ndcap **Pix**el Detector or
HGC: **H**igh **G**ranularity **C**alorimeter, planned CMS forward calorimeter based on silicon pad detectors allowing a fine granularity. It contains the **EE** Endcap Electromagnetic part, the **FH F**orward **H**adron Calorimeter and the **BH B**acking **H**adron Calorimeter
IBL Insertable **B**-**L**ayer innermost pixel layer of ATLAS

Detector Types

CCD: **C**harge **C**oupling **D**evice; another silicon pixelated device, not discussed in this book but very popular in former digital cameras
SoI: **S**ilicon **o**n **I**nsulator
LGAD: **L**ow **G**ain **A**mplifier **D**etector
DC: **D**rift **C**hamber: Gaseous tracking detector; wires in gas, where wire position plus drift time of electrons and ions from the ionising traversing particle define particle position, or **DT D**rift **T**ube

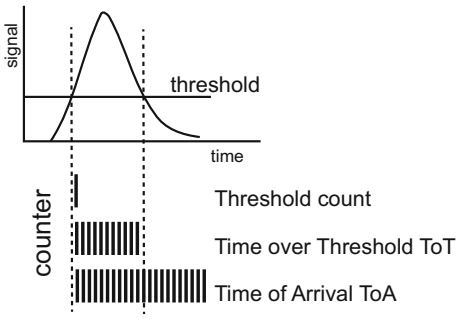
GEM: **G**as **E**lectron **M**ultiplier
ECAL: **E**lectronic **Cal**orimeter: Measurement device to measure energy of electrons and photons
HCAL: **H**adronic **Cal**orimeter: Measurement device to measure energy of hadrons, mesons and baryons
RICH: **R**ing **I**maging **Če**renkov Counter (or detector)
Si-PM Silicon-**P**hoto-**M**ultiplier
TPC: **T**ime **P**rojection **C**hamber, a gaseous tracking detector, utilizing time/drift information in a large volume for the z-coordinate
MSGC: **M**icro**S**trip **G**as **C**hamber a Micro-pattern gas detector gaseous. A fine resolution tracking gas detector.

Abbreviations for Electronic Chips

APV: **A**nalog **P**ipeline **V**oltage. Readout chip in use in the CMS experiment
CBC: **C**MS **B**inary **C**hip for CMS at HL-LHC.
MPA Macro-**P**ixel-**A**SIC for CMS at HL-LHC.
SSA Short **S**trip **A**SIC for CMS at HL-LHC.
Readout chips for the HL-LHC-CMS experiment
MX: **M**icrople**x** used in the LEP experiments, especially in DELPHI. Precursor of the TRIPLEX chip
SVX1–3: **S**ilicon **V**erte**x**: chips used in TEVATRON experiments
VLSI: **V**ery **L**arge **S**cale **I**ntegration stands for the process of creating **I**ntegrated **C**ircuits (ICs) by combining thousands of transistor-based circuits into a single chip
ASIC: **A**pplication-**S**pecific **I**ntegrated **C**ircuit, a Custom Chip; every front-end chip in HEP is an ASIC
FPGA: A **F**ield-**P**rogrammable **G**ate **A**rray is an integrated circuit designed to be configured after manufacturing. FPGAs are used in the back-end electronic in HEP, e.g. to sparsify data or in the high level trigger (HLT). FPGA are too bulky and not radiation tolerant enough to be used in the detector as front-end electronics
CMS Pixel chips: PSI46 - analogue chip of RUN I; PSIDIG - digital chip with ADC for Layer 2-4 of the CMS Pixel detector 2017 onwards; PROC600 similar to PSIDIG but with higher rate capability (600 MHz/cm^2) for Layer 1

Readout Modes See Also Figure Below

Charge sampling: The charge of the signal is being integrated
ToT: **T**ime **o**ver **T**hreshold: Circuit counts (clock cycles) the time while the signal is above threshold which is proportional to the signal itself. The threshold value is normally configurable per channel
Binary Threshold Count: Counts one when the signal passes the threshold
ToA: **T**ime **o**f **A**rrival: Starts counting (electronic clock cycles) when the signal passes the threshold until the readout cycle is over. The count is then proportional to the arrival time. This can be very precise



Abbreviations

BB Bump **B**onding - often called flip chip bonding or in case of industrial large pitch bonding **C4 – C**ontrolled **C**ollapse **C**hip **C**onnection – CCCC
CCE: **C**harge **C**ollection **E**fficiency
CMOS Complementary **M**etal-**O**xide **S**emiconductor
CV & IV: Capacitance or current vs. voltage characteristic – a measurement scan of the global silicon properties
CVD: **C**hemical **V**apour **D**eposition
DAQ: **D**ata **Acq**uisition
DRIE Deep **R**eactive **I**on **E**tching
ENC Equivalent **N**oise **C**harge
FEC: **F**ront **E**nd **C**controller – off-detector control board
FED: **F**ront **E**nd **D**river – off-detector readout board
DTC: **D**ata **T**rigger **C**ontrol Board
FZ Float **Z**one Crystal technique - wafer type
mCz magnetic **Cz**ochralski - wafer type
GCD Gate **C**ontrolled **D**iodes
HEP: **H**igh **E**nergy **P**hysics – Collider Physics
HLT: **H**igh **L**evel **T**rigger
LCR: Short for a **L**, **C** and **R** metre
MIP Minimum **I**onizing **P**article
MOS: **M**etal **O**xide **S**emiconductor, in the sensor case mainly "aluminium-SiO_2-silicon"
b-**tag**: The identification of *b* quarks or heavy quarks in general
τ-**tag**: The identification of a τ lepton

SCSI space **c**harge **s**ign **i**nversion
SCR space **c**harge **r**egion
UBM Under **B**ump **M**etallization

Measurement Techniques

DLTS: **D**eep **L**evel **T**ransient **S**pectroscopy
come in several flavours, mainly C-DLTS (capacitance) and I-DLTS (current)

TSC: **T**hermally **S**timulated **C**urrents
TCT: **T**ransient **C**urrent **T**echnique; also **e-TCT e**dge-TCT or **TPA-TCT T**wo **P**hoton **A**bsorption TCT
CV/IV: Capacitance/Current- Voltage measurements
SIMS Secondary **I**on **M**ass **S**pectrometry

Constants and Formulas

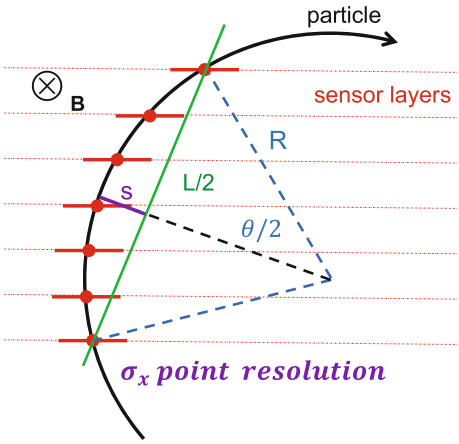
Symbol	Definition	Units or value
c	Speed of light	$2.997924580 \cdot 10^{8}$ ms^{-1}
h	Planck constant	$6.62606896 \cdot 10^{-34}$ kgm^2.s^{-1}
μ_0	Permeability of vacuum	$12.566370614 \cdot 10^{-7}$ kgms^{-2}A^{-2}
ε_0	Permittivity of vacuum	$08.854187817 \cdot 10^{-12}$ kg^{-1}m^{-3}s^{4}A^{-2}
N_A	Avogadro's number	$6.0221415 \cdot 10^{23}$
$m_e c^2$	Electron Mass $\times c^2$	0.511 MeV
r_e	Classical electron radius	2.818 fm
ε_0	Dielectric constant	$8.85 \cdot 10^{-14}$ Fcm^{-1}
k_B	Boltzmann constant	$1.38065 \cdot 10^{-23}$ JK^{-1}
e	Electron charge	1.602×10^{-19} C
β		$\frac{v}{c}$
γ		$\frac{1}{\sqrt{1-\beta^2}}$

V_{FD}: Full depletion voltage, the voltage, where the sensor is just depleted from charge carriers, visible in a kink in the CV characteristic
V_{bias}: The operating voltage to fully deplete the sensors, $V_{\text{bias}} > V_{FD}$
Pseudo rapidity: $\eta = -\ln(\tan \Theta/2)$ replaces the azimuthal angle Θ and is invariant versus Lorentz addition. In hadron colliders, e.g. CDF and CMS, the z-coordinate of the primary vertex varies. Therefore, the quantity η is used in calculations and also to define and construct the detector elements.

Track, Transverse Momentum Resolution and Impact Parameter Resolution

Track: A helix is a 5D object, defined by two positions, two angles and its curvature. A track state can be represented as a point in 5D linear space. To complete, the track is a measured (fitted) object, and has uncertainties (errors) on it's parameters. A track state is fully described by 5 parameters and a 5×5 symmetric error matrix.
Transverse Momentum Resolution The momentum resolution of a moving charged particle in a B field is given by its curvature path. With $s = L^2/8R$ and $B \cdot R = p/q$ one gets the momentum resolution as $\frac{\Delta p}{p} \approx \frac{\Delta s[\mu m]}{(L[cm])^2 B[T]} p[GeV]$ – simple approximation. This immediately tells us that (a) intrinsic position resolution has to be good

to resolve s and that (b) the B field strength gives a linear improvement, while (c) a larger lever arm improves momentum resolution quadratically. With increasing p_T the resolution gets worse again and with an error of 100% the particle charge cannot anymore be identified.



The full $\frac{\sigma p_T}{p_T}$ description: The Glückstern formula (rf. [118]) plus the multiple scattering term.

$$\frac{\sigma p_T}{p_T} = \frac{\sigma p_T}{p_T}_{|res} \otimes \frac{\sigma p_T}{p_T}_{|MS} = \sigma_x \frac{p_T}{0.3B \cdot L^2}\sqrt{\frac{720}{N+4}} \otimes \frac{0.0136}{0.3B \cdot L}\sqrt{\frac{x}{X_0}} \quad \text{(A.1)}$$

with L lever arm $\sim$ full detector radius; B magnetic field in Tesla; N number of track layers, σ_x point resolution, x/X_0 fractions of radiation lengths.

Impact Parameter Resolution: How to design a tracking detector to achieve good impact parameter resolution is illustrated in the following formula:

$$\sigma_{d_0}^2 = \sigma_{MS}^2 + \sigma_{geom}^2 \text{ with } \sigma_{geom}^2 = \left(\frac{\sigma_1 r_2}{r_2 - r_1}\right)^2 + \left(\frac{\sigma_2 r_1}{r_2 - r_1}\right)^2 \text{ and}$$

$$\sigma_{MS}^2 = \sum_{j=1}^{n_{scatt}} (R_j \Delta\Theta_j)^2 \quad \text{(A.2)}$$

The resulting design parameters to optimize are

- small r_1
- large r_2
- good space point resolution σ_1 and σ_2
- low material budget to minimize σ_{MS}^2

Appendix B
Some Additional Figures

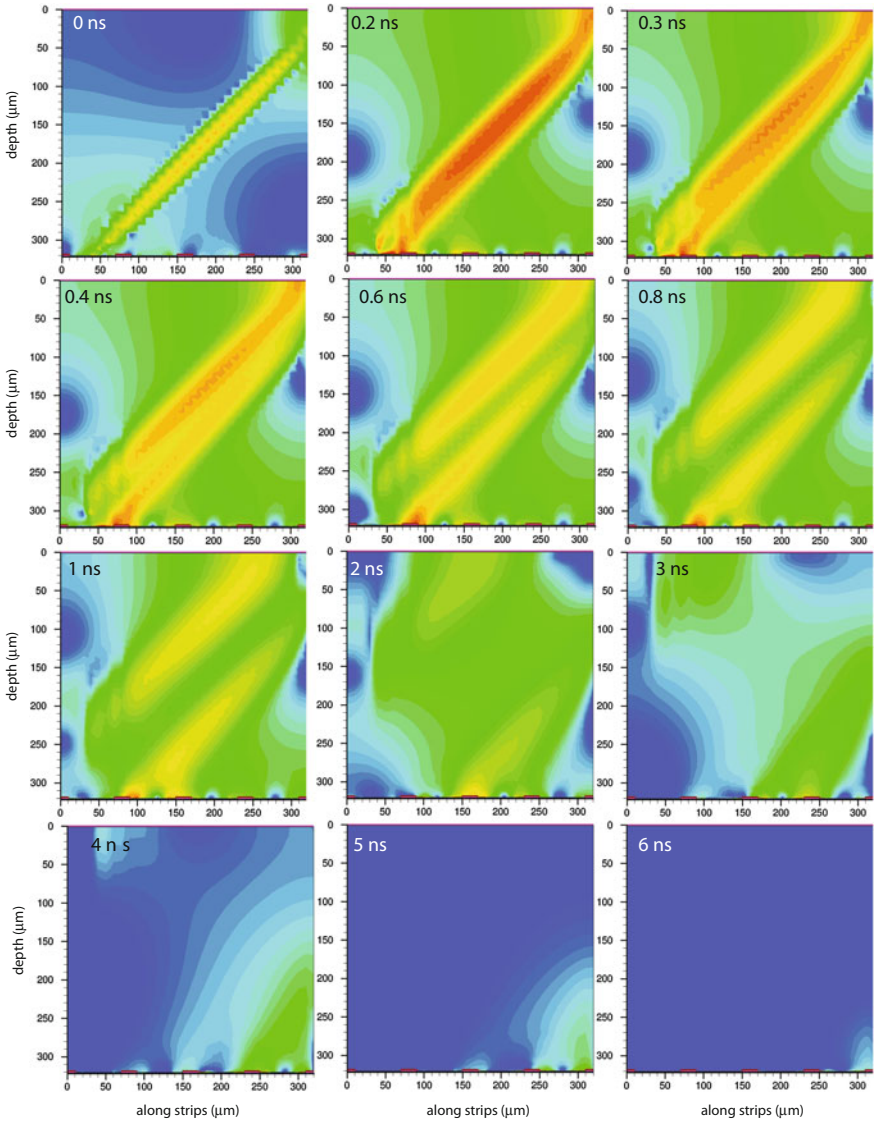


Fig. B.1 An ionising particle traverses the sensor at a 45° angle, disturbing the static situation. The 12 plots show the absolute sum of the charge densities (electrons and holes) for different times [Simulation]

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3

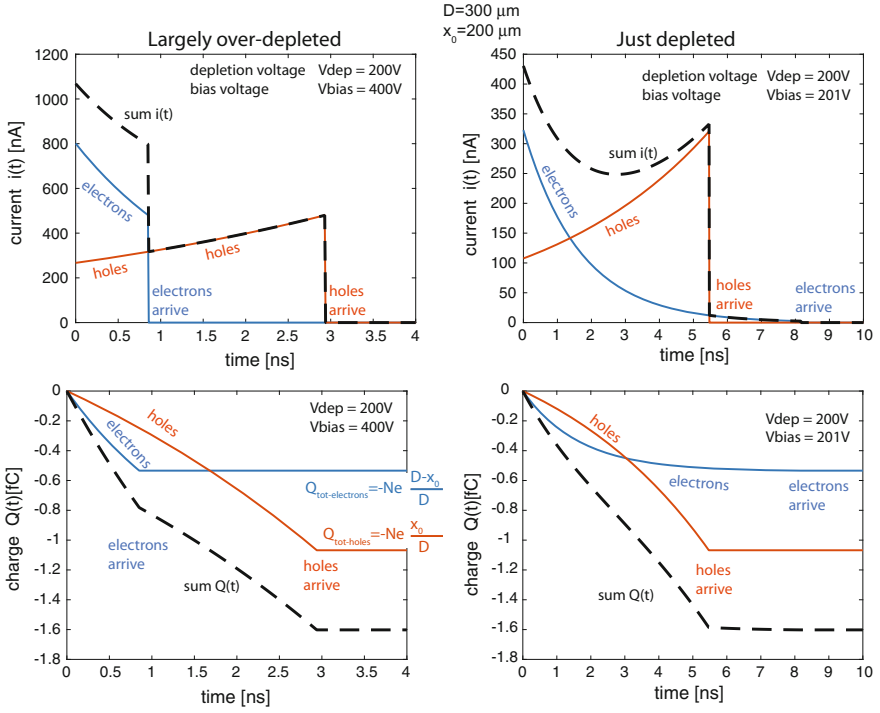


Fig. B.2 Signal currents and integrated charges of a $D = 300$ µm thick diode are plotted in a time resolved manner for different bias voltages. N=10000 elementary charges have been deposited at 2/3 thickness (200 µm away from p-electrode). The figure also illustrates the value of over-depletion defining largely signal shape and duration (*left*). With barely the depletion voltage the electric field at the n-electrode is that low that electrons arrive late (*right*). Differently to Fig. 1.22 the summed up charge due to electrons and holes are not the same. Reminder the weight field $E_W = \frac{1}{D}$is constant while the weighting potential $V_W(x) \propto x$ is linear within the bulk and with boundaries $V_W(x = 0) = 1$ and $V_W(x = D) = 0$ at p-, n-electrode respectively, hence $V_W(x_0 = 200\ \mu\text{m}) = \frac{1}{3}$. Therefore, with formula Eq. 1.32, $Q_{\text{hole}} = Nq_h\,(V_W(0) - V_W(x = 200\,\mu\text{m})) = Nq_h\left(1 - \frac{1}{3}\right) = Nq_h\frac{2}{3} = -\frac{2}{3}Ne$ moving towards the p-electrode and $Q_{\text{electron}} = Ne\,(V_W(D) - V_W(x_0 = 200\,\mu\text{m})) = Ne(0 - \frac{1}{3}) = -\frac{1}{3}Ne$ summing up to $Q_{\text{tot}} = -Ne(\frac{1}{3} + \frac{2}{3})$. The weighting potentials for a diode and a strip sensor is displayed in Fig. 1.23 on p. 30

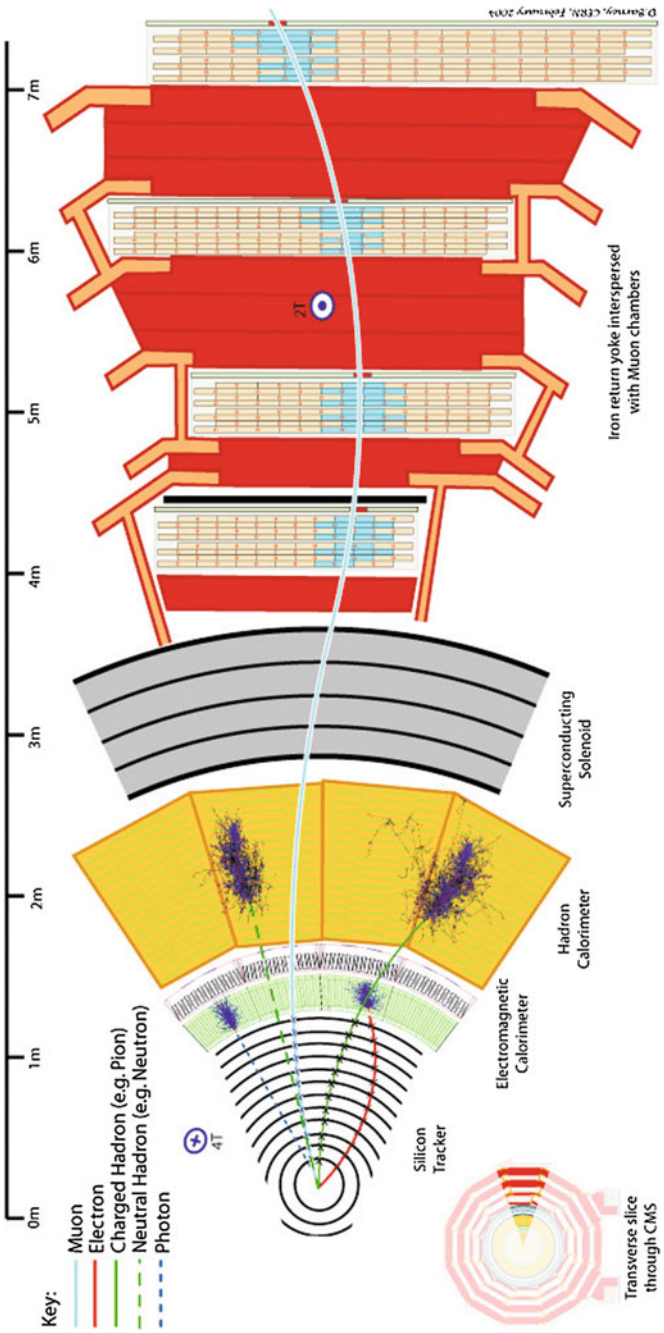


Fig. B.3 A slice through the CMS detector. CMS consists of a central silicon Tracker (Pixel and Strips), an Electromagnetic Calorimeter, a Hadron Calorimeter, a Superconducting Solenoid Magnet, a massive iron return yoke instrumented with Muon Chambers. The depicted interactions present an ideal detector behaviour for the different particles μ, e, charged and neutral hadrons, and γ. A more detailed introduction how HEP detectors work in general can be found in [135]. [Courtesy of CMS/CERN [329]]

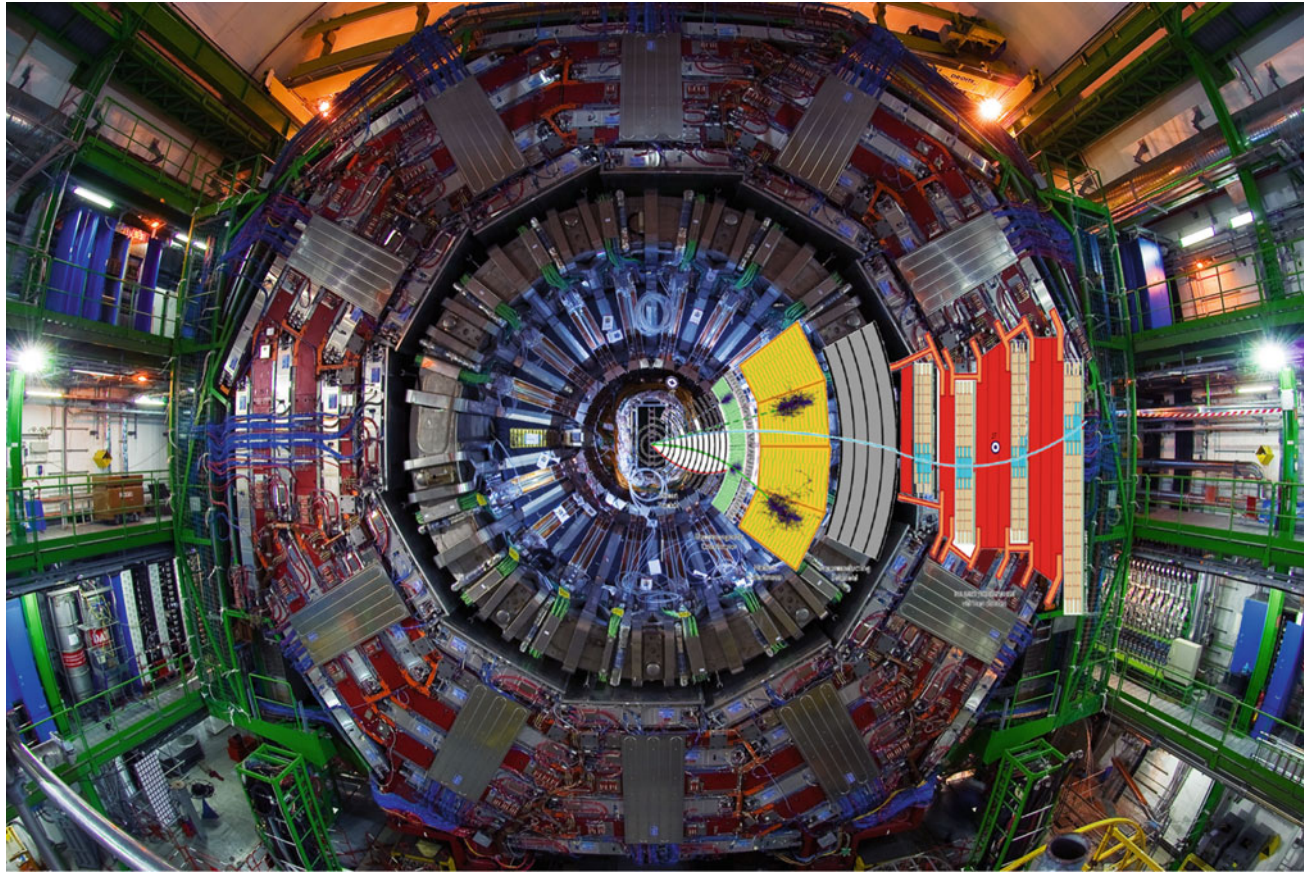


Fig. B.4 The schematic of the CMS detector demonstrating how particles are identified with combinations of detector elements, laid over a photo of the open CMS detector. [Courtesy of CMS/CERN [329]]

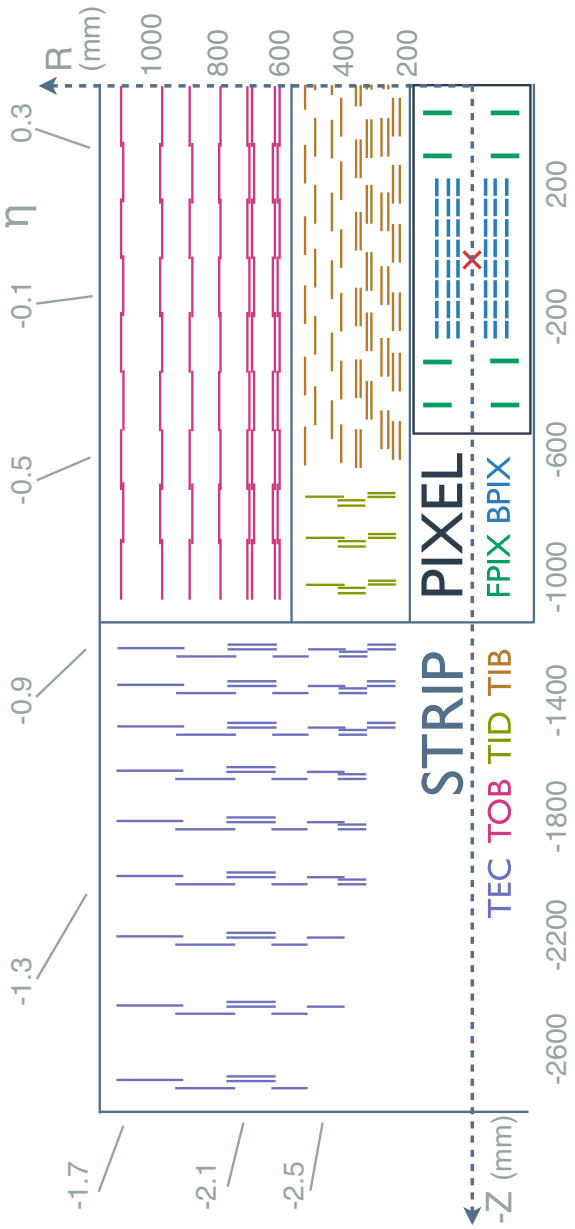


Fig. B.5 A quarter of the CMS Tracker with individual modules displayed (2008 – 2016). Each line shows an individual module. The module-overlaps to ensure hermeticity and the layers with double modules are shown in detail. Plot from [25]

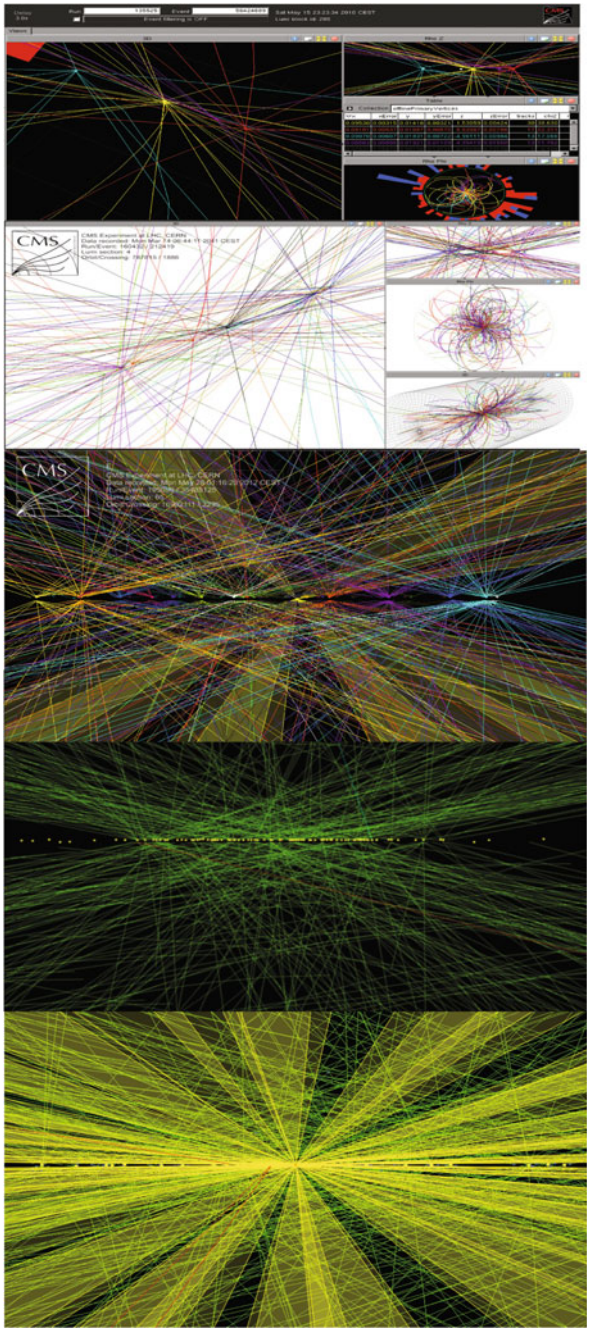


Fig. B.6 Pile-up of three at initial low luminosity $\mathcal{L} \sim 10^{29}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$; then PU = 7; then nominal with about 500 tracks; a dedicated high PU = 78 run with a small number of bunched and the situation during the Heavy Ion HI run. The frequency of events during HI is smaller and no zero suppression happens on the back-end electronics level but in the High Level Trigger farm. [Courtesy of CMS/CERN]

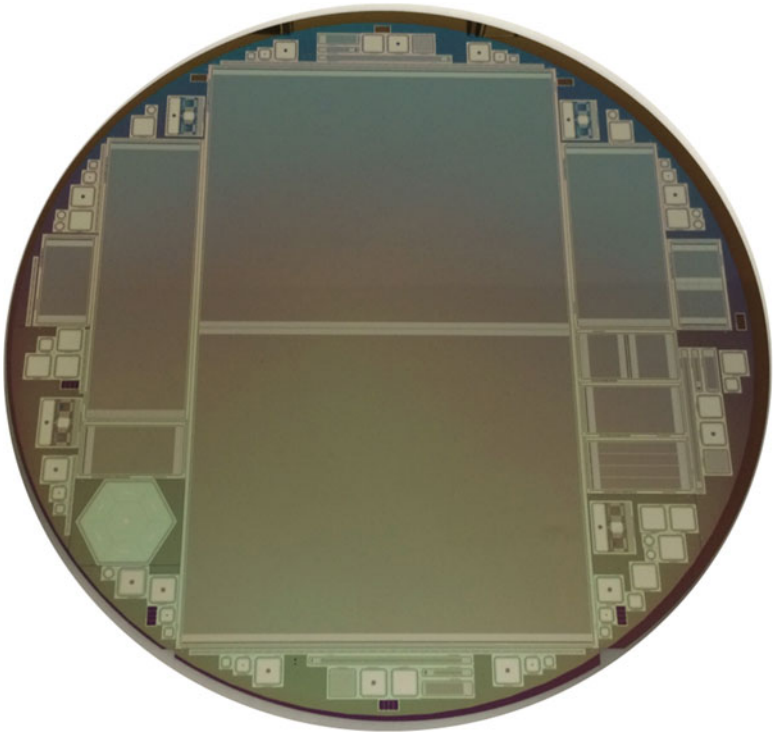


Fig. B.7 The first HEP sensor processed on an 8 inch wafer (scale 1:2). A design study of a longer 2S sensor with $16 \times 10\,\mathrm{cm}^2$ instead of $10 \times 10\,\mathrm{cm}^2$. The 2S-sensor concept is nicely visible with strip spanning only have of the sensor length where then readout chips will be located on both sensor ends. Only 6 inch sensors will be used for the CMS HL-LHC tracker, while 8 inch wafer processing is the baseline for the CMS high granularity endcap calorimeter [Courtesy of HEPHY/CMS]

References

1. A Multi-Tev Linear Collider Based on CLIC – Technology CLIC Conceptual Design Report, SLAC-R-985, KEK Report 2012-1, PSI-12-01, JAI-2012-001, CERN-2012-007, 12 Oct 2012
2. Abashian, A., et al.: The Belle detector. NIM **A479**, 127–232 (2002)
3. Abbaneo, D.: Upgrade of the CMS tracker with tracking trigger. J. Instrum. **6**; JINST **6**, C12065 (2011)
4. Abbaneo, D., et al.: Layout and performance of the CMS silicon strip tracker. NIM **A518**, 331–335 (2004)
5. Abe, F., et al.: Evidence for top quark production in p $\overline{p}$ collisions at sqrts = 1.8TeV. Phys. Rev. D **50**, 2966–3026 (1994)
6. Abe, F., et al.: Observation of the top quark production in p $\overline{p}$ collisions at $\sqrt{s}$ =1.8TeV. Phys. Rev. Lett. **74**, 2626 (1995)
7. Abreu, P., et al.: Phys. Lett. B **267**, 422 (1991)
8. Abulencia, A., et al.: Measurement of the $B_s^o \overline{B}_S^o$ oscillation frequency. Phys. Rev. Lett. **97**, 062003 (2006)
9. Acciarri, M., et al.: The L3 silicon microvertex detector: installation and results on 1993 performance. NIM **A360**, 103–109 (1995)
10. Adapted from idea of Frank Meier Aeschbacher
11. Affolder, A., Allport, P., Casse, G.: Collected charge of planar silicon detectors after pion and proton irradiations up to $2.2 \times 10^{16} n_{eq} cm^{-2}$, *NIMA* (2010). https://doi.org/10.1016/j.nima.2010.02.187
12. Affolder, A.: On behalf of the CMS Silicon Strip Tracker (SST) Collaboration: The CMS silicon strip tracker: design and production status. Nucl. Phys. B (Proc. Suppl.) **150**, 118–123 (2006)
13. Agram, J.-L., et al.: The silicon sensors for the compact muon solenoid tracker - design and qualification procedure, CMS Note 2003/015. NIM **A517**, 77–93 (2004)
14. Albergini-Quaranta, Martini, M., Ottaviani, G., Zanarini, G.: Experimental results on the information available from the rise time of pulses supplied by the semiconductor particle detector. IEEE Trans. Nuc. Sci. (1965)
15. Amendolia, S.R., et al.: A multi-electrode silicon detector for high energy experiments. NIM **A176**, 457–460 (1980)
16. Anderson, S., et al.: The extended OPAL silicon strip microvertex detector. NIM A **403**, 326–350 (1998)
17. Andreas Koerpert Messungen an Halbleiterdetektoren für die Herstellung des '96 Upgrade des DELPHI Vertexdetektors. Diplomarbeit Universität Karlsruhe (1995)

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3

18. Anzivino, G., et al.: The DELPHI silicon strip microvertex detector. NIM **A263**, 215–220 (1988)
19. Ardelean, J., Hrisoho, A., Truong, K., Seller, P.: Noise evaluation and improvement of the LAL-RAL microplex readout chip for the DELPHI μ-vertex detector. NIM **A315**, 393–396 (1992)
20. Auzinger, G.: For the CMS Tracker Collaboration: Analysis of testbeam data of irradiated silicon prototype sensors for the CMS tracker upgrade. NIM **A730**, 195 (2013). https://doi.org/10.1016/j.nima.2013.03.018
21. Bailey, R.: For the ACCMOR Collaboration: Measurement of mass and lifetime of hadronically produced charmed f mesons. Phys. Lett. **139B**(4), 320 (1984)
22. Barberis, E., et al.: Capacitances in silicon microstrip detectors. NIM **A342**, 90 (1994)
23. Barnett, R.M., et al.: Particle data group. Phys. Rev. D **54**, 1268 (1996)
24. Barth, C.: For the CMS Tracker Collaboration: Evolution of silicon sensor characteristics of the CMS silicon strip tracker. NIM **A699**, 178–183 (2013)
25. Bartosik, N.: Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} =$ 8 TeV as test of QCD and background tt^+ + Higgs production. https://doi.org/10.3204/DESY-THESIS-2015-035, CERN-THESIS-2015-120
26. Bateman, J.E., Hogg, W.R.: NIM **79**, 134 (1970)
27. Bedeschi, F., et al.: A silicon vertex detector for CDF, CDF Note 362 (1985)
28. Bedeschi, F., et al.: Estimated performance of the forward silicon detectors, CDF Note 213 (1983)
29. Behari, S., et al.: Operational experience, improvements, and performance of the CDF Run II silicon vertex detector. NIM A **729**, 153–181 (2013)
30. Behnke, T., Brau, J.E., Foster, B., Fuster, J., Harrison, M., McEwan Paterson, J., Peskin, M., Stanitzki, M., Walker, N., Yamamoto, H.: The international linear collider technical design report - volume 1: executive summary. arXiv:1306.6327 [physics.acc-ph]
31. Belau, E., et al.: Silicon detectors with 5 μ m spatial resolution for high energy particles. NIM **A217** (1983)
32. Belle II Technical Design Report. arXiv:1011.0352 [physics.ins-det]
33. Bellini, G., et al.: Live target performances in coherent production experiment. NIM **107** (1973)
34. Bellucci, L.: Silicon detector developments for CMS. NIM **A462**, 243–248 (2001)
35. Bergauer, T.: Process quality control of silicon strip detectors for the CMS tracker, Masters thesis, TU Wien (2004)
36. Bergauer, T., Dragicevic, M., König, A., Hacker, J., Bartl, U.: First thin AC-coupledsilicon strip sensors on 8-inch wafers. NIM **A830**, 473–479 (2016)
37. Bhardwaj, A., Eber, R., Eichhorn, T., Lalwani, K., Messineo, A., Peltola, T.H.T., Printz, M., Ranjan, K., Ranjeet, R.: Development of radiation damage models for irradiated silicon sensors using TCAD tools. PoS (TIPP2014), 276
38. Bianchi, G.: tkLayout: a design tool for innovative silicon tracking detectors. JINST **9** (2014). https://doi.org/10.1088/1748-0221/9/03/C03054/meta
39. Billoir, P.: Progressive track recognition with a Kalman like fitting procedure. Comput. Phys. Commun. **57**, 390 (1989). https://doi.org/10.1016/0010-4655(89)90249-X
40. Billoir, P., Qian, S.: Simultaneous pattern recognition and track fitting by the Kalman filtering method. NIM A **294**, 219 (1990). https://doi.org/10.1016/0168-9002(90)91835-Y
41. Bingefors, N., et al.: The DELPHI microvertex detector. NIM **A328**, 447–471 (1993)
42. Bolla, G.: Private communication
43. Bolla, G., et al.: Silicon microstrip detectors on 6 in technology. NIM **A435**, 51–57 (1999)
44. Bolla, G., et al.: Design and test of pixel sensors for the CMS experiment. NIM **A461**, 182 (2001)
45. Bolla, G., et al.: Status of the CDF silicon detecto. Eur. Phys. J C **33**(s01), s1020–s1022 (2004)
46. Bolla, G., et al.: Wire-bond failures induced by resonant vibrations in the CDF silicon detector. NIM **A518**, 277–280 (2004)

47. Borrello, L., Messineo, A., Focardi, E., Macchiolo, A.: Sensor design for the CMS Silicon Strip Tracker, CMS Note 2003/020
48. Bräunlich, P. (ed.): Thermally Stimulated Relaxation in Solids. Springer, Berlin, Heidelberg, New York (1979). ISBN 3-540-09595-0
49. Braibant, S., et al.: Investigation of design parameters for radiation hard silicon microstrip detector. NIM A **485**, 343–361 (2002)
50. Brau, J., et al.: Silicon detectors at the ILC. NIM **A579**, 567–571 (2007)
51. Broennimann, C., Glaus, F., Gobrecht, J., Heising, S., Horisberger, M., Horisberger, R., Kästli, H., Lehmann, J., Rohe, T., Streuli, S.: Development of an indium bump bond process for silicon pixel detectors at PSI. NIM. A **565**(1), 303–308 (2006)
52. Butz, E.: For the CMS Tracker Collaboration: CMS Strip Detector: Operational Experience and Run1 to Run2 Transition. PoS, 004 (Vertex2014)
53. Caccia, M., et al.: NIM **A260**, 124 (1987)
54. Caminada, Lea: Recent developments of HEP pixel detector readout chips. Phys. Procedia **37**, 1644–1653 (2012)
55. Cartiglia, N., et al.: Beam test results of a 16 ps timing system based on ultra-fast silicon detectors. NIM **A850**, 83 (2017)
56. Cartiglia, N., et al.: Tracking in 4 dimensions. NIM **A845**, 47–51 (2017)
57. Casse, G., Allport, P.P., Biagi, S.F., Bowcock, T.J.V., Greenall, A., Turner, P.R.: Charge collection and charge sharing in heavily irradiated n-side read-out silicon microstrip detectors - Research. NIM A **511**(1–2), 112–117 (2003)
58. Casse, G., Affolder, A., Allport, P.P., Wormald, M.: Measurements of charge collection efficiency with microstrip detectors made on various substrates after irradiations with neutrons and protons with different energies. PoS, 036 (VERTEX 2008)
59. Casse, G., et al.: Enhanced efficiency of segmented silicon detectors of different thicknesses after proton irradiations up to $1 \times 10^{16} neqcm^2$. NIM A. https://doi.org/10.1016/j.nima.2010.02.134
60. Cattai, A., et al.: Large-scale module production for the CMS silicon strip tracker. NIM **A541**, 104–109 (2005)
61. CDF II Collaboration: Technical Design Report, FERMILAB-Pub-96/390-E (1996)
62. Centro Nacional de Microelectr'onica, CNM-IMB (CSIC), Barcelona, Spain
63. Chabaud, V., et al.: The DELPHI silicon strip microvertex detector with double sided readout. NIM **A368**, 314–332 (1996)
64. Chilingarov, A.: Temperature dependence of the current generated in Si bulk. JINST **8**, P10003 (2013)
65. Chiochia, V., et al.: Simulation of heavily irradiated silicon pixel sensors and comparison with test beam measurements. IEEE Trans. Nucl. Sci. **52**(4) (2005)
66. Cogenda simulation tool, partially also available as open access, so particularly useful – https://www.cogenda.com/
67. Collins, P.: Experience with silicon detectors at the DELPHI experiment. LEP, NIM **A383**, 1–6 (1996)
68. Collins, P.: Semiconductor detectors for high-luminosity environment. NIM **A581**, 38–47 (2007)
69. Contarato, D., Kramberger, G.: Simulation of signal in irradiated silicon pixel detectors. NIM **A511**, 82–87 (2003)
70. Creanze, D., et al.: The new ALEPH silicon vertex detector. NIM **A409**, 157–160 (1998)
71. Curras, E., et al.: Radiation hardness and precision timing study of silicon detectors for the CMS High Granularity Calorimeter (HGC). NIM A (2016). https://doi.org/10.1016/j.nima.2016.05.008i
72. D'Alessandro, R.: On behalf of the CMS Silicon Tracker Collaboration: The CMS silicon strip tracker: experience in integration and commissioning. NIM **A582**, 739–743 (2007)
73. de Boer, W., Dierlamm, A., Hartmann, F., Hoffmann, K.-H., Kornmayer, A., Müller, T., Nürnberg, A., Pellegrini, G., Printz, M., Quirion, D.: A fourfold segmented silicon strip sensor with read-out at the edges. NIM **A788**, 154–160. 11 July 2015. https://doi.org/10.1016/j.nima.2015.03.082. ISSN 0168-9002

74. de Boer, W., et al.: Lorentz angle measurements in silicon detectors. NIM **A478**, 300–332 (2002)
75. De Cosa, A.: CMS Pixel detector: operational experience and run 1 to run 2 transition. PoS, 003 (Vertex2014)
76. Dearnaley, G.: Harwell, UKAEA Report WERE-R 3874 (1961), quoted by D.A. Bromley. IRE Trans. Nucl. Sci. **NS-9**, 135 (1962)
77. Demaria, N., et al.: New results on silicon microstrip detectors of CMS. NIM **A447**, 142 (2000)
78. Dierlamm, A.: For the CMS TK collaboration: CMS Strip Tracker R&D, ECFA High Luminosity LHC Experiments Workshop - 2016; Aix-Les-Bains (2016)
79. Dierlamm, A.: Irradiation qualification of CMS tracker components. 4th Int. Conf. Radiat. Effects Semicond. Mater. Detect. Devices, Florence 2002, NIM **A514** (2003)
80. Dierlamm, A.: Untersuchungen zur Strahlenhärte von Siliziumsensoren, IEKP-KA/2003-23, Ph.D. Thesis (2003)
81. Dierlamm, A.: Silicon detectors for the SLHC - An overview of recent RD50 results. NIM **A624**, 396–400 (2010)
82. Dierlamm, A.: For the CMS Tracker Collaboration: Silicon sensor developments for the CMS tracker upgrade. JINST **7**, C01110 (2012). https://doi.org/10.1088/1748-0221/7/01/C01110
83. Dierlamm, A.: For the CMS Tracker Collaboration: Characterisation of silicon sensor materials and designs for the CMS Tracker Upgrade. PoS Vertex **2012**, 016 (2013)
84. Dijkstra, H., et al.: Physics aspects of the DELPHI vertex detector. NIM Phys. Res. **A277**, 160–169 (1989). North-Holland, Amsterdam
85. Dirkes, G.: On behalf of the CMS Silicon Tracker Collaboration: Construction of the CMS silicon strip tracker. NIM **A581**, 299–302 (2007)
86. Doležal, Z., et al.: SiLC R&D: design, present status and perspectives. NIM **A579**, 750–753 (2007)
87. Dominguez, A., et al.: The CMS pixel detector. NIM **A581**, 343–346 (2007)
88. Dorokhov, A., et al.: Tests of silicon sensors for the CMS pixel detector. NIM A **530**, 71–76 (2004)
89. Eber, R.: Investigations of new Sensor Designs and Development of an effective Radiation Damage Model for the Simulation of highly irradiated Silicon Particle Detectors, Ph.D. Karlsruhe Institute of Technology, 2013, IEKP-KA/2013-27
90. Eber, R.: Private communication
91. Eränen, S., Kalliopuska, J., Orava, R., van Remortel, N., Virolainen, T.: 3D processing on 6 in. high resistive SOI wafers: Fabrication of edgeless strip and pixel detectors. NIM **A607**, 85–88 (2009)
92. Eremin, V., Li, Z., Iljashenko, I.: Trapping induced $_{N}eff$ and electrical-field transformation at different temperatures in neutron-irradiated high-resistivity silicon detectors. NIM **A360**, 458 (1995)
93. Eremin, V., Strokan, N., Verbitskaja, E., Li, Z.: Development of transient current and charge techniques for the measurement of effective net concentration of ionized charges (N_{eff}) in the space charge region of p-n junction detectors. NIM **A372**, 388 (1996)
94. Eremin, V., Verbitskaya, E., Li, Z.: The origin of double peak electric field distribution in heavily irradiated silicon detectors. NIM **A476**, 556–564 (2002)
95. Eremin, V., Roe, S., Ruggiero, G., Weilhammer, P.: The charge collection in single side silicon microstrip detectors. NIM A **500**(1–3), 121–231 (2003)
96. Eusebi, R., et al.: Operational experience with the CDF run II silicon detector. NIM A **729**, 153–181 (2013)
97. Fadeyev, V., et al.: Scribe-cleave-passivate (SCP) slim edge technology for silicon sensors. NIM **A731**, 260–265 (2013)
98. Fahrer, M.: Beam loss studies on silicon strip detector modules for the CMS experiment, IEKP-KA/2006-9, Ph.D. Thesis
99. Fahrer, M., Dirkes, G., Hartmann, F., Heier, S., Macpherson, A., Muller, T., Weiler, T.: Beam-loss-induced electrical stresstest on CMS silicon strip modules. NIM **A518**, 328–330 (2004)

100. Feld, L., Karpinski, W., Klein, K., Lipinski, M., Preuten, M., Rauch, M.: St. Schmitz and M. Wlochal The DC-DC conversion power system of the CMS Phase-1 pixel upgrade. J. Instrum. **10** (2015)
101. Fernandez-Martinez, P., Flores, D., Hidalgo, S., Greco, V., Merlos, A., Pellegrini, G., Quirion, D.: Design and fabrication of an optimum peripheral region for low gain avalanche detectors. arxiv.org/pdf/1510.08626.pdf
102. Ferrari, A.: FLUKA: a multi-particle transport code, CERN 2005-10, INFN/TC_05/ 11, SLAC-R-773 (2005)
103. Fiori, F.: CMS tracker operational experience. PoS, 006 (Vertex 2016)
104. Fleming, J.S.: NIM **150**, 417 (1978)
105. Fleming, R.M., Seager, C.H., Lang, D.V., Bielejec, E., Campbell, J.M.: Appl. Phys. Lett. 90, 172105 (2007); Fleming, R.M., Seager, C.H., Lang, D.V., Bielejec, E., Campbell, J.M. J. Appl. Phys. **104**, 083702 (2008)
106. Forbes, L., Sah, C.T.: On the determination of deep level center energy and concentration by thermally stimulated conductivity measured using reverse- biased p-n junctions. Solid-State Electron. **14**, 182–183 (1971)
107. Fretwurst, E., Feick, H., Glaser, M., Gossling, C., Heijne, E.H.M., Hess, A., Lemeilleur, F., Lindstrom, G., Miihlmann, K.H., Rolf, A., Schulz, T., Soave, C.: Reverse annealing of the effective impurity concentration and long term operational scenario for silicon detectors in future collider experiments. NIM **A342**, 119–125 (1994)
108. Fretwurst, E., Eremin, V., Feick, H., Gerhardt, J., Li, Z., Lindström, G.: Investigation of damage-induced defects in silicon by TCT. NIM **A388**, 356–360 (1997)
109. Friedl, M.: The CMS silicon strip tracker and its electronic readout. Ph.D. Thesis, Vienna University of Technology (2001)
110. Frühwirth, R.: Application of Kalman filtering to track and vertex fitting. NIM **A262**, 444 (1987). https://doi.org/10.1016/0168-9002(87)90887-4
111. Furgeri, A.J., de Boer, W., Hartmann, F.: Results of irradiation quality assurance of CMS silicon microstrip detectors. NIM **A573**, 264–267 (2007)
112. Gaillard, M.K., et al.: Rev. Mod. Phys. **47**, 277 (1975)
113. Garcia, M.F., Sanchez, J.G., Echeverría, R.J., Moll, M., Santos, R.M., Moya, D., Pinto, R.P., Vila, I.: High-resolution three-dimensional imaging of a depleted CMOS sensor using an edge Transient current technique based on the two photon absorption process (TPA-eTCT). NIM A **845**(11), 69–71 (2017)
114. Garcia, M.F., Sanchez, J.G., Echeverría, R.J., Moll, M., Santos, R.M., Moya, D., Pinto, R.P., Vila, I.: On the determination of the substrate effective doping concentration of irradiated HV-CMOS sensors using an edge-TCT technique based on the Two-Photon-Absorption process. JINST **12**, C01038 (2017)
115. Garcia-Sciveres, M., et al.: The SVX3D integrated circuit, NIM **A435**(1+2)
116. Garcia-Sciveres, M., et al.: The SVX4 integrated circuit. NIM **A511**, 171–173 (2003)
117. Gatti, E., Rehak, P.: Semiconductor drift chamber - an application of a novel charge transport scheme. NIM **A226**, 375 (1984)
118. Glückstern, R.L.: Uncertainties in track momentum and direction, due to multiple scattering and measurement errors. NIM **24**, 381 (1963)
119. Gonella, L., Barbero, M., Hügging, F., Krüger, H., Wermes, N.: The shunt-LDO regulator to power the upgraded ATLAS pixel detector. JINST **7** (2012)
120. Green, M.A., Keevers, M.J.: Optical properties of intrinsic silicon at 300 K. Research and Applications, Progress in Photovoltaics (1995)
121. Haase, E.L., et al.: Striped semiconductor detectors for digital position encoding. NIM **97**, 465 (1971)
122. Haber, C., et al.: The CDF SVX: a silicon vertex detector for a hadron collider. NIM **A289**, 388–399 (1990)
123. Haber, C., et al.: The discovery of the top quark: Instruments and methods **471**, 1217 (2001)
124. Hall, G.: Calculating parameters for the Pixel and Tracker upgrade performance studies (Tracker Week). http://bit.ly/eXvi8L

125. Hall, G.: LHC frontend electronics. NIM **A453**, 353–364 (2000)
126. Hall, R.N.: Electron-hole recombination in germanium. Phys. Rev. E **87**, 387 (1952)
127. Hansen, N.J., et al.: NIM **105**, 293 (1972)
128. Hartmann, F.: On behalf of the CMS Silicon Tracker Collaboration: The construction of the CMS tracker, NIM **A46120** (2000)
129. Hartmann, F.: On behalf of the CMS Silicon Tracker Collaboration: The CMS all-silicon tracker - strategies to ensure a high quality and radiation hard silicon detector. NIM **A478**, 285–287 (2002)
130. Hartmann, F.: On behalf of the CMS Silicon Tracker Collaboration: Construction issues for large detector systems with industry involvement. NIM **A549**, 171–177 (2005)
131. Hartmann, F.: Recent advances in the development of semiconductor detectors for very high luminosity colliders. NIM **A617**, 543–545 (2010)
132. Hartmann, F.: Semiconductor sensors. NIM A **628**(1), 40–49 (2010). https://doi.org/10.1016/j.nima.2010.06.282
133. Hartmann, F.: Evolution of silicon parameters due to irradiation at the LHC. PoS, 036 (Vertex 2011)
134. Hartmann, F.: Advances in tracking detectors. Annu. Rev. Nucl. Part. Sci. **61**, 197–221 (2011). https://doi.org/10.1146/annurev-nucl-102010-130052
135. Hartmann, F.: Sharma, Archana: Multipurpose detectors for high energy physics, an introduction. NIM **A666**, 1–9 (2012)
136. Hartmann, F.: Entwicklungsarbeit am Spurendetektor für das CDF Experiment am Tevatron, IEKP-KA/00-11, Ph.D. Thesis
137. Hartmann, F.: Messungen, Analyse und Optimierung von Silizium-Halbleiterdetektoren für die Herstellung des '96 Upgrade des DELPHI Vertexdetektors, Diploma Thesis IEKP-KA/96-09
138. Hartmann, F., et al.: The intermediate silicon layers detector at CDFII design and progress – proceedings of VERTEX '98, NIM **A435**(1+2), 44–50 (1999)
139. Hartmann, F., Bergauer, T., Fontaine, J.-C., Frey, M., Furgeri, A., Krammer, M.: Corrosion on silicon sensors. NIM A **569**(1), 80–83 (2006)
140. Havranek, Hemperek, K., et al.: DMAPS: a fully depleted monolithic active pixel sensor - analog performance characterization. JINST **10**(02), P02013 (2015)
141. Heijne, E.H.M., et al.: A silicon surface barrier microstrip detector designed for high energy physics. NIM **A178**, 331 (1980)
142. Heijne, E.H.M.: Semiconductor detectors in the low countries. NIM **A509**, 1–16 (2003)
143. Heine, V.: Studien zur Entwicklung von strahlenharten Siliziumsensoren für den CMS Detektor am HL-LHC, Diploma Thesis, IEKP-KA/2014-05
144. Heuser, J.M.: Construction, Operation and Application of the DELPHI Pixel Detector at LEP2, Dissertation at Bergische Universität-Gesamthochschule Wuppertal, WUB-DIS 99-1
145. Hocker, A.: Private communication during a visit at FERMILAB to understand the CDF Control and Safety System
146. Hoffmann, K.-H.: For the CMS Tracker Sensor Collaboration: Campaign to identify the future CMS tracker baseline. NIM **A658**, 30 (2011). https://doi.org/10.1016/j.nima.2011.05.028
147. Hollik, W., Müller, Th.: Das Top-Quark: sein Nachweis durch Theorie und Experiment. Phys. Bl. **53**, Nr.2 (1997)
148. Hopman, P.: Silicon detectors for physics. Talk, IEEE Short Course on Detectors for High Energy Physics (1997)
149. Horisberger, R.: Readout architectures for pixel detectors. NIM **A465**, 148–152 (2001)
150. Huhtinen, M.: Simulation of Non-Ionising Energy Loss and Defect Formation in Silicon, ROSE/TN/2001-02, CERN, CH-1211 Geneva 23, Switzerland (2001)
151. Hyams, B., et al.: The earliest successful use of SMDs in high energy physics experiments was by the NA11 collaboration at CERN, which did so much to develop these devices; A silicon counter telescope to study short-lived particles in high-energy hadronic interactions. NIM **205**, 99 (1983)

152. Hynds, D.: On behalf of the CLICdp collaboration: Silicon pixel R&D for the CLIC detector, PoS, 049 (Vertex 2016)
153. I: [Pintilie, I., et al.: Appl. Phys. Lett. **82**, 2169 (2003)] H defects: [Pintilie, I., et al.: Appl. Phys. Lett. **92**, 024101 (2008)] BD: [Pintilie, I., et al.: NIM **A514**, 18 (2003)] & [Pintilie, I., et al.: NIM **A556**(1), 197 (2006)] & [Fretwurst, E., et al.: NIM **A583**, 58 (2007)] E30: [Pintilie, I., et al.: NIM **A611**, 52–68 (2009)]
154. Incandela, J.: Private communication
155. Jarron, P., et al.: Deep submicron CMOS technologies for the LHC experiments, Nucl. Phys. B (Proc. Suppl.) **78**, 625–634 (1999)
156. Jones, L.L., et al.: The APV25 deep submicron readout chip for CMS detectors. 1999. In: Prepared for 5th Workshop on Electronics for the LHC Experiments (LEB 99). Snowmass, Colorado, 20–24 Sept (1999)
157. Joyce, D., et al.: Silicon detector tests with the ral microplex readout chip. NIM **A279**, 189–194 (1989)
158. Junkes, A.: Influence of radiation induced defect clusters on silicon particle detectors, Ph.D. thesis University of Hamburg (2011)
159. Junkes, A., Eckstein, D., Pintilie, I., Makarenko, L.F., Fretwurst, E.: Annealing study of a bistable cluster defect. NIMA **612**(3), 525–529 (2010)
160. Junkes, A., Pintilie, I., Fretwurst, E., Eckstein, D.: A contribution to the identification of the E5 defect level as tri-vacancy (V3). Phys. B **407**, 3013–3015 (2012)
161. Kajfasz, E.: The D0 silicon microstrip tracker for Run IIa. NIM **A511**, 1619 (2003)
162. Karimäki, V.: CMS Note 1997/064 [NIM **A410**, 284 (1998)] NIM **A305**, 187 (1991)
163. Kästli, H.C., et al.: Design and performance of the CMS pixel detector readout chip. NIM **A565**, 188 (2006)
164. Kästli, H.C., et al.: CMS pixel upgrade project. PoS. Vertex **2010**, 039 (2011)
165. Kemmer, J.: The earliest successful use of SMDs in high energy physics experiments was by the NA11 collaboration at CERN, which did so much to develop these devices. NIM **169**, 499 (1980)
166. Kemmer, J., Lutz, G.: NIM **A253**, 365 (1987)
167. Kemmer, J., et al.: NIM **A288**, 92 (1990)
168. Kemmer, J., et al.: Streifendetektor, Patentoffenlegungsschrift DE 19620081 A1. Munich, Germany **21** (1997)
169. Kittel, C.: Introduction to Solid State Physics, 7th edn. Wiley, New York, London, Sydney, Toronto
170. Knoblauch, D.: Diplomarbeit Aufbau und Test der Mikrostreifenmodule für den Upgrade des DELPHI Vertexdetektors, IEKP-KA/96-01
171. Kramberger, G.: Signal development in irradiated silicon detectors. Doctoral Thesis, University of Ljubljana (2001)
172. Kramberger, G.: Advanced transient current technique systems, PoS, 032 (Vertex2014)
173. Kramberger, G.: Solid State Detectors for High Radiation Environments, section 7 in Landolt-Börnstein. https://doi.org/10.1007/978-3-642-14142-3_7
174. Kramberger, G., Cindro, V., Mandić, I., Mikuž, M., Zavrtanik, M.: Determination of effective trapping time of holes in irradiated silicon, NIM **A476**(3) (2002). https://doi.org/10.1016/S0168-9002(01)01653-9
175. Kramberger, G., et al.: Effective trapping time of electrons and holes in different silicon materials irradiated with neutrons, protons and pions. NIM **A481**, 297–305 (2002)
176. Kramberger, G., Cindro, V., Mandic, I., Mikuž, M., Zavrtanik, M.: Field engineering by continuous carrier injection in silicon detectors irradiated with neutrons. NIM **A497**, 440 (2003)
177. Kramberger, G., Cindro, V., Dolenc, I., Mandic, I., Mikuž, M., Zavrtanik, M.: Performance of silicon pad detectors after mixed irradiations with neutrons and fast charged hadrons. NIM **A609**, 142–148 (2009)

178. Kramberger, G., Mandić, I.: Talks at the 15th RD50 workshop 2009 in Geneva http://rd50.web.cern.ch/rd50/15th-workshop; published later Kramberger, G., Cindro, V., Mandić, I., Mikuž, M., Milovanović, M., Zavrtanik, M., **Z**agar, K.: Investigation of irradiated silicon detectors by edge-TCT. IEEE Trans. Nucl. Sci. **57**(4) (2010)
179. Kramberger, G., Cindro, V., Mandić, I., Mikuž, M., Milovanović, M., Zavrtanika, M.: Modeling of electric field in silicon micro-strip detectors irradiated with neutrons and pions. JINST **9** P10016 (2014)
180. Kramberger, G., et al.: Radiation effects in low gain avalanche detectors after hadron irradiations. JINST **10**, P07006 (2015)
181. Krammer, M.: On behalf of the CMS Tracker Collaboration: Module production for the CMS tracker: problems and achieved quality. NIM **A582**, 766–770 (2007)
182. Krammer, M., Bergauer, T.: Private communication
183. Krammer, M., et al.: Experience with silicon sensor performance and quality control for a large-area detector. NIM **A511**, 136–144 (2003)
184. Krasel, O.: Charge Collection in Irradiated Silicon Detectors, Ph.D. Thesis, Uni Dortmund (2004)
185. Lampert, M., Mark, P.: Current Injection in Solids. Academic Press, New York (1970)
186. Lamport, J.E., et al.: NIM **134**, 71 (1976)
187. Lang, D.V.: Deep level transient spectroscopy: A new method to characterize traps in semiconductors. Journal of Applied Physics **45**, 3023 (1974). https://doi.org/10.1063/1.1663719
188. Lange, J., et al.: Charge collection studies of proton-irradiated n- and p-type epitaxial silicon detectors. NIM A **624**, 405–409 (2010)
189. Layer 00 was first proposed by J. Incandela, N. Bacchetta, R. Snider, D. Stuart. It is described in C.S. Hill, CDF Collaboration, NIM **A530**, 1 (2004)
190. Lindhard, J., Nielsen, V.: Nuclear collisions and ionisation fluctuations in charged particle detectors. Phys. Lett. **2**(5), 209–211 (1962)
191. Lindstrom, G., et al.: Radiation hard silicon detectors developments by the RD48 (ROSE) Collaboration. NIM **A466**, 308–326 (2001)
192. Linssen, L., Miyamoto, A., Stanitzki, M., Weerts, H. (eds.): Physics and Detectors at CLIC: CLIC Conceptual Design Report, CERN-2012-003 ; ANL-HEP-TR-12-01 ; DESY-12-008 ; KEK-Report-2011-7
193. Litke, A., et al.: A silicon strip vertex detector for the mark II experiment at the SLAC linear collider. NIM **A265**, 93–98 (1988)
194. Lutz, G.: Semiconductor Radiation Detectors. Springer, Berlin (1999)
195. Lutz, G., et al.: DEPFET-detectors: New developments, Frontier detectors for frontier physics. Proc. 10th Pisa Meetings Adv. Detect. NIM **A572**, 311–315 (2007)
196. Lutzer, B.: Characterization of irradiated test structures for the CMS tracker upgrade. NIM **A730**, 204–209 (2013). https://doi.org/10.1016/j.nima.2013.07.014
197. Macchiolo, A., et al.: Control of the fabrication process for the sensors of the CMS silicon strip tracker. NIM **A530**, 54–58 (2004)
198. Maehlum, G.: A measurement of lifetimes of b-hadrons using a silicon microstrip detector at LEP, University of Oslo (1993)
199. Maki, T., et al.: Status and performance of the CDF run II silicon detector. NIM **A579**, 723–725 (2007)
200. Mallows, S., et al.: Monte Carlo simulations of the radiation environment for the CMS experiment. NIM A **824**(11), 30–32 (2016). https://doi.org/10.1016/j.nima.2015.11.044
201. Mandić, I.: For the RD50 collaboration, Silicon sensors for HL-LHC tracking detectors. NIM A **732**, 126–129 (2013)
202. Mandić, I.: Silicon sensors for HL-LHC tracking detectors. NIM **A732**, 126–129 (2013)
203. Mandić, I., et al.: Observation of full charge collection efficiency in heavily irradiated n+p strip detectors irradiated up to $3 \times 10^{15} neq/cm^2$. NIM **A612**, 474–477 (2010)
204. Mangano, M., et al.: B physics at CDf: results and prospects. NIM **A333**, 57–22 (1993)
205. Manghisoni, M., et al.: 130 and 90 nm CMOS technologies for detector front-end applications. NIM **A572**, 368–370 (2007)

206. Mankel, R.: A Concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system. NIM **A395**, 169 (1997). https://doi.org/10.1016/S0168-9002(97)00705-5
207. Mankel, R.: Pattern recognition and event reconstruction in particle physics experiments. Rept. Prog. Phys. **67**, 553 (2004). https://doi.org/10.1088/0034-4885/67/4/R03. arXiv:physics/0402039
208. Marczewski, J.: In: Ambroziak, C.A. (ed.) Bulk silicon detectors of ionizing radiation. the role of the depletion layer. Institute of Electron Technology, Warsaw (2005). ISBN 83-914179-5-6
209. Marczewski, J., Domanski, K., Grabiec, P., Grodner, M., Jaroszewicz, B., Kociubinski, A., Kucharski, K., Tomaszewski, D., Kucewicz, W., Kuta, S., Machowski, W., Niemiec, H., Sapor, M., Caccia, M.: In: 2003 IEEE NSS Symposium SOI Active Pixel Detectors of Ionizing Radiation – Technology and Design Development, vol. NS-51, pp. 1025–1028. published in IEEE Trans. Nuclear Science, Portland, USA, 19–25 Oct 2003 (2004)
210. Materials Used in Semiconductor Manufacturing (Page 2). semicon matls2.htm (2001). http://www.semiconfareast.com/
211. McKay, K.G.: Electron-hole production in germanium by alpha-particles. Phys. Rev. **84**, 829–832 (1951)
212. Meier, B., et al.: Pixel Front-end development for CMS. PoS Vertex **2009**, 028 (2010)
213. Menichelli, D., Bruzzi, M., Li, Z., Eremin, V.: Modelling of observed double-junction effect. NIM **A425**, 135–139 (1999)
214. Menzione, A., et al.: Feasibility of operating silicon detectors inside the collider vacuum pipe, CDF Note No. 60 (1983)
215. Meschi, E., et al.: SVT: an online silicon vertex tracker for the CDF upgrade. NIM **A409**, 658–661 (1998)
216. Miller, L.F.: IBM. J. Res. Dev. **13**(3), 239 (1969)
217. Mittag, G.: For the CMS Collaboration: Alignment of the CMS tracker: latest results from LHC Run-II, CMS Conference Report, CMS CR -2017/011
218. Mnich, J.: Detectors for a linear collider. NIM **A579**, 567–571 (2007)
219. Moll, M.: Radiation Damage in Silicon Particle Detectors, Universität Hamburg, DESYTHESIS-1999-040 (1999)
220. Moll, M., et al.: Development of radiation tolerant semiconductor detectors for the Super-LHC. NIM A **546**, 99–107 (2005)
221. Nelson, T.K.: CDF collaboration. Int. J. Mod. Phys. A **16S1C**, 1091 (2001)
222. Nürnberg, A.: Studien an bestrahlten Siliziumsensoren für den CMS Spurdetektor am HL-LHC, IEKP-KA/2014-04 (2014)
223. Nürnberg, A., Schneider, T.: Lorentz angle measurements as part of the sensor R&D for the CMS Tracker upgrade. JINST **8**, C01001 (2013)
224. Nygård, E., et al.: CMOS low noise amplifier for microstrip readout design and results. NIM **A301**, 506–516 (1991)
225. Ohsugi, T., et al.: Micro-discharges of AC-coupled silicon strip sensors. NIM **A342**, 22 (1994)
226. Ohsugi, T., et al.: Micro-discharge at strip edge of silicon microstrip sensors. NIM **A383**, 116 (1996)
227. Ohsugi, T., et al.: Design optimization of radiation-hard, double sided, double metal. AC-coupled silicon sensors. NIM **A436**, 272–280 (1999)
228. Palmonari, F.M.: For the CMS Tracker Collaboration: CMS tracker performance. NIM **A699**, 144–148 (2013)
229. Parker, S.I., Kenney, C.J., Segal, J.: NIM **A395**, 328 (1997)
230. Parzefall, U.: Silicon for high-luminosity tracking detectors - recent RD50 results. Phys. Procedia **37**, 899–906 (2012)
231. Peisert, A.: Silicon Microstrip Detector. Instrumentation in High-Energy Physics. World Scientific, Singapore (1992)
232. Pellegrini, G.: NIM **A765**, 12–16 (2014)

233. Pellegrini, G., Baselga, M., Carulla, M., Fadeyev, V., Fernandez-Martinez, P., Fernandez-Garcia, M. , Flores, D., Galloway, Z., Gallrapp, C., Hidalgo, S., Liang, Z., Merlos, A., Moll, M., Quirion, D., Sadrozinski, H., Stricker, M., Vila, I.: Recent Technological Developments on LGAD and iLGAD Detectors for Tracking and Timing Applications. arXiv:1511.07175 [physics.ins-det]
234. Peric, I.: A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology. NIM **A582**, 876–885 (2007)
235. Peric, I.: Hybrid pixel particle-detector without bump interconnection. IEEE Trans. Nucl. Sci. **56**(2), 519–528 (2009)
236. Peric, I.: Strip technology and HVMAPS. PoS, 021 (Vertex 2012)
237. Peric, I., Kreidl, C., Fischer, P.: Hybrid pixel detector based on capacitive chip to chip signal transmission. NIM A **617**(1–3), 576–581 (2010)
238. Peric, I., et al.: High-voltage pixel sensors for ATLAS upgrade. NIM **A765**, 172–176 (2014)
239. Pernegger, H.: The DELPHI silicon tracker. Nucl. Phys. B (Proc. Suppl.) **54B**, 43–48 (1997)
240. Pesaresi, M., et al.: An FPGA-based track finder for the L1 trigger of the cms experiment at the high luminosity LHC, to be submitted to JINST (2017)
241. Pintilie, I., Fretwurst, E., Lindström, G., Junkes, A.: Appl. Phys. Lett. **92**, 024101 (2008)
242. Pintilie, I., et al.: Radiation-induced point- and cluster-related defects with strong impact on damage properties of silicon detectors. NIM A **611**(1), 52–68 (2009)
243. Printz, M.: T-CAD analysis of electric fields in n-in-p silicon strip detectors in dependence on the p-stop pattern and doping concentration. JINST **10**, C01048 (2015)
244. Printz, M.: On behalf of the CMS Tracker Collaboration: Radiation hard sensor materials for the CMS tracker phase II upgrade - charge collection of different bulk polarities. NIM **A765**, P29–34
245. Ξ_b Phys. Rev. Lett. **99**, 1052001 (2007); Charge Asymmetry in Inclusive B_sDecays, PRD **74**, 092001 (2006); Λ_b lifetime. arXiv:hep-ex/0609021v1; Mass: $m(\Lambda_b)$ PRL **96**, 202001 (2006); Ξ_b Mass Measurement Phys. Rev. Lett. **99**, 052002 (2007); Rare Decays (CDF). arXiv:0712.1708; Σ_b Published PRL **99**, 202001 (2007)
246. Radeka, V.: Low-noise techniques in detectors. Ann. Rev. Nucl. Part. Sci **38**, 217–77 (1988)
247. Radeka, V.: Signal Processing for Particle Detectors, section 3 in Landolt-Börnstein. https://doi.org/10.1007/978-3-642-03606-4_10
248. Radiation Effects on the SVX3 Chip, CDF Internal Note 4461
249. Radu, R., Fretwurst, E., Klanner, R., Lindstroem, G., Pintilie, I.: Radiation damage in n-type silicon diodes after electron irradiation with energies between 1.5 MeV and 15 MeV. NIM **A730**, 84–90 (2013). https://doi.org/10.1016/j.nima.2013.04.080
250. Radu, R., et al.: J. Appl. Physics **117**, 164503 (2015)
251. Ramo, S.: Proc. IRE **27**, 584 (1939)
252. Ravera, F.: on behalf of the CMS and TOTEM collaborations: The CT-PPS tracking system with 3D pixel detectors. JINST **11**, C11027 (2016)
253. RD19 CERN Detector R&D Collaboration: Development of Hybrid and Monolithic Silicon Micropattern Detectors, CERN/DRDC/P22/90-81, Anghinolfi, F., et al.: A 1006 element hybrid silicon pixel detector with strobed binary output. IEEE Trans. Nucl. Sci. **NS-39**, 650 (1992)
254. RD19 CERN Detector R&D Collaboration: Large scale pixel detectors for DELPHI at LEP200 and ATLAS at LHC. NIM **A342**, 233–239 (1994)
255. Regler, M., Frühwirth, R., Mitaroff, W.: Filter methods in track and vertex reconstruction. Int. J. Mod. Phys. C **7**, 521 (1996). https://doi.org/10.1142/S0129183196000454
256. Riegler, W.: Challenges for Particle Detectors at Future Circular Colliders, talk at the RD50 collaboration meeting, December 2015 & M.W. Riegler, CERN FCC Academic Training Lectures, Aleksa (2016)
257. Riegler, W., Aglieri Rinella, G.: Point charge potential and weighting field of a pixel or pad in a plane condenser. NIM **A767**, 267–270 (2014)
258. Rohe, T., et al.: NIM **A552**, 232 (2005)

259. Rossi, L., Fischer, P., Rohe, T., Wermes, N.: Pixel Detectors. Springer, Berlin (2006). ISBN-10:3-540-28332-3
260. Rovere, M.: Private communication
261. Ruggiero, G., et al.: IEEE Trans. Nucl. Sci. **52**, 1899 (2005)
262. Sadrozinski, H.: NIM **A765**, 7–11 (2014)
263. Sadrozinski, H.F.-W., et al.: Sensors for ultra-fast silicon detectors. NIM **A765**, 7–11 (2014)
264. Sauvage, D., Clemens, J.C. , Cohen-Solal, M., Raymond, M., Delpierre, P.A.: A Pixel Detector for the '95 Upgrade of the DELPHI MicroVertex Detector, CPPM-95-05, CERN SCAN-9601178
265. Schäfer, B.: Lehrbuch der Experimentalphysik Band 6 FESTKÖRPER. walter de Gruyter, Berlin, New York (1992)
266. Schulz, T.: Investigations on the Long Term Behaviour of Damage Effects and Corresponding Defects in Detector Grade Silicon after Neutron Irradiation. Ph.D. Thesis, University of Hamburg (1996)
267. Schulz, T., Feick, H., Fretwurst, E., Lindstrom, G., Moll, M., Mahlmann, K.H.: Long term reverse annealing in silicon detectors. IEEE Trans. Nucl. Sci. **41**(4), 791–795 (1994)
268. Schwarz, A.S.: Heavy flavour physics at colliders with silicon vertex detectors. Phys. Rep. **238**(1 & 2), 1–133 (1994). North-Holland
269. Sguazzonia, G.: On behalf of the CMS Silicon Tracker Collaboration: The construction of the CMS silicon srip tracker. Nucl. Phys. B (Proc. Suppl.) 177–178, 328–329 (2008)
270. Shockley, W., Read, W.T.: Statistics of the recombinations of holes and electrons. Phys. Rev. **87**, 835 (1952)
271. SiD Collaboration and ILD detector group: (Marcel Stanitzki (DESY) for the collaboration) Vertexing at the ILC. PoS VERTEX **2015**, 037 (2015)
272. Sill, A., et al.: CDF Run II silicon tracking projects. NIM **A447**, 1–8 (2000)
273. Silvaco simulutaion tool. http://www.silvaco.com
274. Silvestris, L.: Silicon tracking in CMS. Nucl. Phys. B (Proc. Suppl.) **120**, 239–248 (2003)
275. Spannagel, S.: On behalf of the CLICdp collaboration: Silicon Technologies for the CLIC Vertex Detector. JINST **12**, C06006 (2017)
276. Spieler, H.: Semiconductor Detector Systems, 2nd Printing 2006. Oxford University Press, Oxford (2005). ISBN-13:978-0198527848
277. Starodumov, A., Berger, P., Meinhard, M.: On behalf of the CMS Collaboration: High rate capability and radiation tolerance of the new CMS pixel detector readout chip PROC600, CERN CMS-CR-2016-269, to be submitted to JINST
278. Stewart, G., Bates, R., Fleta, C., Kramberger, G., Lozano, M., Milovanovic, M., Pellegrini, G.: Analysis of edge and surface TCTs for irradiated 3D silicon strip detectors. JINST **8**, P03002 (2013)
279. Strandlie, A., Frühwirth, R.: Track and vertex reconstruction: From classical to adaptive methods. Rev. Mod. Phys. **82**, 1419 (2010). https://doi.org/10.1103/RevModPhys.82.1419
280. Swartz, M.: A detailed simulation of the CMS pixel [23] sensor. [Online] (2002). http://cmsdoc.cern.ch/doc/notes/docs/NOTE2002_027
281. Swartz, M.: CMS pixel simulations. NIM **A511**, 88–91 (2003)
282. Synopsys simulutaion tool. http://www.synopsys.com
283. Sze, S.M.: Physics of Semiconductor Devices. Wiley, New York (1985)
284. The ALICE Collaboration: Technical Design Report for the Upgrade of the ALICE Inner Tracking System. J. Phys. G **41**, 087002 (2014). CERN-LHCC-2013-024; ALICE-TDR-017. https://doi.org/10.1088/0954-3899/41/8/087002
285. The ATLAS and CMS Collaborations: Combined measurement of the Higgs boson mass in pp collisions at $sqrts$ = 7 and 8 TeV with the ATLAS and CMS experiments 26 Mar 2015, Phys. Rev. Lett. **114** (2015) 191803 In summary, a combined measurement of the Higgs boson mass is performed in the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4l$ channels using the LHC Run 1 data sets of the ATLAS and CMS experiments, with minimal reliance on the assumption that the Higgs boson behaves as predicted by the SM. The result is $m_H = 125.09 \pm 0.24GeV = 125.09 \pm 0.21(stat) \pm 0.11(syst)GeV$, where the total uncertainty is

dominated by the statistical term, with the systematic uncertainty dominated by effects related to the photon, electron, and muon energy or momentum scales and resolutions. Compatibility tests are performed to ascertain whether the measurements are consistent with each other, both between the different decay channels and between the two experiments. All tests on the combined results indicate consistency of the different measurements within 1σ, while the four Higgs boson mass measurements in the two channels of the two experiments agree within 2σ. The combined measurement of the Higgs boson mass improves upon the results from the individual experiments and is the most precise measurement to date of this fundamental parameter of the newly discovered particle

286. The ATLAS Collaboration: ATLAS Technical Proposal for a General Purpose pp experiment at the Large Hadron Collider, CERN/LHCC/94-93 (1994)
287. The ATLAS Collaboration: Technical Design report – The Insertable B-Layer, CERN-LHCC-2010-013 ATLAS TDR 19
288. The CALICE Collaboration: Tests of a Particle Flow Algorithm with CALICE test beam data. JINST **6**, P07005 (2011)
289. The CDF Collaboration: Measurement of the $B_S^o - \overline{B}_S^o$ oscillation frequency. Phys. Rev. Lett. **97**, 062003 (2006)
290. The CMS Collaboration: Technical Proposal, CERN/LHCC 94-38 (1994)
291. The CMS Collaboration: Plots from early Posters explaining the CMS Detector and its physics (2000). http://cms.web.cern.ch/content/cms-photo-book-and-brochures
292. The CMS Collaboration: The CMS experiment at the CERN LHC. JINST **3**, S08004 (2008)
293. The CMS Collaboration: Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and Emiss, CMS Physics Analysis Summary CMS-PAS-PFT-09-001 (2009)
294. The CMS Collaboration: Commissioning and Performance of the CMS Pixel Tracker with Cosmic Ray Muons (2010). arXiv:0911.5434v2 [physics.ins-det]
295. The CMS Collaboration, Description and performance of track and primary-vertex reconstruction with the CMS tracker. CMS-TRK-11-001, CERN-PH-EP-2014-070. JINST **9**, P10009 (2014). https://doi.org/10.1088/1748-0221/9/10/P10009
296. The CMS Collaboration: HLT plot from CMS presents new 13 TeV results at end-of-year jamboree (2015)
297. The CMS Collaboration: Mechanical stability of the CMS strip tracker measured with a laser alignment system. J. Instrum. **12** (2017)
298. The CMS Collaboration Detector Performance and Software, Physics Technical Design Report, Vol I, CERN/LHCC 2006-001 CMS TDR 8.1
299. The CMS Collaboration: Addendum to the CMS Tracker, CERN/LHCC 200-016
300. The CMS Collaboration: Approved CMS alignment plots. https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerAlignmentLHCC2015#Primary_Vertex_Validation
301. The CMS Collaboration: CMS Phase II Tracker Conceptual Design Report - slides
302. The CMS Collaboration: CMS Phase II Upgrade Scope Document, CERN-LHCC-2015-019: LHCC-G-165
303. The CMS Collaboration: CMS Technical Design Report for the Phase-2 Tracker Upgrade, CERN-LHCC-2017-009, CMS-TDR-17-001, 1 July 2017
304. The CMS Collaboration: CMS Technical Design Report for the Pixel Detector Upgrade, CERN-LHCC-2012-016; CMS-TDR-11
305. The CMS Collaboration: Particle-flow reconstruction and global event description with the CMS detector, CMS-PRF-14-001, CERN-EP-2017-110. arXiv:1706.04965
306. The CMS Collaboration: Silicon strip tracker detector performance with cosmic ray data at the tracker integration facility, CMS NOTE 2008/032
307. The CMS Collaboration: *Studies of Tracker Material*, CMS PAS TRK-10-003
308. The CMS Collaboration: Technical Proposal for the Phase-II Upgrade of the CMS Detector; CERN-LHCC-2015-010; LHCC-P-008; ISBN 9789290834168
309. The CMS Collaboration: Technical Proposal for the Upgrade of the CMS Detector through 2020; CERN-LHCC-2011-006; CMS-UG-TP-1; LHCC-P-004

310. The CMS Collaboration: CMS Tracking Performance Results from early LHC Operation. Eur. Phys. J. C **70**, 1165–1192 (2010). https://doi.org/10.1140/epjc/s10052-010-1491-3
311. The CMS Collaboration: Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, CMS-HIG-12-028; CERN-PH-EP-2012-220, 31 July 2012. Phys. Lett. B **716**, 30 (2012)
312. The CMS Collaboration: Identification of b-quark jets with the CMS experiment. JINST **8**, P04013 (2013). https://doi.org/10.1088/1748-0221/8/04/P04013. arXiv:1211.4462
313. The CMS Collaboration: Alignment of the CMS tracker with LHC and cosmic ray data. JINST **9**, P06009 (2014). https://doi.org/10.1088/1748-0221/9/06/P06009
314. The CMS Tracker Collaboration: Planar silicon sensors for the CMS tracker upgrade. NIM **A732**, 113 (2013). https://doi.org/10.1016/j.nima.2013.07.058
315. The CMS Tracker Collaboration: The CMS tracker, CMS note
316. The CMS Tracker Collaboration, Dalal, R., et al.: Comparison of radiation hardness properties of p+n- and n+p- Si strip sensors using simulation approaches. In: 23rd RD50 Workshop (2013)
317. The D0 Collaboration: NIM **A338**, 185, FERMILABPUB-93/179-E (1994)
318. The DELPHI Detector at LEP: DELPHI collaboration. NIM **A303**, 233–276 (1991)
319. The DELPHI Silicon Tracker Group: THE DELPHI tracker at LEP2. NIM **A412**, 304–328 (1998)
320. The ILC Collaboration, ILC Technical Design Report: Volume 4, ILC-REPORT-2013-040 (2013)
321. The ILD Concept Group: The International Large Detector - Letter of Intent, DESY 2009-87, FERMILAB-PUB-09-682-e, KEK Report 2009-6
322. The ISL was first proposed by D. Stuart, J. Incandela, A. Yagil. It is described in A. Affolder et al., NIM **A453**, 84 (2000)
323. The LHCb Collaboration: The LHCb detector at the LHC. JINST **3**, S08005 (2008)
324. The LHCb VELO Group: Performance of the LHCb Vertex Locator, CERN-LHCb-DP-2014-001, 2014 JINST 9 P09007
325. The LHCb VELO Group: Radiation damage in the LHCb Vertex Locator, CERN-LHCb-DP-2012-005, arXiv:1302.5259v1 [physics.ins-det] 21 Feb 2013, 2013 JINST 8 P08002
326. The Particle Data Group. http://pdg.lbl.gov/
327. The STAR experiment at Brookhaven. http://www.bnl.gov/RHIC/STAR.htm
328. The Tracker Group of the CMS Collaboration: *P-Type Silicon Strip Sensors for the new CMS Tracker at HL-LHC*, JINST_12_P06018. https://doi.org/10.1088/1748-0221/12/06/P06018
329. The URL of CERN: http://cms.web.cern.ch
330. The URL of the ALICE Collaborations: http://aliceinfo.cern.ch/
331. The URL of the Berkely National Labs. www.bnl.gov
332. The URL of the CDF Collaboration. http://www-cdf.fnal.gov
333. The URL of the CMS Collaboration. http://cms.cern.ch/
334. The URL of the FCC Collaborations: https://fcc.web.cern.ch/
335. The URL of the Linear Collider Collaborations: https://www.linearcollider.org
336. The URL of the RD50 Collaboration. http://www.cern.ch/rd50/
337. The URL of the RD53 Collaborations: https://rd53.web.cern.ch/rd53/
338. The URL of the XEUS Experiment. www.rssd.esa.int/Xeus
339. Title CLIC and CLICdp Collaboration, Updated baseline for a staged Compact Linear Collider, CERN-2016-004; CERN-2016-004. arXiv:1608.07537
340. Tracker Operation and Performance at the Magnet Test and Cosmic Challenge CMS Note 2006/000, (1 Nov 2006)
341. Turini, N., et al.: CDF evidence for the Top quark: detector issues. Nucl. Phys. B (Proc. Suppl.) 44, 3–11 97 (1995)
342. van Heerden, P.J., et al.: The crystal counter. Physica **16** (1950)
343. van Lint, V.A.J., Flanagan, T.M., Leadon, R.E., Naber, L.A., Rogers, V.C.: Mechanisms of Radiation Effects in Electronic Materials. Wiley, New York (1980)
344. Veszpremi, V.: For the CMS Collaboration: Operation and performance of the CMS tracker. JINST **9**, C03005 (2014). https://doi.org/10.1088/1748-0221/9/03/C03005

345. Veszpremi, V.: For the CMS Collaboration: Radiation experience with the CMS pixel detector. JINST **10**, C04039 (2015). https://doi.org/10.1088/1748-0221/10/04/C04039
346. Virdee, J.: Requirements from the experiments in Year 1. In: Proceedings of LHC Performance Workshop Chamonix XII, Chamonix CERN-AB-2003-008 ADM, 3–8 Mar 2003
347. Visit at the Instytut Technologii Elektronowej ITE in Warsaw
348. Wagner, J., et al.: Operational experience with the silicon detector at CDF. NIM **A576**, 220–222 (2007)
349. Walker, J.T., Parker, S., Hyams, B.D., Shapiro, S.L.: NIM **226**, 200 (1984)
350. Weiler, T.: Studies on silicon strip detectors for the CMS experiment, IEKP-KA/2004-22, Ph.D. Thesis (2004)
351. Weiler, T., et al.: The performance of irradiated CMS silicon micro-strip detector modules, CMS Note 2006/049
352. Weilhammer, P.: Double-sided Si strip sensors for LEP vertex detectors. NIM **A342**, 1–15 (1994)
353. Weisberg, L.R., Schade, H.: A technique for trap determinations m low-resistivity semiconductors. J. Appl. Phys. **39**(11), 5149–5151 (1968)
354. WODEAN project (initiated in, : 10 RD50 institutes, guided by G. Lindstroem, Hamburg) (2006)
355. Worm, S.: Vertex detector for the ILC. NIM **A569**, 123–126 (2006)
356. Wunstorf, R.: DESY-FH1K-92-01, Ph.D. Thesis (1992)
357. Wunstorf, R.: Radiation hardness of silicon detectors: current status. IEEE Trans. Nucl. Sci. **44**(3), 806–814 (1997)
358. Wüstenfeld, J., et al.: Surface effects before and after ionisation-induced surface defects. In: 6th Rose Workshop on Radiation Hardening of Silicon Detectors (2001)
359. Yohay, R.: For the CMS Tracker Collaboration: Operation and Performance of the CMS Silicon Tracker, 14th ICATPP Conference on Astroparticle, Particle, Space Physics and Detectors for Physics Applications, pp. 546–550. Como, Italy, 23-27 Sep 2013. https://doi.org/10.1142/9789814603164_0085
360. Zimmerman, T., et al.: SVX3: a deadtimeless readout chip for silicon strip detectors. NIM **A409**, 369–374 (1998)

Index

© Springer International Publishing AG 2017

F. Hartmann, *Evolution of Silicon Sensor Technology in Particle Physics*,
Springer Tracts in Modern Physics 275, DOI 10.1007/978-3-319-64436-3