Status and Perspectives of Silicon Detectors

Alberto Cervelli†

1 Istituto Nazionale di Fisica Nucleare - Sezione di Bologna, Viale Berti Pichat 6/2, 40138 Bologna, Italy

Abstract. In the following a review of the present status of silicon tracking and vertexing systems and their future developments will be presented. We will show the modern detectors used in present day experiments both in nuclear and elementary particle physics, and their achieved performances. Later we present a review of the near-future systems which are now being designed, built, or commissioned together with an outlook on the future developments for next-generation silicon detectors.

1 Introduction

Silicon detectors have been used in tracking since the start of the 80s. The development of silicon detectors in high energy physics followed closely the one in the microelectronic industry, which share its technology and design with the detectors used nowadays for tracking and vertexing.

What should be kept in mind is that for tracking purposes, and especially for reconstructing secondary vertices, silicon trackers have to be placed very close to the interaction point, in order to achieve the best results, should it be on a fixed target or near colliding beams. Their operating environment is always characterized by a large number of particles traversing the detector. This results in mainly two requirements for any silicon tracker: the need for a finely segmented detector, to cope with large number of hits, without reducing the performances due to occupancy or dead-time, and the possibility to withstand incredibly high radiation doses. These two requirements need to be tailored to the particular need of each experiment, both in terms of its physics case, and depending on the environment provided by the accelerator in use. For example hadron colliders have higher dose-rates compared to electron-positron colliders due to the production of baryons, conversely electron-positron colliders provide harsher conditions for the on-detector electronics caused by high fluxes of photons produced by synchrotron radiation.

The first silicon trackers had active elements connected via thin wires to peripheral readout, and then, from 1985 onwards, larger active silicon sensor were produced thanks to very large-scale integration (VLSI).

During the 90s and early 2000s many new and interesting results were obtained thanks to the new development in silicon trackers: the results obtained by the experiments taking data at LEP[1], with precision measurement at the Z resonance, the top discovery by CDF[2], or the measurement of indirect and direct CP violation in the B-sector by BaBar and Belle experiments[3]. Today the currently operating detectors in the LHC experiments [4–7], and B-factories[8] reached unprecedented precision, complexity, the possibility to continue working without a reduction in performances in environments with tremendous amount of radiation, and to record a large amount of data with great speed.

Silicon detectors for tracking evolved in complexity, and reduced their production cost together with the development of smaller pitches used in the microelectronic industry: this allowed to have an exponential increase in the surface covered and with the increase in density in the electronics it allowed for a larger number of channels, high read-out speeds, and always better performances. The evolution of the covered surface of silicon trackers in time is shown in figure 1.

Figure 1. Evolution of silicon trackers surface over the last 4 decades, both silicon strips and pixels had an exponential increase in covered areas and number of channels.

The evolution on both the sensor and their electronics is ever evolving, with many new improvement foreseen for silicon trackers in the near future: after almost 50 years,
silicon tracker systems are still one of the most advanced detectors present in many high energy physics detectors.

In the following discussion we will focus on pixel detectors for tracking, as they represent one of the most and active fields of research in tracker detector development.

2 Different Designs

To better understand how each detector, and so its design, should be tailored to the environment it is bound to work in a good example is provided by the different designs of the Belle2 tracker, and the ongoing design of the EIC tracking system.

Despite a very different physic goal, both EIC and Belle2 need to measure the $D^+ \rightarrow D^0 \pi^+ \rightarrow (K_S^0 \pi^0)\pi^+_K$; however in the case of EIC, an electron-ion collider, the main backgrounds arises from collision products, resulting in high charged particles multiplicity, leading to the requirement for high granularity. On the other hand, in the case of Belle2 the background is dominated by beam induced signal, which instead requires a better timing resolution, rather than better granularity.

A similar situation arises from the design of the beam line and the spectrometer magnet: EIC foresee a larger beam pipe and higher B-Field, requiring a better spatial resolution for tracking especially at low transverse momentum. Belle2 conversely has a smaller beam pipe, weaker magnet so the momentum resolution depends less strongly on the spatial resolution, on the other hand its electronics has to withstand a large amount of radiation due to the closeness of the system to the beam.

In conclusion, even the optimization for the detection of the same process will need different approaches in the design and the research of a particular tracking system, so in sections 5-7 we will provide a description of the building blocks for the future generation of pixel tracking systems.

3 The LHC-era

The LHC experiments provided one of the best benchmarks for tracking systems working in high radiation and high charged track multiplicity environments, the three 4-π experiments at LHC (ALICE, ATLAS, CMS) shared a similar approach in their design for the first runs of data-taking. These detectors have tracking systems with an innermost part instrumented with pixels and an outermost part with silicon strip sensor, all designed to survive doses in excess of $10^{12}$n.eq. Due to the differences in their respective physics case ATLAS and CMS have systems able to reach a resolution of $O(10µm)$, while ALICE system has a poorer position resolution $O(10mm)$, but can cover a wider range of transverse momenta.

In the coming years these three systems will undergo an upgrade to copewith the higher beam intensities, and the resulting higher luminosity provided by the LHC. For the pixel sub-detectors of ATLAS and CMS the upgrade foresee a similar approach: the pseudorapidity coverage of the system will be extended, with a detector surface increase following the increase of coverage, the technology for the pixel sensors will not change, relying on hybrid pixels with new design for both sensor and front-end electronics. The goal is to achieve better performances, or at least on-par with previous run of data-taking in harsher pile-up conditions[9]. ALICE, on the other hand, will undergo a major change in its tracking system, replacing its innermost tracker with a monolithic pixel detector. The new detector will be the first to be used in such a large scale in a high energy experiment at a collider[10], increasing both the position resolution performances. Stand-alone tracking at low momenta will be possible thanks to a lower material budget provided by the thin monolithic sensors.

4 Pixel architectures

In order to understand the status and the research and development of silicon pixels for the future generations of high energy physics experiments it is important to have a look at the two possible approaches leading to a sensor and front-end assembly. The two broad families constituting pixels are the hybrid and monolithic architectures for pixel design.

Hybrid pixels are characterized by having the sensor built on a different silicon layer than the front-end electronics. The two silicon substrates are usually interconnected trough the use of bump-bonding (i.e. the deposition of small amounts of conducting material between contact pads on the two substrates). Monolithic pixel instead take advantage of the modern lithographic processes allowing to make deep wells in the silicon bulk, where the front-end electronic is placed. A schematic view of both Hybrid and monolithic sensors is shown in figure 2.

![Figure 2. Sketch design for Hybrid pixel (left) where the second substrate harboring the FE is visible on top, and monolithic pixels (right), where the electronics is placed on the same substrate as the sensor.](https://doi.org/10.1051/epjconf/202329010007)
the budget material of the detector requiring O(500µm) of silicon for each module, increasing the multiple scattering, so making the detector less resolute for low momenta particles. On the other hand, hybrid modules allow for the use of high resistivity bulk in the sensor layer, largely increasing the radiation hardness of the sensor, without affecting the front-end performances, for the lack of technology commonality also the size of the full module is not strictly limited in surface, as it would happen for instance in a fully CMOS technology sensor, allowing for large area coverage.

In Monolithic sensors front-end and sensor are designed on the same substrate: this requires a commonality between the design of the two components, usually CMOS technology: the price to pay to avoid interconnection, and reduce the sensor thickness thanks to the lack of a second layer of silicon harboring the front-end, is represented by the need of a shared design and production process. Overall monolithic pixels have typically better performances in environment requiring low material budget (i.e. low momentum tracks), with lower (with respect to hybrid pixels) level of radiations: this limitation comes from the usually lower resistivity of monolithic sensor bulks, limiting the charge collection efficiency after absorbing high doses of radiation. The other main limitation comes from the small module sizes allowed for monolithic processes, which lead to a larger number of sensors to cover a given instrumented surface.

5 Hybrid R&D

As stated in the previous section, Hybrid pixels provide the best solution for high radiation environment where large areas should be instrumented: the main lines of research go in the direction of having even radiation-harder sensors, and being able to operate properly after large accumulated doses. On the other hand there is a will to overcome the drawbacks of hybrid pixels, trying to reduce the interconnection spacing due to the bump-bonding, or thinning the sensor itself.

**HV protection at edges**

Nowadays sensors built to operate in high radiation environment, such as those in use at LHC detectors, are built with n-in-p technology. This choice is made so to have sensors and guard rings on the same side of the pixel, a cheaper solution than n-in-n, but more importantly so that the bulk does not experience a type inversion during the operations. While a p substrate does not experience type inversion, nevertheless the operating voltage will increase exponentially with the accumulated dose. During most of their operations Hybrid pixel operate at voltages up to 600V, while the front-end electronics, placed just few microns above, operates at low voltages. The difference in operating voltages is critical at the edges of the sensors, where a high voltage difference can cause discharges between the edges of the sensor and the one of the front-end, causing irreversible damage to the latter.

In recent times two paths have been explored with good results to mitigate this effect, so allowing for even higher sensor voltages to allow operations following higher accumulated doses. The first consist in the deposition of an insulator in the lithographic process of the sensor, this can be done without the need to protect the front-end structures, but increase the complexity of the sensor production process making it more expensive. An alternative consists in the deposition of parylene over the module assembly, after it has been bump-bonded, while cheaper, this procedure need the front-end to be protected during the deposition, making it riskier. Both procedures have been tested with good results, with the latter going to possibly be the standard procedure for the construction of the new ATLAS tracker.

**3D technology**

One of the main effects of radiation damage for silicon trackers is the increase of defects in the sensor bulk. These defects reduce the charge collection efficiency, and strongly depend on the drift length of the charge inside the sensor itself. In the last decade it became possible to produce 3D structures inside the silicon bulk: using these processes the sensor electrodes are deposited in columns inside the sensor, opposed to the usual planar configuration with electrodes deposited on the faces of the sensors. A graphic comparison between 3D sensors and planar sensor geometry is shown in figure 3.

![Figure 3](https://example.com/figure3.png)

Figure 3. Comparison of planar (left) and 3D (right) electrode geometry, highlighting the charge drift between electrodes.

The 3D geometry decouples the charge drift length and the sensor thickness, providing a better radiation hardiness by reducing charge drift inside the detector. While the first prototypes suffered from inefficiencies along the electrode columns, that reduced the available signal for collection, the problem has been mitigated by tilting the sensor with respect to the most probable directions of the incoming particles. The complex structure of the electrodes require an expensive and rather low yield production processes, with yields of O(80%), however more complex structures allow also for reducing the pitch between neighboring electrodes: the latest prototypes achieved a 50µm×50µm sensor cells. The first use in large scale of 3D sensors in a collider experiment was the ATLAS insertable B-Layer [11], which is operating since 2015.
**Thin sensors**

Since they rely on two silicon substrates, hybrid detectors tend to be inherently thick, their high material budget limiting their precision in tracking low momentum tracks. Moreover a thin sensor is easier to operate after irradiation: a lower voltage is needed to fully deplete the sensor after irradiation, due to the reduced volume of the bulk, the shorter collection distance reduce drift length; it allow for larger electric fields, high speed collection, and lower power dissipation. Thinner sensors on the other hand are more mechanically fragile, and the thinning process can resultin lower yields with respect to thicker sensors.

Recent results [12] have shown that thin sensors, with thickness down to 75µm can operate with high efficiencies after irradiation and at very high voltages, without incurring in breakdown or charge collection losses, as shown in figure 4. These sensors can then operate in high radiation environment for long time and good reliability, and they are also a great platform for timing measurements for 4D tracking, as will be described in section 7.

**Small pitches**

Bump-bonding process drives the interconnection pitch dimension: this has a direct effect on the minimum pitch between two neighboring pixel sensor elements. Over the years new technologies and techniques reduced the pitch dimension driven by the bonding processes, as shown in figure 5. The physical dimension of the soldering item can achieve a pitch between soldering of the order of 10µm: new prototypes are testing Hybrid modules electrically connected through the use of anisotropic conductive films (ACF)[13]. ACFs provide a cheap solution to the expensive flip-chipping processes and decouples the sensor pitch and the bump-bonding size by providing a thin film with conducting micro-particles in an insulating medium: in fact only the micro-particles in contact to both the sensor electrode and the front-end provide a viable electrical connection, while cross talk and shortings between neig-

**Figure 4.** Collected charge as a function of bias voltage for irradiated sensors with thicknesses ranging from 75 µm to 285 µm [12].

**6 Monolithic R&D**

Monolithic active pixels sensors (MAPS) are characterized by the use of a single silicon substrate hosting both the sensor and the electronics: the need for a common sensor and front-end technology can be achieved predominantly using CMOS technology. CMOS processes are cheap, widely available, and allow for a large number of detector elements to be produced, thanks to the common availability of producers. The front-end electronics are usually placed in a deep n-well while the sensor is formed by a depleted p-substrate with high resistivity.

The MAPS sensors can be very thin, resulting in detectors with 0.1%X0 thickness per layer, their thinness is paired with a thin bulk, which can be fully or partially depleted. In both cases the depletion region is rather shallow, with respect to hybrid pixels, with depleted region depths ranging between 10-50µm. The drawback of their thinness is given by the small signal produced in the shallow depletion region, a limited radiation tolerance, due to long collection times, which is particularly true if the sensor collect charge by diffusion.

In order to mitigate some of the drawbacks and exploit to the best the potential of each MAPS design, one of the following three approaches for electrode placement can be followed: large electrodes, small electrodes, or buried electrodes. A schematic view of the three designs can be found in figure 6.

**Figure 5.** Soldering dimension for different bonding techniques ordered in decreasing size, from the ordinary flip-chipping to indium bumps.

**Figure 6.** Different electrode design for MAPS, from top left, clockwise: large electrode design, small electrode design, buried electrodes.

Large electrode design uses the well holding the electronics as the collection electrode for the charge released in
the bulk by traversing particles. With this approach there are little zones of low field in the collection region, and large areas for collection allow short drift for the charges to be collected, allowing for better radiation hardness. On the other hand large electrodes result higher capacitance for the sensor, and hence higher noise. Moreover the closeness between the collection electrode and the electronics are a source of potential cross-talk between the digital and analog electronic component, interlocking even more the need for a common approach in the design of the sensor and front-end.

Some of the shortcomings of large electrodes design can be overcome through the use of smaller electrodes. In this way the electrode is decoupled from the front end, as charge is collected on an implant different from the deep well holding the analog circuitry. This design does not suffer from cross talk, and provides smaller sensor capacitances, allowing for higher signal-to-noise ratios and faster response. However the longer drift collection length makes it less radiation hard, and need adjustments to CMOS processes to maintain good performances after irradiation.

Another way of decoupling the electrodes and the circuitry is the use of buried electrodes: in this design schema, the electronics and the collection electrodes are separated by a silicon oxide layer, and the connection between the electrode and the overlaying electronics is provided through conductive vias. This approach decouples the design of sensor and electronics, and at the same time offer the possibility to use lower or higher resistivity substrates for the depletion region. The main liability of the design come from the charge build-up in the oxide layer after irradiation.

Diffusion MAPS

MAPS relying on diffusion for charge collection are coming to be used in large apparatuses in the coming years, in particular the ALICE upgrade foresee the use of monolithic sensors for its innermost detector. The high radiation doses it has to survive high doses and so a moderate reverse bias voltage will be applied to help the charge collection. The detector will be closer to the interaction point with respect to the present one, moving from 39mm to 21mm and will reduce the total budget for the first tracker layer from 1.14% $X_0$ to 0.3% $X_0$ together with a reduction of the pixel size to 30µm × 30µm pixels. ALICE MAPS will use large n-type collection electrodes placed in high resistivity epitaxial layer. And the new systems will provide standalone tracking for low momenta particle, which was hardly possible with the system used at present.

Depleted MAPS

Fully depleted MAPS are yet to be used for the first time in large apparatuses, however Mu3e experiment is going to employ fully depleted radiation hard CMOS pixels for its operations, starting in less than 5 years. Depleted MAPS have clear advantages over non-depleted ones: while sharing the same light weight characteristic, they are inherently faster with a collection time reduction proportional to the bias voltage ($V_B$), which makes them suitable for their use also in high repetition environments or where track multiplicity is very high (i.e. at LHC-like experiments). The depletion voltage scales with $\sqrt{V_B}$ and so makes it possible to fully deplete thin detectors without having to apply voltages as high as in hybrid pixels, even after irradiation.

In order to achieve the performance and operation goals for fully depleted maps, some technological challenges need to be overcome: in particular the need for radiation hard sensors require deep submicron technologies and multiple well process are needed to decouple front-end electronics from the sensitive region. This would be allowed by 130-18µm feature size. Beside very fine features radiation hard can be achieved only if they are coupled with high resistivity substrates, and processes allowing for high voltage usage in the sensors.

With these technology challenges tackled Mu3e experiment will be able to reach unprecedented precision in the measurement of lepton flavor violating process $\mu^+ \rightarrow e^+ e^- e^+$, thanks to a detector with a 80µm×80µm pixel size, a time resolution better than 20ns, and a material budget lower than 0.1 $X_0$ per layer. This configuration will allow for reliable and precise tracking of low momentum electrons, in an environment densely populated by tracks from decaying muons.

7 4D detectors

Silicon trackers are inherently fast, with collection times of the order of tenths of nanoseconds, however they are not fast enough for resolving hits from particles produced during a bunch-crossing in a highly luminous collider. In the past, the ghost-hits problems in strip detector through the pixellization of the detector. By making the tracker having a time resolution of the order of tenths of picoseconds we could solve the ambiguity in track reconstruction each hit to the track based on the time of charge release in the detector, possibly using radiation-hard detectors. This process is usually called 4D tracking, as it adds time information to the spatial one.

Timing resolution can be written as:

$$\sigma_t = \left( \frac{\text{Noise}}{dV/dt} \right)^2 + \Delta_{\text{ionization}}^2 + \Delta_{\text{shape}}^2 + \sigma_{\text{TD Gy}}^2 \quad (1)$$

if we focus on the first term, which is the leading term in our resolution, it becomes clear that a large slew rate for the signal is needed. As ratio between the peak current (S) and the time raise is constant (i.e. Thicker detectors have longer signals, not higher signals), the consequence is silicon detectors should be thinner to have better time resolution, as shown in figure 7[14].

In order to maximize the slew rate of the signal it is important to increase the peak current for it: in order to do so, and reach the goal timing resolutions a gain layer for the sensor can be beneficial. In order to do so a thin (few nm) p+-type doping is added below the n-type collected electrode: in this way a local very large field is put in place allowing for large gains (factor of 10 to 30 before
irradiation). This design is usually referred to as low gain avalanche diodes (LGAD), and are one of the most promising designs for 4D tracking. Their main drawback comes from their gain layer itself: as the gain strongly depends on the doping density in the gain layer: the doping densities are strongly affected by radiation damage, resulting in lower gains after irradiation, and hence a lower timing resolution, as the gain is one of the leading terms for the slew rate of the signal sensors.

Even with all the achieved goals and the improvements reached in the last decade, silicon pixel trackers are pushing their limits by the day: new technologies to reach higher radiation tolerance, extending the lifetime and reliability, and improving the system resolution have been developed for the next phase of experiment upgrades at LHC and for near-future experiments such as EIC or Mu3e, pushing the boundaries for both the consolidated Hybrid pixel modules, and paving the road for large-scale use of monolithic pixel detectors.

Hopefully the next generation of trackers will also be able to perform time measurement with unprecedented precision, allowing the trackers to match both the space and time coordinates for the track, to make them even more better performing in high track density conditions.

To conclude, silicon trackers are the most advanced, complex, and precise detectors for tracking charged particles, and probably still be in the coming decade, and their advancements are starting to open new paradigms in track reconstruction, and operation in harsh environments.

Figure 7. Timing resolution on LGAD detectors for different thicknesses and production processes.

8 Outlook

Silicon trackers have been a staple in high energy physics experiments since the eighties, and followed closely the development happened in the microelectronics industry and they evolved to ever bigger and more complex apparatuses. Since the nineties they have been crucial for some of the most important discoveries and measurements in our field, and they continue to do so today.

Different technical solutions have been devised for different physics cases and different operating conditions, and in particular the development of pixelated detectors improved the already high precision in position measurement achieved by silicon strips detector.

Even with all the achieved goals and the improvements reached in the last decade, silicon pixel trackers are pushing their limits by the day: new technologies to reach higher radiation tolerance, extending the lifetime and reliability, and improving the system resolution have been developed for the next phase of experiment upgrades at LHC and for near-future experiments such as EIC or Mu3e, pushing the boundaries for both the consolidated Hybrid pixel modules, and paving the road for large-scale use of monolithic pixel detectors.

Hopefully the next generation of trackers will also be able to perform time measurement with unprecedented precision, allowing the trackers to match both the space and time coordinates for the track, to make them even more better performing in high track density conditions.

To conclude, silicon trackers are the most advanced, complex, and precise detectors for tracking charged particles, and probably still be in the coming decade, and their advancements are starting to open new paradigms in track reconstruction, and operation in harsh environments.

References