Chapter 1

The CMS pixel detector

W. Erdmann for the CMS pixel group
Paul Scherrer Institute, Laboratory for Particle Physics
5232 Villigen PSI, Switzerland
wolfram.erdmann@psi.ch

The inner tracking detectors at the LHC operate in a region of unprecedented particle rates. Thousands of hits must be detected, time-stamped and stored in every 25 ns bunch crossing interval. The silicon pixel detector for the CMS experiment has been designed to meet the requirements of position resolution, rate capability and radiation tolerance with a minimal amount of material. Its unique ability to provide three-dimensional high precision space-points plays an important role in the tracking system of the CMS detector.

1.1. Introduction

The identification of long lived particles through the detection of secondary vertices and impact parameters is an important ingredient of a high energy physics experiment. Vertex detection of particles with sub-millimeter decay-length requires precise track measurements close to the production point. The feasibility of a tracking detector in the hostile environment near the beam of the LHC was unclear during the initial stages of planning of the CMS experiment.[1]

Energy and luminosity targets for the LHC lead to particle fluxes far beyond those encountered at other colliders. Every cm^2 of a detector close to the beam pipe is expected to be hit by 100 million particles per second when the LHC runs at the full design luminosity of $10^{34}\mathrm{cm}^{-2}s^{-1}$. The smallest radius at which a traditional silicon microstrip detector can function in such an environment is limited by occupancy and radiation damage. Reducing the size of the sensor elements brings occupancy to a manageable level. It also mitigates the problems of radiation damage of the sensors. Smaller detector capacitance permits high signal-to-noise ratios even when charge collection degrades. Leakage currents per channel remain small enough to be absorbed in the front-ends. Extremely short strips or pixels are therefore an inevitable choice for a vertex detector at the LHC.

Pixels with comparable dimensions in the transverse plane and the z-direction become two-dimensional arrays and introduce a new quality to tracking and tech-

nology. Hits detected by the pixel detector are true space points with considerable benefits for pattern recognition. The absence of ambiguities connected with projections makes the pixel detector the preferred starting point for track finding in CMS.[2] The relative simplicity and robustness of reconstruction have also lead to applications in the High Level Triggers.[3] The CCD detectors of SLD and NA32 had proven the power of pixelated tracking detectors, but this technology cannot be pushed to cope with LHC rates. New territory for electronics, sensor and interconnect technology was entered for building the CMS pixel detector.

The CMS pixel project started in 1994 as a collaboration of PSI with Swiss and Austrian universities[a]. A parallel effort existed at UC Davis. US groups[b] took responsibility for the forward endcaps of the pixel detector.

The work of the first years focused on very basic R&D on sensors and radiation hard readout chip concepts. Detailed studies of charge collection and Lorentz drift in heavily irradiated silicon sensors were made using the grazing angle method.[4,5] Power consumption, speed, and noise of analog stages realized in radiation hard CMOS processes were evaluated.[6,7] Building blocks of the readout and small prototype chips followed. A fine pitch bump-bonding process suitable for the target pixel size was developed at PSI.[8]

The mechanical detector layout and a detailed module design were developed for the Technical Design Report[9] (TDR) in 1998. A first realization of the readout architecture in a full readout chip was a major milestone in 2002. The PSI43 chip,[10] which uses the radiation hard 0.8μm DMILL[11] process, came close to meeting the requirements but would have needed at least one more iteration. By then it had become clear that deep sub-micron processes can provide the required radiation tolerance[12] while offering a number of advantages.[13] The PSI43 architecture was improved and translated to a commercial 0.25μm process. After test submissions in 2003 and 2004 the main production of the PSI46 chip[14] started in 2006.

Detector modules were produced between mid 2006 and the end of 2007 at PSI for the barrel, and in the US for the endcaps. The forward detector was integrated at Fermilab and shipped to CERN where it was fully commissioned and tested in 2008. The barrel detector was assembled and tested at PSI and shipped to CERN in June 2008. Both detector parts were installed and commissioned[15–17] successfully in July 2008.

The following section gives an overview of the pixel system and defining requirements. More detailed discussion of individual topics will follow in subsequent sections.

[a]ETH Zürich, University of Zürich, University of Basel, HEPHY Vienna

[b]Northwestern, Rutgers, FNAL, Purdue, Johns Hopkins, Mississippi, Davis. Later joined by Nebraska, Iowa, Kansas State, Cornell, Milan

1.2. Overview

One main motivation for a tracking detector close to the beam-line is the identification of secondary vertices of b quarks and possibly τ leptons. This requires an impact parameter resolution on the order of 100 μm or better. The determining factors for the impact parameter measurement are the position resolution of the innermost detector and the extrapolation uncertainty. The latter is largely due to multiple scattering in the beam pipe and detector and it grows with the distance between the interaction point and the first measurement. The obvious primary goal is therefore to provide a position measurement with a precision much better than 100μm at the smallest possible radius with the smallest possible amount of material. The power dissipation of the electronics must be minimized because cooling and cabling make a significant contribution to the material budget. Material in outer layers is less important for the impact parameter resolution, but must be kept minimal, too, in order not to degrade the overall performance of the experiment.

More specific for a CMS pixel detector are the requirements of radiation hardness to at least $6 \times 10^{14} n_{eq}/cm^2$ and (Section 1.2.3) a readout architecture that handles 40 MHz bunch crossing frequency with 20 simultaneous pp collisions. Only a small fraction of the bunch crossings will be read out, but all hits must be stored during the 3 μs latency of the CMS trigger. Trigger rates up to 100 kHz are foreseen.

1.2.1. *The pixel shape*

A central decision in the development of the CMS pixel detector was to exploit charge sharing among pixels to improve the position resolution. Significant charge sharing is a consequence of the Lorentz drift in the strong magnetic field of 4 T inside CMS. Charge carriers released by the ionizing particle in the silicon sensor do not follow the electric field lines to the collection electrodes, but are deflected by the Lorentz force (Fig. 1.1). The n-side readout of the silicon sensors, chosen for its radiation tolerance, leads to a larger Lorentz-angle because of the higher mobility of electrons compared to holes. Interpolating positions between pixels based on the amount of the collected charge requires the transmission of analog pulse-height information. The analog optical link developed for the CMS strip tracker naturally offers that possibility without the need for digitization in the front-end.

The position resolution of single-pixel hits is given by the pitch divided by $\sqrt{12}$. Two-pixel clusters and interpolation allow a much better resolution, limited only by fluctuations of the charge deposition. Dividing the signal charge among more than two pixels increases the data rate and reduces the signal charge per pixel but does not improve the resolution. An ideal choice of the pixel size in the direction perpendicular to the magnetic field ($r\phi$) is therefore given by length over which charges are spread when they reach the surface of the sensor. For the usual $\sim 300\mu$m sensor thickness and a Lorentz angle of 25° this amounts to $\sim 150\mu$m. A slightly smaller size of 100 μm was chosen to maintain charge sharing, and hence resolution,

after irradiation (Section 1.3).

The area of a pixel must be large enough to accommodate the readout electronics. With one dimension fixed by the Lorentz drift, this leads to a more or less quadratic shape of 100μm(rϕ) $\times$ 150μm(z). Traditionally, collider detectors have much better resolution in the $r\phi$ plane where the transverse momentum is measured from the bending of charged tracks. From the point of view of vertex reconstruction or pattern recognition there is no reason to favor the transverse coordinate. The pixel shape of the CMS pixel detector results in comparable resolution in both directions. Another advantage of an (almost) square shape is the small circumference for a given area. This minimizes the pixel capacitance which is important for noise, speed, and power dissipation.

There is no Lorentz drift in the direction parallel to the magnetic field (z-direction), but sufficiently inclined tracks are detected in more than one pixel, allowing interpolation in both directions. At high rapidity, where tracks hit the barrel detector at low angles, the small z-size is a disadvantage because increasing cluster size in the z-direction is only beneficial for the z-resolution until it exceeds two pixels. Higher multiplicities put a burden on the readout system without improving the resolution. Pixel disks therefore complement the barrel detector. Tilting the disk sensors away from the $r\phi$–plane introduces an angle between electric and magnetic field and hence Lorentz-drift. Sufficient charge sharing is achieved with a tilt angle of 20° to reach resolutions of approximately 15 μm.

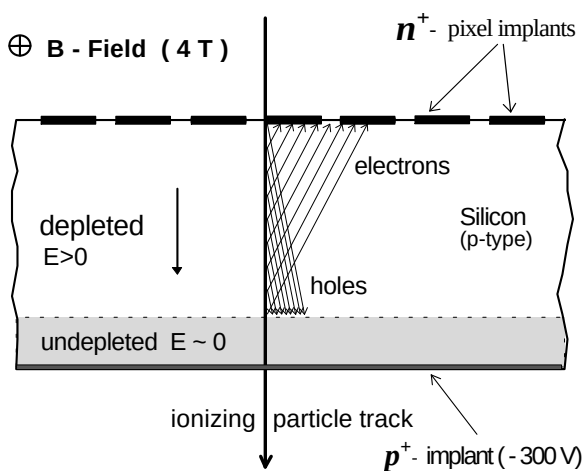

Fig. 1.1. Illustration of charge sharing induced by Lorentz drift in the CMS pixel barrel detector.[9] For improved position resolution the charge should be collected by at least two pixels. After irradiation, the detector stays functional beyond the point where it cannot be fully depleted, but the amount of charge and the charge sharing are reduced.

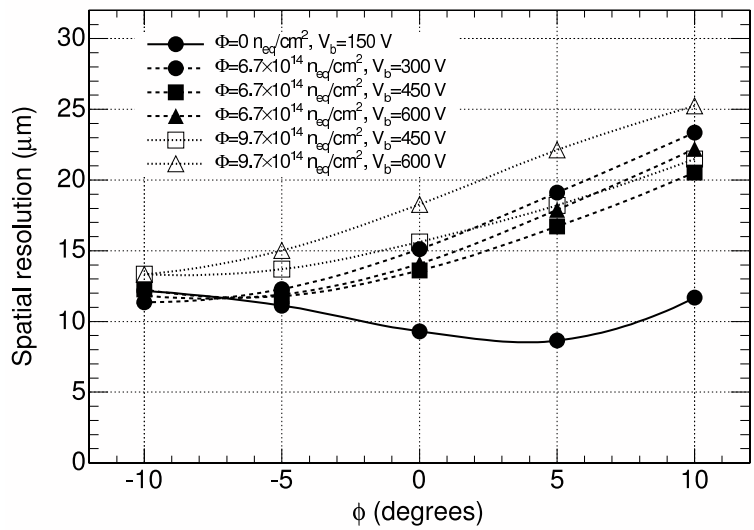

Fig. 1.2. Expected position resolution of the CMS pixel barrel in the $r - \phi$ plane as a function of the track angle of incidence.[2] The resolution degrades with increasing fluence and bias voltage because of the reduced charge sharing.

1.2.2. *Detector layout*

The inner layer of the detector should be mounted as close as possible to the interaction point. The smallest possible radius allowed by the CMS beam pipe is 4 cm. Further layers should be as far away as possible, as long as reliable matching of the outer track stub with the innermost hit is possible. Two additional layers are placed at radii 7 cm and 11 cm. The full radius of the available volume was not used in order to limit the total area (and hence cost) while providing three pixel hits per track.

The CMS tracker covers a pseudorapidity range of $|\eta| < 2.5$. The 4.4 cm layer, which is essential for the impact parameter resolution, nominally covers this range with its length of 56 cm. The acceptance begins to drop a little bit below $|\eta| = 2.5$ because of the length of the interaction region. A longer barrel is disfavored because the tracks crossing the detector at very shallow angles traverse more material and produce longer clusters.

The full η coverage in the outer layers would require a barrel length of up to 130 cm. The same coverage is achieved with much less detector area, and hence material, by supplementing disks in the forward region. A disadvantage of this configuration is that the bulkhead of the barrel where a lot of material resides comes to lie within the tracking volume. Two disks are placed on each side of the barrel at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm. They have an inner radius of 6 cm and an outer radius of 15 cm. The disk positions optimize the probability of measuring two or more pixel

hits on any track inside the tracker acceptance.

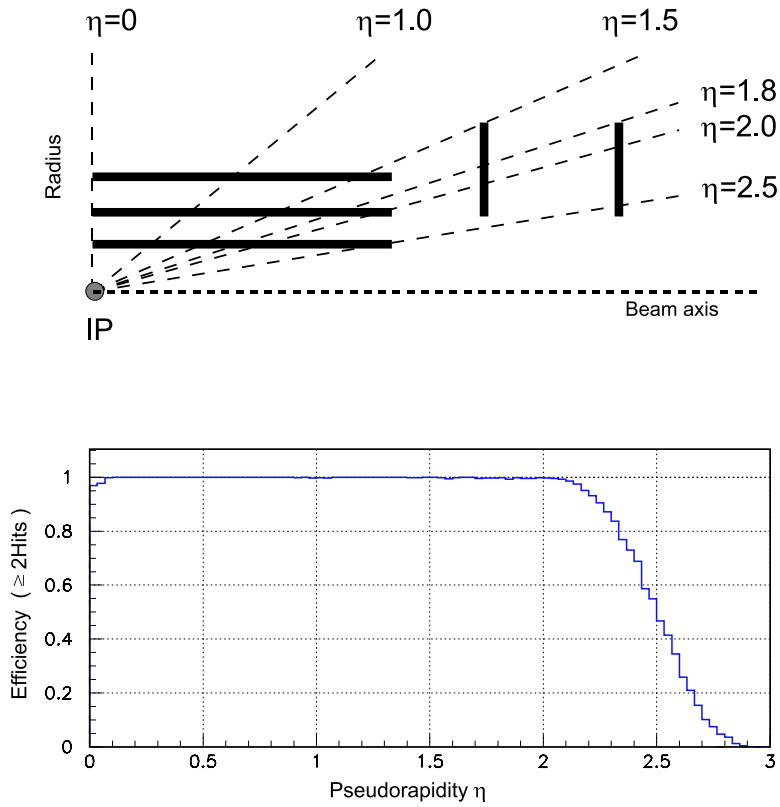

Fig. 1.3. Layout of the CMS pixel detector[18] (top). Three barrel layers at radii 4 cm, 7 cm and 11 cm cover the central region. Two endcap discs at $z = 34$cm and $z = 46$ cm extend the coverage to high rapidity. The region in which tracks are measured with at least two hits in the pixel detector (lower plot) is well matched to the acceptance of the outer tracker of CMS.

1.2.3. *Radiation hardness*

The radiation inside the CMS detector is a spectrum of charged and neutral particles coming from primary proton-proton collisions and from secondary interactions in the material of the detector.

Ionizing radiation leads to the accumulation of charges in CMOS electronics and sensor surface structures. The consequences of that must be mitigated by appropriate designs. All components of the radiation generate bulk damage in the silicon of the sensor which determines the lifetime of the detector. The amount of damage depends on the particle type and energy. It is collectively expressed in terms of an equivalent number of neutrons with 1 MeV kinetic energy (n_{eq}).

The 4 cm layer in CMS is expected to be exposed to a fluence of $3 \times 10^{14} n_{eq}/cm^2$

per year when the LHC operates at full luminosity. Radiation hard electronics available at the time of the TDR was believed to be functional up to $6 \times 10^{14} n_{eq}/cm^2$. This was taken as the target tolerance to be achieved by sensors and other components. The yearly fluences for the 7 cm and 11 cm layers are expected to be $1.2 \times 10^{14} n_{eq}/cm^2$ and $0.6 \times 10^{14} n_{eq}/cm^2$, respectively. While the lifetime of the 11 cm layer exceeds the projected 10 years of LHC operation, the 7 cm layer would need to be replaced at least once and the 4.4 cm layer every two years. Extracting the pixel detector in such intervals or even more frequently is also suggested by the possible need to bake out the beam pipe. Accommodating extraction and re-insertion during a few months shutdown has an impact on the mechanical structure and the cabling of the pixel detector and CMS. Details of the installation procedure needed to be defined at an early stage of the project.

Originally, a staged installation was foreseen in which only the inner two layers were present during a low luminosity phase of the LHC. Later, when the LHC had reached full luminosity, the 11 cm layer would have been added and the damaged 4.4 cm layer removed. The staging scenario eventually became obsolete as the LHC startup slipped and the full three-layer detector was installed in 2008.

Advances in sensor material engineering and the availability of very radiation hard deep sub-micron CMOS technology have led to a detector that will be able to withstand significantly higher doses than anticipated at the time of the TDR.

1.3. Sensors

The sensors for the forward and barrel detectors were developed independently in cooperation with two different vendors. Both adopted the n-in-n concept where the pixels are formed by high dose n-implants introduced into a highly resistive n-substrate. The junction is formed with a p-implant on the back-side. After space charge sign inversion, the bias voltage needed to fully deplete the sensor rises with irradiation. In contrast to a p-side readout the n-in-n design allows significantly under-depleted operation because the region of the high electric field has moved to the readout side after irradiation. Extremely high operating voltages can therefore be avoided. This reduces the problems of leakage currents and high voltage breakdowns in a highly miniaturized environment.

The n-in-n sensor implies the collection of electrons with their larger mobility compared to holes. This leads to the relatively large Lorentz-angle that is a central part of the detector concept. It also reduces trapping, which begins to play a role at the highest targeted fluences.

The pn-junction on the back-side of the sensors must not extend into the diced edges of the sensor. Therefore, the back-side must be structured and double-sided processing of the wafers is mandatory. This leads to significantly higher costs but has another important advantage. The guard ring structure implemented on the back side keeps the sensor edges near ground potential. This removes the risk of

high voltage sparks across the 15 μm air gap between the sensor edge and the readout chip.

Additional processing is also needed on the readout side to electrically isolate the n-implants from each other. The electron accumulation layer induced by ionizing radiation otherwise tends to short-circuit the pixel implants. Two inter-pixel isolation techniques were evaluated with good results. Open p-stops[19] were selected for the forward pixel detector and moderated p-spray[20] for the barrel detector.

P-stops are rings of high dose p-implants surrounding the pixels that interrupt the conducting layer. P-spray consists of a medium dose p-implantation "sprayed" without mask over the whole wafer. In the region of the lateral pn-junction between n-implant and spray, the dose is reduced by a factor 2-3 ("moderated") to improve the high voltage stability of un-irradiated devices.

In order to keep unconnected pixels at a well defined potential, the p-stop rings have openings allowing a controlled high resistive path between pixels. This path is important for testing of sensors before they are bump-bonded to readout chips. In case of individual bump-bonding failures it prevents the unconnected pixel from affecting the neighbors. The pixels in the p-spray technology are connected to a bias grid via a punch-through resistor for the same purpose. The resistor is formed by a small n-implant in the corner of each pixel that is separated from the pixel implant by a small gap.

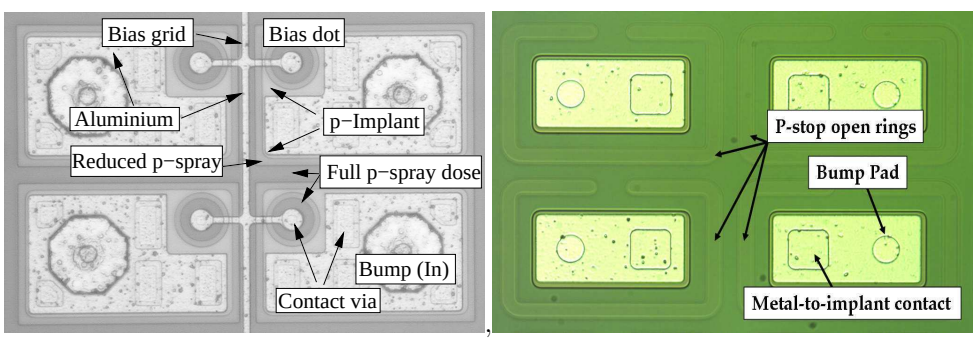

Fig. 1.4. Photo of four pixel cells with moderated p-spray (left) and open p-stop design (right). The pixel size is $100\mu m \times 150\mu m$.

The area between the pixels and the edge of the sensors is covered by an n-implant and connected to the chip ground via bump bonds. This ensures that the sensor edge is kept at ground potential and drains leakage currents from the guard ring region. On the back-side the high voltage is dropped over a series of guard rings between the edge and the p-implant of the active region. The distance between the outermost pixels and the edge of the silicon is 1.2 mm, which is small enough to allow wirebonding to the readout chips. The barrel sensors are arranged without overlap in the z-direction and to minimize the dead region an even smaller distance of 0.9 mm was chosen for those sides.

The high voltage stability of the sensors exceeds the maximum operating voltage of 600 V of the installed high voltage supplies. The detectors are going to be operated below room temperature to reduce the effects of reverse annealing. Oxygen diffusion as recommended by the ROSE collaboration[21] was applied to reduce the depletion voltage of highly irradiated sensors. Both types of sensors have been shown to be functional up to the original target fluence and beyond. A detection efficiency above 95% was observed after a fluence of $1.2 \times 10^{15} n_{eq}/\mathrm{cm}^2$ with a detection threshold of 3000 electrons.[22]

1.4. Front-end electronics

The front-end chip is a key component of the pixel detector and its development accounts for a large fraction of the detector R&D. It's basic tasks are:

- registering the signals produced by particles in the sensor
- storing time, position, and the amount of collected charge of all channels during the trigger latency
- sending out data for bunch crossings selected by the first level trigger

A full readout of all channels, like it is used in the strip tracker, is ruled out by the large number of pixels. Zero-suppression at a very early stage is mandatory because buffering or transferring every channel for every bunch crossing is impossible. Ignoring signals of insignificant magnitude reduces local data rates by more than 3 orders of magnitudes. However, good control of detection thresholds is needed in order to maintain efficiency and – because charge sharing is exploited – resolution of the detector. Active compensation of inevitable pixel-to-pixel and chip-to-chip variations complicates the design and operation of the detector considerably.

The size of the readout chip is limited by the manufacturer's reticule and yield considerations and is coupled to the size of the detector module. Multiple readout chips need to be grouped together to form modules of practical size. The floorplan is fixed by the need to tile chips without insensitive gaps. It is a matrix of pixel cells extending all the way to the edges of the chip on three sides. All auxiliary components, including input/output, are in the periphery on only one side (Fig.1.5). High volume data-flow from the pixels to the periphery suggests a column oriented design.

The required radiation hardness originally seemed to make the use of a dedicated radiation hard process mandatory. Early developments started with the Honeywell and DMILL process. The DMILL technology is a 0.8μm process with two metal layers. Compared to more recent deep sub-micron technology those processes had large device sizes and very limited connectivity. The initial choice of the pixel size and readout architecture were made with these limitations in mind. To ensure the feasibility of the full readout chip, all layouts of test devices and prototypes were done from the beginning with radiation hard processes.

The transition from DMILL to a 0.25μm process brought a number of advantages. The smaller feature size is only one of them and it was deliberately not fully exploited. The translation permits roughly a factor 6 increase in device density even when the requirements of radiation tolerant layout, like ring gates and guard rings, are taken into account.[13] Instead of consequently using minimal layout rules and highest device count, the design emphasized good yield and compatibility with multiple vendors. It was still possible to reduce the pixel size from 150μm $\times$ 150μm to 100μm $\times$ 150μm. In view of the cluster sizes, no advantages were expected from reducing the pixel size further. The size of the periphery, on the other hand, was reduced significantly from 2.8 mm to 1.8 mm despite an increase of buffer capacity.

Five metal layers and finer trace pitches permit significant further improvements, although the use of the top metal layers is severely restricted by fill factor limitations and the need to completely shield the sensor from cross-talk.

While the DMILL chip was never intended to be operated in the 4 cm layer at full design luminosity, the improvements allowed by the 0.25μm technology made this possible without a fundamental redesign of the readout architecture.

Last but not least, the power consumption of the chip went down by a factor of 4. The PSI46 consumes 30 μW per pixel.

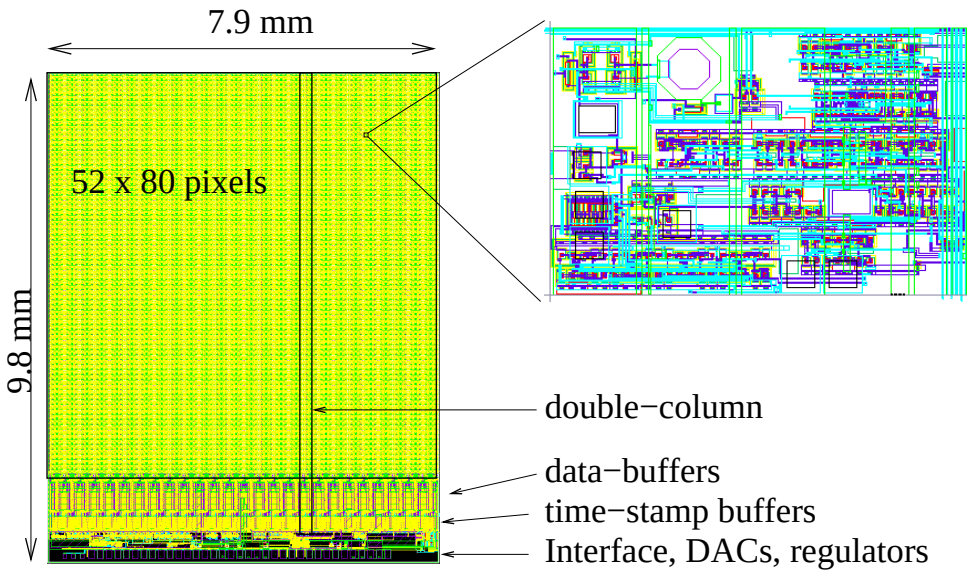

Fig. 1.5. Floorplan of the PSI46 readout chip. The active matrix of 80 $\times$ 52 pixels has a size of 8 mm $\times$ 7.8 mm. The periphery has a length of 1.8 mm, dominated by the size of buffers. Also shown is the "left" pixel of a double column. The other half is essentially mirror symmetric with a common vertical bus in between. The analog section with the bump-pad is in the left part of the pixel, as far away as possible from the bus with its frequent digital activity.

1.4.1. *Analog section*

The smallness of the input capacitance of the pixel makes the intrinsic noise of a charge sensitive amplifier almost irrelevant. More important than the analog quality of the amplifier is the speed needed to separate the LHC bunch crossings. Maximizing speed requires maximizing transconductance for a given current and minimizing parasitic capacitance where it slows down the amplifier.

The amplifiers are push-pull stages. The transconductance of $p-$ and $n-$transistors adds up and is higher than in other configurations. The input transistors operate in the weak inversion regime where the transconductance depends only on the drain current I_D and g_m/I_D is maximal. A push-pull amplifier is in principle quite susceptible to power supply ripple. However, with the on-chip regulators the system is sufficiently robust. The working point of the amplifiers is in the middle between the power rails and independent of voltage drops across the chip because the analog power rails are kept symmetric and separate from other power rails. This minimizes systematic variations of thresholds and feedback transistors.

The charge sensitive pre-amplifier has sufficient gain for signals of a few thousands of electrons when the integrating feedback capacitance is a few fF. For larger values, the gain is reduced but the integration is faster (or the same speed is achieved with lower current). In terms of the total power, a solution with a rather large feedback of $C_f = 20$ fF and an additional gain stage was preferred over a single-stage design with small C_f.

The pre-amplifier input is DC coupled to the sensor pixels. Its feedback must absorb the expected sensor leakage current of 10 nA per pixel. Active and passive feedback networks have been considered. The value R_f of a passive feedback must be small enough to absorb the leakage current without causing an offset of more than ~ 100 mV. On the other hand it must be large enough to keep the shaping time ($R_f C_f$) bigger than the amplifier risetime to avoid loss of signal. Both requirements are fulfilled by a $\sim$MΩ resistor. Passive feedback was therefore chosen because of its simplicity and robustness. The resistor is implemented as a weak p-transistor[c] operating in its linear region. The working point of the push-pull stage is at one transistor threshold voltage above analog ground, such that the feedback resistor is properly biased when the gate is near ground. Both gate and substrate potential are made adjustable to control the resistance of the feedback resistors.

The second stage ("shaper") has the same push-pull design and is AC-coupled to the pre-amplifier. The AC coupling serves a dual purpose. It removes offsets caused by leakage current and it is part of the gain stage. The gain is the ratio of coupling-capacitance to feedback-capacitance.

The shaper output is connected to the comparator and the sample-and-hold circuit. A buffer in front of the sampling capacitor reduces the load of the shaper output (not shown in Fig. 1.6) which is critical to keep the input to the comparator

[c]such weak n-transistors are not possible in radiation hard layout

fast. The comparator threshold is adjustable with an eight-bit DAC. Production related random variations of transistors leads to pixel-to-pixel signal threshold variations, but good threshold uniformity is essential to obtain a low global threshold without driving a large number of pixels into saturation. Additional four-bit DACs in each pixel can compensate these variations and make fine-adjustments for individual pixel thresholds. The un-trimmed threshold dispersion in the 0.25μm chip is approximately 450 electrons which can be reduced by trimming to 80 electrons.

The analog pulse-height information is stored for later readout on the sample-and-hold capacitor. There is no bunch crossing clock in the pixels and the sampling point is simply given by a delayed version of the comparator output. Non-linearities related to this and other effects are compensated off-line. Only a modest accuracy equivalent to 4 bits is needed because charge fluctuation dominate the resolution obtained by interpolation.[23] Saturation of the amplifier occurs near 2 times the average signal of a minimal ionizing particle. Larger signals occur frequently but are not useful for interpolation because they are due to long range delta-electrons.

The finite risetimes in amplifier and comparator lead to time-walk. Larger signals are detected earlier than small ones. The peaking time is slightly larger than the bunch crossing separation which means that not all pulses will reach the detection threshold within the same bunch crossing. This determines the effective threshold for useful signals. For a nominal threshold of 3000 electrons the effective threshold is 3800 electrons with an analog current of 5 μA per pixel.

1.4.2. *Readout architecture*

Minimizing the logic inside pixels and the number of bus lines in the pixel array was essential for the original DMILL layout and the readout architecture developed for the PSI43 chip reflects this. It was successfully implemented and verified in a beam test. Although those requirements are less stringent for the 0.25μm technology, the readout architecture described here is essentially unchanged. Some of the data-losses that were consciously accepted in the DMILL chip were reduced when it was possible with small modifications.

The latency of the first level trigger in CMS is approximately 4 microseconds, corresponding to 150 bunch crossing intervals of 25 ns. At high luminosity the probability of more than one hit occurring within the trigger latency in the same pixel can be several percent. Keeping the hits inside the pixel during the latency would require a buffer to avoid unacceptable deadtimes. It would also require the distribution of bunch-crossing numbers throughout the chip to allow the association of hits and bunch-crossings. Placing all buffers in the periphery, outside of the sensitive region of the sensors allows a simpler (and hence smaller) pixel cell and makes more efficient use of the buffers because they can be shared by more pixels. The data rate into the buffer should not exceed more than a few percent of the transfer capability to keep waiting times of pixel cells small. Simulations show that grouping two columns of pixels leads to manageable rates per buffer. The readout

chip is therefore organized into double-columns of pixels that operate essentially independently. The pixels communicate the detection of hits over a wired-OR. No clock or bunch crossing numbers need to be distributed over the pixel array.

The periphery synchronizes the wired-OR with the LHC clock and latches the current bunch crossing number in a time-stamp buffer whenever a hit is found. The state of the pixels that have a hit at that time is frozen and their data is subsequently collected. A token passing from pixel to pixel controls the transfer. Even though empty pixels are skipped relatively rapidly (asynchronously), finding the next pixel with a hit can take longer than one bunch crossing. Hit pixels therefore must re-synchronize with the periphery. The handshake reduces the transfer speed to 20 MHz. However, in most cases the number of hits per column is small and the total transfer time has a significant contribution from skipping empty pixels anyway.

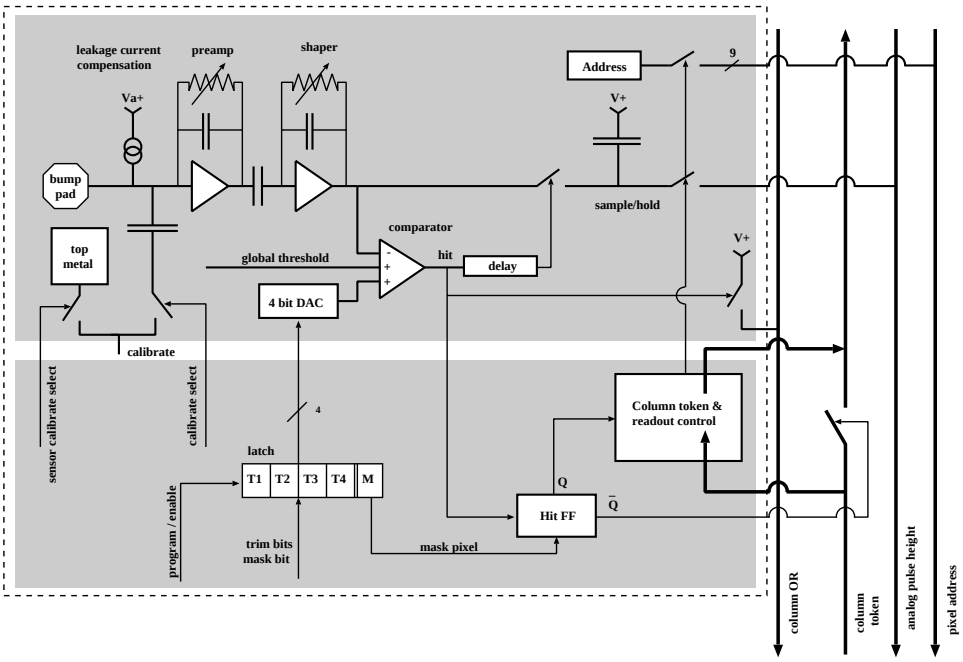

Fig. 1.6. Schematic view of a pixel cell.

The data transmitted and stored in the data-buffers are the pixel (row) address and the analog pulse-height. Marker bits in the data buffers keep track of the association between time-stamps and hits. Maintaining this synchronization under all conceivable circumstances is crucial. A flaw that could cause the loss of a marker when an event was triggered while the buffer was just overflowing was discovered during high rate test-beam and fixed in the production version of the PSI46. High rate beam tests were an important part in the development cycle of the readout chip because the effect of random hit patterns on a chip with data-driven architecture

cannot be tested fully in the laboratory.

Hits and time-stamps are stored in the buffers during the latency of the trigger. If no trigger signal arrives when a time-stamp expires, the corresponding buffer cells are again available for new hits. The buffers overflow when more than 12 time-stamps or more than 32 hits occur during the trigger latency, in which case some data are lost. If a time-stamp is triggered, the double-column stops data-taking and waits for readout. More details about the detector readout are given in Section 1.6.

1.4.3. *Control*

During data-taking, the readout chip is controlled by the bunch crossing clock and the trigger. All control and configuration data are transmitted by differential lines similar to LVDS at 40 MHz. As usual in CMS, a trigger signal is three clock cycles long and can encode different commands. In addition to the first level trigger this is used for soft resets and the injection of calibration signals. Configuration data uses a serial protocol similar to I2C but operating at 40 MHz. The fast transfer was chosen in order to keep the detector setup-times short and to permit frequent reloading of the configuration in case that single event upset (SEU) should make this necessary.

SEU was originally assumed to be a problem but is now expected to be much less serious. A simple protection scheme using a capacitive load in the feedback of the storage cells has reduced the vulnerability by two orders of magnitude. The SEU cross section of such protected storage cells has been measured to be $\sigma = 2.6 \cdot 10^{-16}\text{cm}^2$ for 300 MeV/c pions. Less than one such storage cell per second is expected to be affected at full LHC luminosity.

Pixel thresholds and masks are the largest part of the configuration data. Chip-wide DAC settings and control registers constitute only 1 % of it. A direct readback of the configuration is not foreseen. DAC adjustments can be verified using the chip header in the readout data (Section 1.6).

A lot of adjustability has been retained for internal operating points and amplifier bias currents. Adjusting the 26 DACs complicates detector operation but it ensures enough flexibility to compensate process variations and to adapt to ambient temperature variations or radiation effects. Most of the DACs control the analog front-end and the analog readout chain. Others control supply voltage regulators and the delay and magnitude of calibration signals. The DACs are referenced to an on-chip, temperature compensated, band-gap reference.

1.5. Modules

The modules of the pixel barrel are shown in Figure 1.7. A full module consists of two rows of 8 readout chips connected to one sensor. This module size permits efficient use of the 4” wafers used by the sensor vendor.

The sensor elements at the chip boundaries have twice the normal width to

avoid dead regions between readout chips. Pixels in the corners are four times larger. The readout chips extend 0.8 mm beyond the sensor edge allowing access to the wire-bond pads. Wire-bonds step up from the readout chip to a thin high density interconnect (HDI) glued onto the back-side of the sensor. Base-strips are glued underneath the readout chips. They allow mounting of modules on the barrel mechanics with small screws and carry no components. The base-strips are made of silicon nitride with a thickness of 250 μm. Silicon nitride has excellent mechanical properties and a coefficient of thermal expansion that matches very well that of silicon. The 4160 bump-bonds per chip have sufficient strength to hold the chips in place and no further mechanical parts are needed.

Half-modules with only one row of 8 readout chips are needed at the boundary of the barrel half-shells to avoid a gap (Fig. 1.10). The three barrel layers consist of 672 modules and 96 half-modules.

Power, control signals, and readout are distributed over the three-layer HDI. The power and ground layers at the bottom reduce coupling of the fast switching signals into the sensor. All digital signals on the HDI are differential with 300 mV amplitude. Un-terminated lines and low-power drivers minimize the power consumption connected with signal distribution. A module controller ASIC, the TBM,[24] receives all external control signals and distributes them to the readout chips. It also coordinates the readout of the 16 readout chips as described in Section 1.6.

Signals to and from the module and the readout are transferred over a 2 layer Kapton/copper compound cable with 21 traces and 300 μm pitch. Unacceptable cross-talk from the digital signals into the analog readout on the cable have led to the use of a very low voltage swing protocol using only 80 mV. Low voltage power and detector bias are brought over 6 copper coated aluminium wires of 250 μm diameter.

A completed full-module has the dimensions 66.6×26.0 mm^2. The power consumption of such a module is $\sim 2W$. It weighs 2.2 g plus up to 1.3 g for cables. The cable length varies between 5 cm and 40 cm depending on the position of the module in the barrel.

The material of the pixel barrel amounts to 5 percent of a radiation length in the central region. Sensors and read-out chips contribute one third of the material, while the support structure and cooling fluid contribute about 50 percent.

Each disk of the endcap detector has 24 wedge shaped segments or blades. The blades extend from 6 cm to 15 cm in radius and have a width between 3 cm and 5 cm. Full coverage of the blade area is achieved by an arrangement of rectangular sub-assemblies (plaquettes) of varying size mounted on two panels as shown in figure 1.12. Two panels are mounted on either side of a cooling channel such that each one covers the insensitive regions of the other. From the point of view of readout and control, a panel is the equivalent of a barrel module with one TBM and up to 24 readout chips. The four disks are made of 96 panels populated by 672 plaquettes

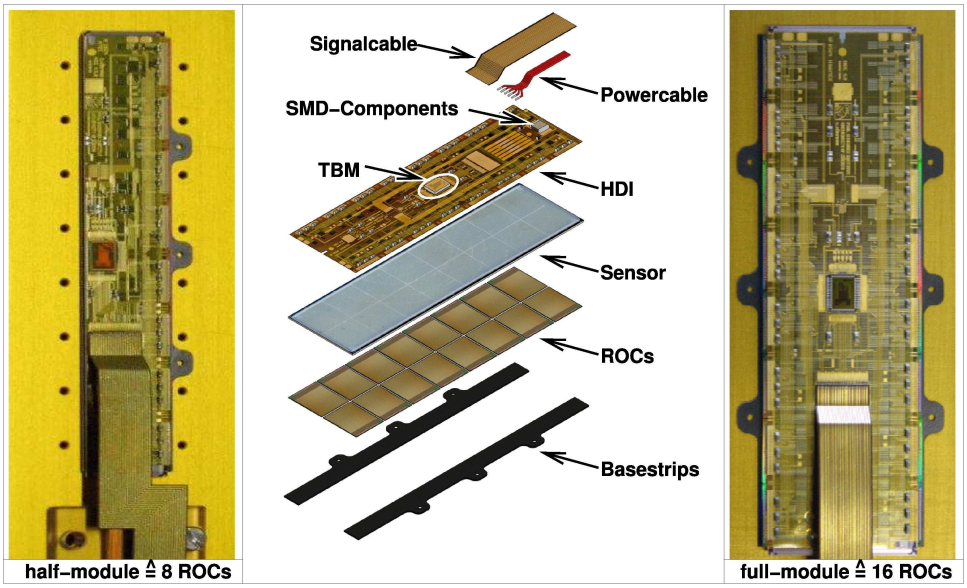

Fig. 1.7. Components of the CMS pixel barrel modules. The full-module (right) has 16 readout chips (ROC) and size of 66 mm × 26 mm. Half-modules (left) have only one row of ROCs and minimize the insensitive region near the vertical split of the detector.

of five different sizes.

Bump bonding of sensors to readout chips was done in-house at PSI using indium bumps[8] for the barrel modules and in industry with Pb/Sn bumps for the endcaps.[25] The in-house process was particularly valuable for evaluating test devices and prototypes.

1.6. Readout

The readout of the pixel data is based on the analog optical links of the CMS strip tracker. Unlike the strip detector data, the pixel data are sparsified which means that pixel addresses need to be transmitted along with the charge information.

For efficient link usage, the addresses are not binary coded, but use a set of 6 discrete analog levels ($\sim$ 2.5 bits/clock). A hit in a given chip is encoded in a sequence of 5+1 clock-cycles: 2 for the double-column, 3 for the row, and one for the pulse-height. Between 8 and 24 chips are serially read out over a common link. The number of chips reflects the expected hit rate in the respective part of the detector.

The serial readout is controlled by a token going in a fixed order from chip to chip. The chip that has the token transmits all hits for a given trigger and then passes the token to the next chip. Each chip adds a header, regardless of whether it actually has hits to send or not. For very low occupancy, this means a significant

overhead, but it is efficient at high occupancy ($\sim$ 1 hit per chip) and avoids the need for transmitting chip IDs.

Inside the chip, the readout is again controlled by a token that goes from double-column to double-column skipping those that are empty. The length of the header, three clock cycles, accommodates skipping through a full chip if necessary. The first two header cycles are used to create an unambiguous pattern identifying chip headers. The third cycle can carry information about DAC adjustment. Figure 1.8 shows the readout of a single chip with exactly one pixel hit (generated by calibration signals).

Triggers and readout tokens are both counted and hits are only read out when the token number matches the readout number: Every token pass collects hits for exactly one trigger. The need for transmitting time-stamp information and event-building is avoided, but it must be ensured that exactly one token for every trigger is issued and that there is never more than one token. This is the task of the Token Bit Manager chip (TBM) on the modules/blades. For every first level trigger it sends a token to the first chip and waits until the token returns from the last chip in the chain. It keeps track of triggers arriving while the readout is still busy with a FIFO. The TBM adds a header with an event number and a trailer with status information to each readout. The most important status information during data-taking is token FIFO overflow warning. A loss of a token destroys the synchronization between module and data-acquisition, effectively leading to the loss of all subsequent data. The trigger counter is used to verify the correct synchronization. A full readout is shown in Figure 1.8.

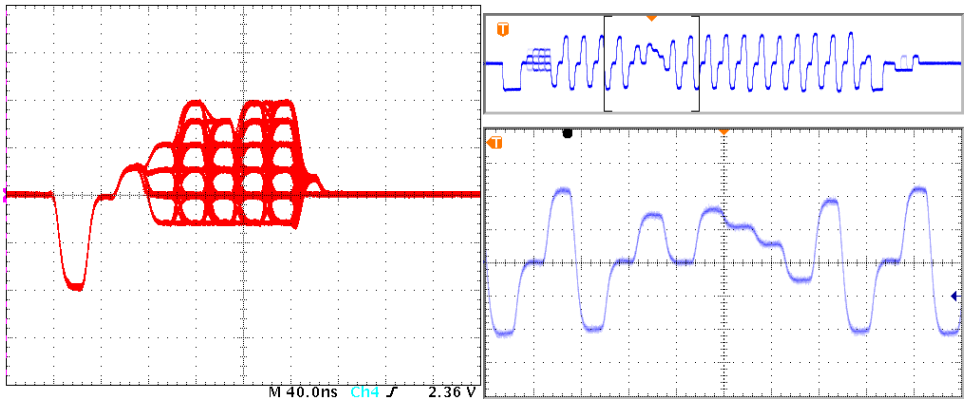

Fig. 1.8. An "Eye diagram" of the readout of a single chip with one pixel hit is shown on the left side. The horizontal axis is time, the vertical axis shows the amplitude of the chip output. All possible pixel positions are scanned and the oscilloscope traces superimposed. The first three 25 ns cycles are the chip header. It is always the same. The following five cycles encode the column and row of the pixel. The last cycle represents the pulse-height, which is constant in this case. The figure on the right shows the readout of a barrel module with a TBM and 16 chips. The TBM header contains an event counter. The fifth chip has one hit, the corresponding section is shown enlarged at the bottom.

The TBM can manage two readout groups in parallel and independently. This is used in the barrel modules of the inner two layers where two readout fibers are needed to have enough bandwidth. A single fiber is sufficient for the third barrel layer and the endcap panels. The TBM is configurable by software such that the same module type can be used in all barrel layers.

The output of the readout chip is differential electrical. The TBM serves as an analog multiplexer and the transmission continues over differential lines for distances up to $\sim 1m$ to the analog optical hybrid where a custom made level translator shifts and amplifies the signal to match the input specifications of the laser driver. The optical signals are received, digitized, and decoded by Front End Drivers (FEDs), after some 40 m of optical fiber.

The robustness of the analog coded addresses was a big concern. To reduce the amount of data, the FED decodes the levels according to predefined address-level windows immediately after digitization. Offline corrections are not possible, the raw data is accessible only in test runs.

The transmission chain is completely transparent from the readout chip to the FED. The accumulated risetime of all transmission stages had to stay below a few ns in order not to spoil the level identification. Impedance matched transmission lines were used at all stages to minimize distortions. The output levels can be adjusted by DACs in each chip to reduce chip-to-chip variations. In addition, the level decoding by the FED is calibrated on a per chip basis. Pixel-to-pixel variations were a problem in the DMILL chip but have been eliminated in the 0.25μm chip. Baseline drifts of the optical link, mostly due to small temperature variations, are continuously monitored and compensated in the FED. A typical separation of 26:1 for the full readout chain (Fig. 1.9) makes the analog coded address transmission robust and reliable. The trigger number and status information in the TBM header and trailer use a similar encoding.

The noise added by the analog transmission to the pulse-height information corresponds to 300 electrons which is negligible for the position resolution.

1.7. Mechanics and Installation

Relatively frequent access for replacement and beam-pipe bake-out has always been a part of the concept of the CMS pixel detector. Mechanical structure and insertion procedure are designed to permit extraction and re-insertion of the detector within a relatively short time and in the presence of the LHC beampipe.

The pixel system is installed inside a cylindrical volume with a diameter of 42 cm and a length of 5.6 m formed by the inner wall of the CMS strip tracker. The beam pipe is cylindrical in the central region followed by a conical section. It is fixed to the CMS detector with vertical supports at $z = \pm 88$ cm.

In order to permit installation of the pixel detector in the presence of the LHC beampipe, the whole pixel system is split vertically and mounted on wheels that

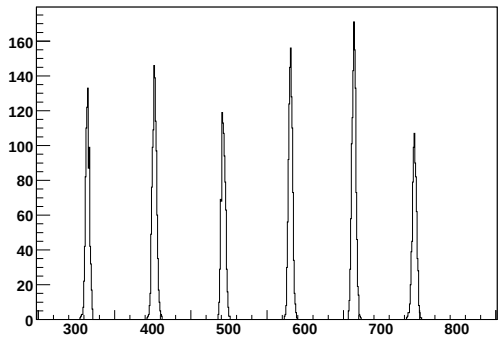

Fig. 1.9. Address level histogram as received by the FED. The horizontal axis is n ADC units corresponding to the analog level. Robust decoding is possible with the RMS to distance ratio of 26:1.

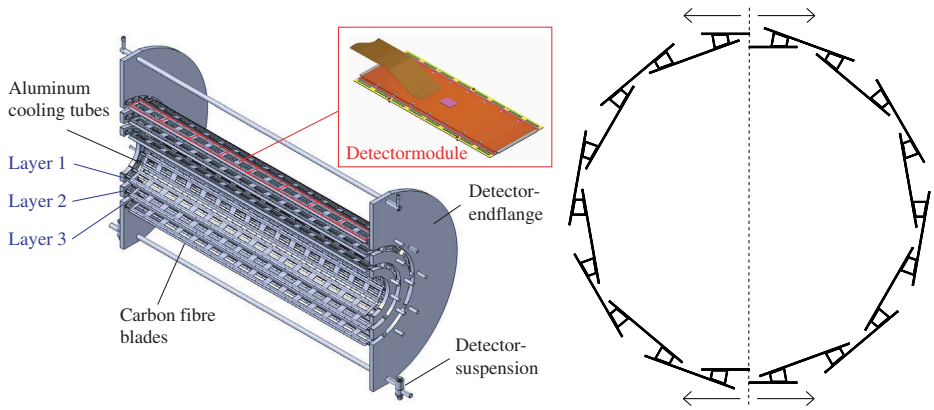

Fig. 1.10. Schematic view of the innermost pixel barrel layer. The modules are mounted with alternating orientation (inside/outside) onto carbon fiber blades held by trapezoidal cooling tubes. Half-modules allow joining the two shells without gap.

are guided by grooves in the floor and ceiling of the support tube. The grooves for the left and right halves are parallel in the center and diverge near the end of the tracker volume to avoid collisions with the conical section of the beam pipe.

One half shell of the barrel mechanics[26] is shown in Figure 1.10. The barrel structure consists of the aluminium cooling tubes connected by thin carbon fiber blades. The blades are glued alternating to the inside and outside of the tubes such that adjacent blades have a small overlap. Eight modules are mounted on each blade, four are read out to the $+z$-side and four to the $-z$-side. The cooling tubes are attached to cooling manifolds embedded in the bulkheads. Wheels for installation are mounted at the top and the bottom of the bulkheads and two carbon fiber rods between the wheels at the $\pm z$ side protect the barrel from horizontal forces during

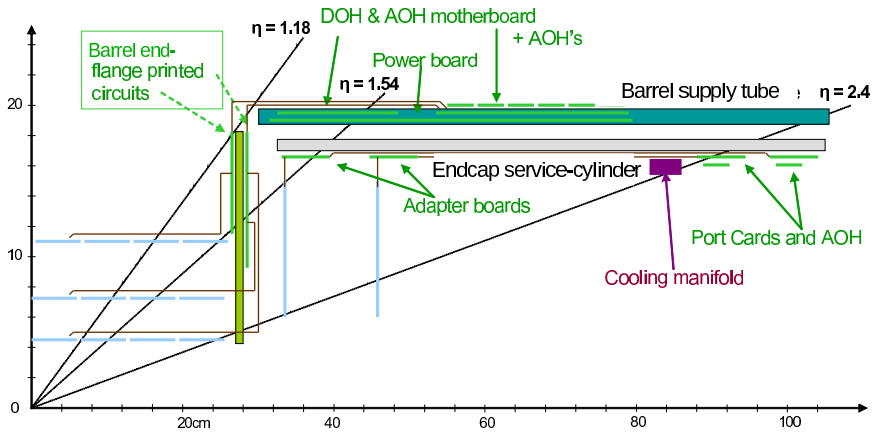

Fig. 1.11. Schematic view of one $r - z$ quadrant of the pixel detector.

insertion. The structure can be separated into three layers for the mounting of the modules. So-called supply tubes with a length of 2.2 m and a thickness between 1 cm and 3 cm bring all services from patch panels outside the tracker volume to the barrel bulkheads. The volume inside of the supply tube is used by the pixel endcaps (Fig. 1.11).

No access to the detector is possible in its final position, all connections must be made before insertion. All connections between barrel and supply tubes must be flexible because the angle changes by a few degrees during insertion. The cooling connections are made with fiberglass reinforced silicone tubes. All electrical cables have enough slack to provide the required flexibility. The electrical/optical conversion is done on the supply tube and no optical fibers go to the barrel.

The endcap detector is divided into four independent half-cylinders for installation reasons. Two half-cylinders are installed from each z-side on a rail system similar to that of the barrel detector. The blades of the endcap detectors are mounted onto aluminum cooling channels that define the turbine geometry (Figure 1.12). Connections to the patch panel are made via service cylinders on which also the optical links and auxiliary electronics are mounted. The endcap detector is inserted after the barrel into the volume inside the barrel supply tube.

1.8. Power

The nominal operating voltage of the 0.25μm process is $2.5V$. The digital logic including the asynchronous parts is still sufficiently fast when operated above 2 V. A similar supply voltage is needed for the discriminator and all parts of the analog readout chain with the exception of pre-amplifier and shaper. A single external supply voltage of approximately $2.5V$ is used to generate these internal voltages with adjustable on-chip regulators. The internal operating voltage of pre-amplifier

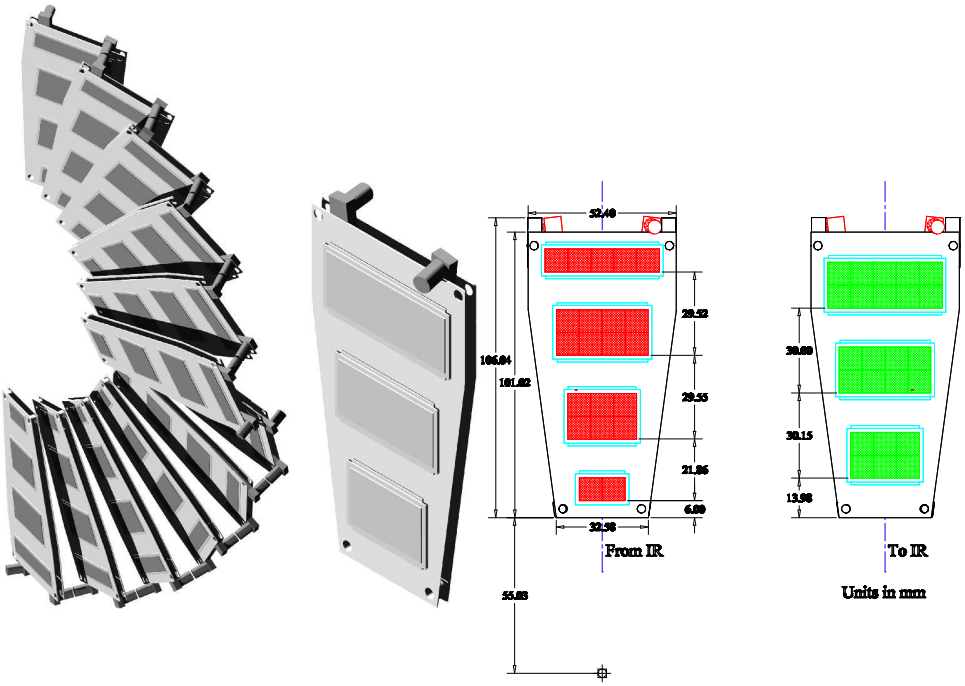

Fig. 1.12. Endcap detector geometry. The left figure shows one half-disk with 12 blades. A blade consists of the cooling channel with detector panels on either side. The gaps between sensor plaquettes on one panel are covered by plaquettes on the second panel.

and shaper is $\approx 1.2V$, much lower that that of the digital section. Deriving this from a separate supply (1.5 V) is more efficient in terms of power consumed in the front-end. A separate analog supply helps isolating the sensitive analog sections from the digital parts. Internally to the chip this isolation is maintained by keeping separated analog and digital nets for power and substrate connections.

Groups of typically 12 modules are powered by a common supply. A voltage drop of $\sim 3V$ over the 40 m of cable is compensated using sense wires at the end of the supply tube. The remaining voltage drop between that point and the modules varies between 100 mV and 400 mV depending on the length of the module cable and the power consumption of the module. The on-chip regulators provide well defined internal operating voltages and reject ripple on the power lines up to 150 kHz.

While the analog power consumption is very constant, the digital power consumption has fast transients and varies substantially with time due to the data-driven architecture. During gaps in the bunch structure of the LHC that last up to 3 μs the power consumption drops up to 30 %. Every readout chip has two external 1 μF SMD capacitors on the HDI that absorb such variations and avoid large voltage fluctuation or inductive effects on the cables. Cross-talk among readout

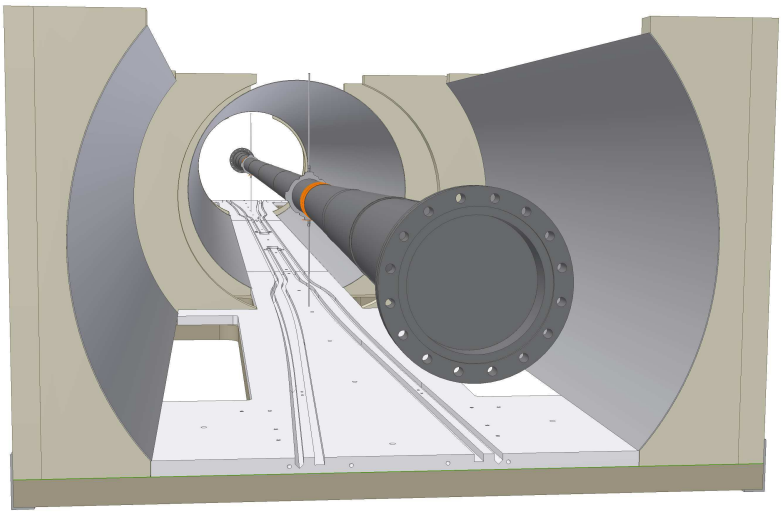

Fig. 1.13. Simplified drawing of the installation volume with beam pipe. The outermost pair of grooves is for the pixel barrel. The barrel is inserted from one side. The inner pairs of grooves are for the pixel endcaps which are inserted from both sides. The volume shown has a length of 5 m.

chips and modules coupling through the power connections was found to be small compared to chip internal cross-talk.

1.9. Conclusion

The challenges of precise tracking near the interaction point of the LHC required a huge step in detector technology. New levels of radiation, track rate, and track density demand a finely segmented pixel detector with a powerful readout architecture and careful design for radiation hardness. The CMS pixel detector has been installed and commissioned in 2008 after more than 10 years of development and construction time. It achieves excellent resolution in the r–ϕ and z direction with almost square pixels and interpolation based on charge sharing. Low occupancy and three-dimensional space-point measurements give the pixel detector a unique role in the tracking of the CMS experiment.

References

1. M. Della Negra et al., CMS: The Compact Muon Solenoid: Letter of intent for a general purpose detector at the LHC. CERN-LHCC-92-03.
2. CMS physics: Technical design report. CERN-LHCC-2006-001.

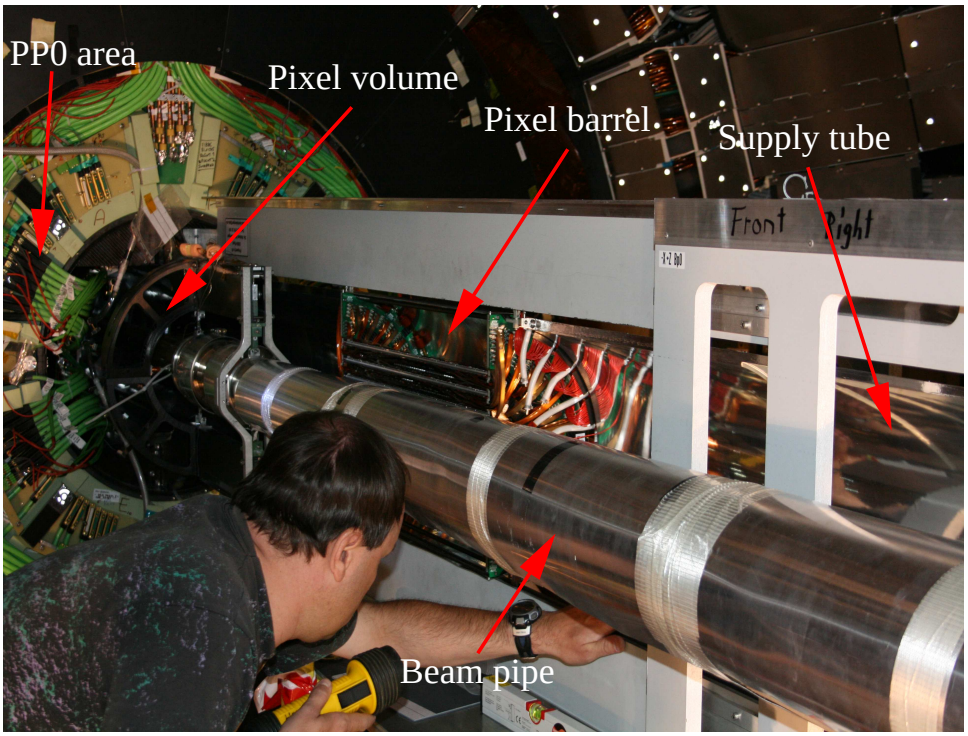

Fig. 1.14. Installation of a pixel barrel shell into the CMS detector.

3. S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, Track reconstruction, primary vertex finding and seed generation with the pixel detector. CERN-CMS-NOTE-2006-026.
4. B. Henrich, W. Bertl, G. K., and R. Horisberger, Depth Profile of Signal Charge Collected in Heavily Irradiated Silicon Pixels. CERN-CMS-NOTE-1997-021.
5. B. Henrich and R. Kaufmann, Lorentz-angle in irradiated silicon, *Nucl. Instrum. Meth.* **A477**, 304–307, (2002). doi: 10.1016/S0168-9002(01)01865-4.
6. R. Horisberger, Design requirements and first results of an LHC adequate analog block for the CMS pixel detector, *Nucl. Instrum. Meth.* **A395**, 310–312, (1997). doi: 10.1016/S0168-9002(96)00000-9.
7. M. Lechner, Development of a radiation hard pixel analog block for the CMS vertex detector and search for the rare decays $\Phi \to \eta e^+e^-$ and $\Phi \to \pi^0 e^+e^-$ at CMD-2. DISS-ETH-12866.
8. C. Broennimann et al., Development of an indium bump bond process for silicon pixel detectors at PSI, *Nucl. Instrum. Meth.* **A565**, 303–308, (2006). doi: 10.1016/j.nima.2006.05.011.
9. CMS, tracker technical design report. CERN-LHCC-98-06.
10. M. Barbero et al., Design and test of the CMS pixel readout chip, *Nucl. Instrum. Meth.* **A517**, 349–359, (2004). doi: 10.1016/j.nima.2003.09.043.
11. M. Dentan et al., Final acceptance of the DMILL technology stabilized at TEMIC/MHS. Given at 4th Workshop on Electronics for LHC Experiments (LEB

98), Rome, Italy, 21-25 Sep 1998.
12. W. Snoeys et al., Layout techniques to enhance the radiation tolerance of standard CMOS technologies demonstrated on a pixel detector readout chip, *Nucl. Instrum. Meth.* **A439**, 349–360, (2000). doi: 10.1016/S0168-9002(99)00899-2.
13. P. Fischer, Design considerations for pixel readout chips, *Nucl. Instrum. Meth.* **A501**, 175–182, (2003). doi: 10.1016/S0168-9002(02)02029-6.
14. H. C. Kastli et al., Design and performance of the CMS pixel detector readout chip, *Nucl. Instrum. Meth.* **A565**, 188–194, (2006). doi: 10.1016/j.nima.2006.05.038.
15. D. Kotlinski, Status of the CMS Pixel detector, *JINST*. **4**, P03019, (2009). doi: 10.1088/1748-0221/4/03/P03019.
16. L. Caminada and A. Starodumov, Building and commissioning of the CMS pixel barrel detector, *JINST*. **4**, P03017, (2009). doi: 10.1088/1748-0221/4/03/P03017.
17. A. Kumar, Commissioning of the CMS forward pixel detector, *JINST*. **4**, P03026, (2009). doi: 10.1088/1748-0221/4/03/P03026.
18. R. Adolphi et al., The CMS experiment at the CERN LHC, *JINST*. **0803**, S08004, (2008). doi: 10.1088/1748-0221/3/08/S08004.
19. G. Bolla et al., Design and test of pixel sensors for the cms experiment, *Nucl. Instrum. Methods*. **A 461**, 182–184, (2001).
20. J. Kemmer et al. Streifendetektor. Patentoffenlegungsschrift DE 19620081 A1.
21. G. Lindström, M. Ahmed, S. Albergo, P. Allport, D. Anderson, L. Andricek, et al., Radiation hard silicon detectors – developments by the RD48 (ROSE) collaboration, *Nucl. Instrum. Methods*. **A 466**, 308–326, (2001).
22. T. Rohe et al., Fluence dependence of charge collection of irradiated pixel sensors, *Nucl. Instrum. Meth.* **A552**, 232–238, (2005). doi: 10.1016/j.nima.2005.06.037.
23. The CMS Collaboration. CMS Tracker. Technical Design Report LHCC 98-6, CERN, Geneva, Switzerland, (1998).
24. E. Bartz, The 0.25-mu-m token bit manager chip for the CMS pixel readout. Prepared for 11th Workshop on Electronics for LHC and Future Experiments (LECC 2005), Heidelberg, Germany, 12-16 September 2005.
25. P. Merkel, Experience with mass production bump bonding with outside vendors in the CMS FPIX project, *Nucl. Instrum. Meth.* **A582**, 771–775, (2007). doi: 10.1016/j.nima.2007.07.089.
26. C. Amsler et al., Mechanical Design and Material Budget of the CMS Barrel Pixel Detector, *JINST*. **0904**, S05003, (2009). doi: 10.1088/1748-0221/4/05/P05003.