PROCEEDINGS OF SCIENCE

Future of Low Mass Pixel Systems with MAPS

Jerome Baudot
IPHC/IN2P3/CNRS, Université de Strasbourg
E-mail: baudot@in2p3.fr

CMOS pixel sensors, also named MAPS for monolithic active pixel sensors, are known to exhibit a high granularity along with a low material budget. Therefore they are natural candidates for vertex detectors where these aspects govern the specifications. Projects for the coming decade or beyond have also requirements for the signal readout speed and the radiation tolerance, which are not yet reached by present MAPS.

After a presentation of the current performances reached by CMOS sensors, we discuss in this paper three technological paths to improve their speed and radiation hardness. We explore first architectures which exploit the newest planar CMOS technologies. We then consider how integration techniques may help to enhance the performances of a single layer. Finally we consider the promises of the latest 3D integration technologies.

We conclude on the timelines to reach the desired performances.

19th International Workshop on Vertex Detectors
June 6 -11 2010
Loch Lomond, Scotland, UK

PoS(VERTEX 2010)001

1. The needs of future vertex detectors, a very brief review

The physics mainly driving the need for a high accuracy track pointing resolution toward the primary vertex, comes from the electroweak decay of strange, charm and beauty quarks and tau leptons. Indeed, reconstructing the secondary vertex produced by such decays, is an efficient and clean way to identify the electroweak flavor of particles, which, in turn, is mandatory to characterize accurately the final quantum state produced in a collision.
Among the previously quoted particles, tau leptons and hadrons made of charm quarks suffer from the smallest decay length due to their respectively $c\tau(c) \sim O(100)$ μm and $c\tau(\tau) = 87$ μm; while bottom hadrons exhibit $c\tau \sim O(500)$ μm and strange hadrons $c\tau$ lie in the cm realm. Consequently, "charm and tau" drive the requirements for several future or existing vertex detectors; either to exclusively reconstruct charm hadrons like in heavy ion ultra-relativistic collision experiments (STAR [1], ALICE [2], CBM [3]), or to inclusively tag the jet flavor for high energy particle physics (ILC [4], CLIC [5]). Experiments studying CP-violation in the B meson system (LHCb [6], Belle-II [7], SuperB [8]) have a specific need for estimating the distance between two B decay vertices, which sits, again, in the 100 μm ball-park.

The granularity, or pitch of the pixels used, is obviously a main figure to reach the desired pointing accuracy. But, in fact, the pixel pitch determines only the single point resolution: $\sigma_{s.p.}$. Another parameter also impacts drastically the pointing accuracy: the material budget of the detection layers, which drives the magnitude of the multiple scatterings.
Let's consider a detector made of two layers, located respectively at radii r_1 and r_2 (with $r_1 < r_2$), both equipped with sensors having a spatial resolution $\sigma_{s.p.}$ and a material budget expressed in terms of radiation length, x/X_0. We estimate the pointing accuracy with the uncertainty $\sigma_{I.P.}$ on the impact parameter of tracks, or distance of closest approach to the primary vertex. In this simplified case, an analytical formulae can be derived for a particle of momentum p crossing the layers at an angle θ:

$$\sigma_{I.P.} = \sigma_{s.p.} \frac{\sqrt{r_1^2 + r_2^2}}{r_2 - r_1} \oplus \frac{r_1 \sqrt{x/X_0}}{p \sin^{3/2}\theta} \times c(x/X_0), \tag{1.1}$$

where $c(x/X_0) \approx 13.6$ MeV/c depends only logarithmically[1] on x/X_0 and introduces a correction ranging typically from 10 to 20%. One should note that the second term impacts the vertexing of low momentum tracks irrespectively of the single point spatial resolution. Hence, both granularity ($\sigma_{s.p.}$) and material budget have to be considered as primary figures of merit for a vertex detector.

Of course, experimental running conditions do impact the specifications for vertex detectors. In the quest for observing increasingly elusive phenomena, the collision rate and thus, the particle rate, are bound to raise. This calls for ever increasing time resolution and radiation tolerance. Time resolution can either be provided by the readout speed or by a time stamping strategy. In the former case, hits observed are allocated to the time slot where their corresponding signal was integrated which coincides approximately with the time needed to read the signals of all pixels. In the latter

[1] This is the well known $13.6(\text{MeV/c}) \times (1 + 0.038 \ln(x/X_0)$ factor appearing in the width of the angular distribution describing multiple scattering.

case, each hit is stamped with an individual time which may feature a resolution well below the time needed to read all the pixel signals along with their time-stamps.

Table 1 provides an overview of the specifications discussed above for the first detection layer of various representative projects. LHC upgrades are not considered here since the extreme fluence of 1 MeV n_{eq}/cm^2 expected, order of 10^{16}, will not be within the reach of MAPS in the immediate future.
An additional parameter was considered: the power dissipation. Indeed, while power may be evacuated by an efficient cooling mechanism, such a cooling will add material in the fiducial volume. The higher the cooling power required, the higher the material necessary. So, in the light of our discussion on material budget, a power dissipation range has its full justification in the table.

	STAR	CBM	ILC	CLIC	SuperB
spatial resolution (μm)	< 10	~ 5	< 3	$\sim$ILC	$\lesssim 15$
Material budget (% X_0)	~ 0.3	~ 0.3	< 0.3	$\sim$ ILC	< 1.0
Hit rate ($\times 10^6$/s/cm^2)	$O(0.1)$	$O(1-10)$	$O(0.2)$	$O(1)$	$O(100)$
Readout speed	200 μs	~ 10 μs	$\sim 10-100$ μs		$O(1)$ μs
or time-stamp				~ 10 ns	
Radiation / year (MRad)	$O(0.2)$	$O(30)$	$O(0.1)$	$O(1)$	$O(20)$
(n_{eq}/cm^2)	$O(10^{12})$	$< 10^{14}$	$O(10^{11})$	$O(10^{11})$	$< 10^{13}$
Power dissipation (W/cm^2)	0.1	$1-2$	0.1	0.1	$1-5$

Table 1: Specifications for the first layer of different vertex detectors. These numbers have to be understood as guide lines for the sensor design, some of them include a security factor ($\times 3$ to 5).

2. Current or standard performances of MAPS

One of the prominent features of CMOS pixel sensors [9] is the small thickness reachable for its sensitive layer. This is rooted to the fact that the CMOS technology allows to include a low noise[2] preamplifier in each pixel, making CMOS sensors effectively monolithic.
The sensitive volume is usually made of an undepleted silicon layer grown by epitaxy and buried directly below the electronics layer (transistors, diodes, capacitors...). In such a material, a minimum ionizing particle creates charges which drift thermally and eventually generates a signal onto a group of collecting diodes (usually one per pixel). For a 10 to 20 μm sensitive layer thickness, the highest pixel signal in the group is typically about a few 100 e^-. This pixel output offers a signal over noise ratio in excess of 15 for the most probable value, which is enough to provide a 100 % detection efficiency. Thinning the whole sensor consists in removing almost utterly the 500 to 700 μm of substrate, keeping only the epitaxial and electronics layers. The operation does not alter the basic principle of operation and may lead to a total sensor thickness from 20 to 30 μm depending on the CMOS process technology.

[2]The equivalent noise charge is usually below 20 e^- at room temperature.

The single point or spatial resolution benefits from the signal spread over several pixels, which results in a sub-pixel size resolution using dedicated algorithms (center of gravity or non-linear algorithm). Due to the smallness of the CMOS process feature size, a pitch well below 50 μm can be easily achieved. An example of performances obtained with MAPS (from the MIMOSA[3] family) with an analogue output is given in figure 1. Each point in the figure corresponds to a detection efficiency above 99.9 %.

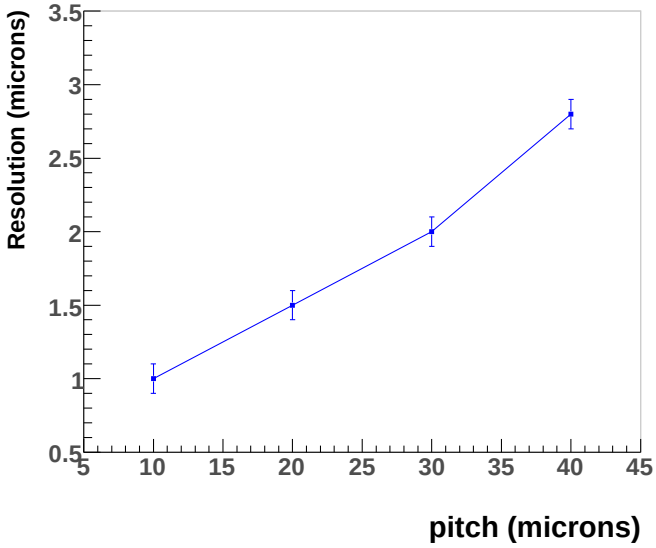

Figure 1: Spatial resolution as a function of the pixel pitch obtained with prototypes of the MIMOSA series featuring analogue readout.

However impressive these performances may be, the standard readout mode and sensitive volume of CMOS pixel sensors call for improvements to comply with the required readout speed and radiation tolerance introduced in the first section, see table 1.
First of all, comparing the pixel area to useful sensor detection surfaces, leads typically to $O(10^6)$ pixels per sensor. In the simplest readout mode, illustrated in figure 2a, where the output of all pixels is passed to the acquisition system, the pixel matrix readout time depends on the clock frequency and the output parallelization level. Experience shows that, realistically, the readout time in such conditions cannot go much below 1 ms for a megapixel sensor. This under-matches, at least by one order of magnitude, the desired specifications.
If we now consider the radiation tolerance, numerous studies have shown that it is limited by two parameters. On one hand, ionizing radiations increase the leakage current and hence the equivalent noise charge of the sensing node by accumulating charges in the transistor gate oxide [10]. The accumulation of charges increases with the oxide thickness. Consequently, the thinner this oxide, the more tolerant the sensor. On the other hand, non-ionizing radiations decrease the charge collection efficiency by inducing damages in the silicon crystal. These damages act as traps for the charge carriers to be collected and limit their lifetime. So, the collection time in a given sensitive

[3]Minimum Ionizing MOS Active pixel sensor

layer drives its tolerance to non-ionizing radiations: the shorter the time, the greater the tolerance. Undepleted volumes are particularly sensitive to this effect since charges drift thermally leading to a relatively long collection time ($O(100)$ ns).
With a standard 0.35 μm CMOS technology, sensor performances were observed [3] not to degrade up to 1 MRad and $1. \times 10^{13}$ n_{eq}/cm^2 when featuring a small pixel pitch (10 μm) and being operated at sub-zero temperatures. While such a tolerance is sufficient for a number of projects (*e.g.* STAR or ILC), though the pixel pitch is not adapted; it fails for others (*e.g.* CBM or SuperB). The sensitivity to ionizing radiation is known to improve drastically when very deep sub-micrometer CMOS processes are used because they feature very thin gate oxide. Hence we shall not consider it as a serious issue. But gaining one or two orders of magnitude in tolerance to cope with fluences of $O(10^{15})$ n_{eq}/cm^2 requires new concepts.

This rapid tour of the standard performances of MAPS clearly points to the need for improvements in two main directions:

- developing a readout mode allowing for faster frame rate,
- decreasing the collection time of charge carriers with a new sensitive volume for a higher non-ionizing radiation tolerance.

On top of that, we did not yet deal with the integration of such sensors in a real detector. MAPS have already been operated in beam telescopes used to characterize new devices, see [11] for instance. They will be also employed in FIRST, a nuclear physics experiment [12] in 2011. Both applications exploit several small area planes of 50 μm thick sensors, stacked in a fixed target fashion where integration is much more simple than in barrel-type experiments on colliders.
We shall discuss, in the next three sections, different strategies to fulfill this program and address the integration issue.

3. Improvements from new planar CMOS process

Being based on an industrial technology, the CMOS process, MAPS performances benefit from the industry progress. We present several developments taking advantage of different technologies to reach the desired improvements, starting with the readout speed and then turning to the radiation tolerance. These technologies might not be necessarily new from the point of view of the micro-electronic industry, nevertheless their usage for MAPS is. We consider in this section, only CMOS processes known as planar, or 2D, in the sense that there is only one level of micro-circuits implanted in the electronic layer.

3.1 Readout speed

Single high energy collisions between particles produce events with a low fill factor on granular detectors, much below 10 %. Under these conditions, a pixel sensor looks like a sparse hits matrix. A time efficient way to read such a matrix consists of reading out only the relevant information, that is the position, signal and potentially time stamp of the pixels which have been fired. Known as zero-suppression or sparsification, this technique has been implemented differently by several groups, among which we present the two main variants.

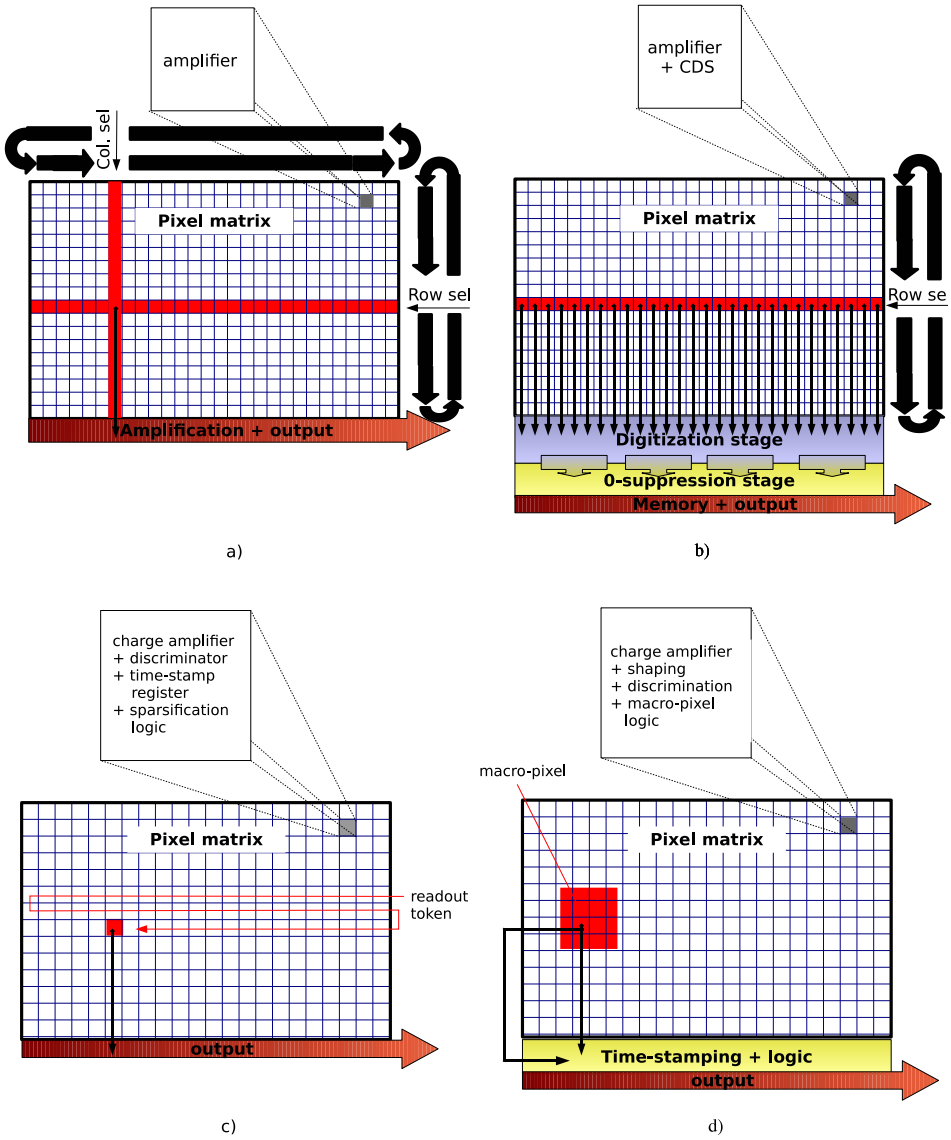

Figure 2: Crude schematic of sensors featuring different level of integrated intelligence: a) no intelligence, b) discrimination and sparcification provided outside the pixel matrix, c) in-pixel discrimination and sparsification, d) in-pixel discrimination but main logic outside the pixel matrix.

3.1.1 Out of pixel discrimination

The first strategy externalizes both the discrimination and the sparsification functions outside the pixel matrix. Each pixel incorporates a pre-amplifier and correlated double sampling (CDS) micro-circuitry. And each column of pixels is terminated by a threshold discriminator at the edge of the matrix, allowing to decide wether the pixel was fired or not. In turn, the discrimination result enters a zero suppression logic stage which stores the fired pixel address in memory for further readout, as depicted in figure 2b.

Part of the readout speed increase comes from the parallel treatment of all columns (or all pixels in a row), the rows being addressed sequentially (rolling shutter mode). Within a fixed number of clock cycles, three main operations are conducted simultaneously: CDS and discrimination of all the pixels in a given row (denoted n), zero suppression of all the discriminator outputs of the previous row ($n-1$) and storage in memory, reading out from the memory of the fired pixel addresses of the previous to previous row ($n-2$). The number of clock cycles required, depends primarily on the column length (number of rows) and the clock frequency but is independent on the row length. The matrix readout time with this architecture is given by the product of the number of rows and the time to read a row. The latter corresponds to a dozen of clock counts in a 0.35 μm CMOS process with 576 rows and a frequency around 100 MHz. Consequently for 1000 pixels long row, the readout time is improved by two orders of magnitude compared to a sequential readout using the same frequency. Up to the discrimination stage, the readout speed is not limited by the hit rate. But, of course, the occupancy level drives strongly the design and size of the zero suppression logic and of the memories. For a fixed pixel size, the higher the occupancy and the desired readout speed, the larger the logic micro-circuitry and the memories.
Additionally, one should note that the rolling shutter mode is efficient with respect to power dissipation, since only a few rows are powered on simultaneously.

The IRFU-Saclay and IPHC-Strasbourg groups have followed this model for the MIMOSA 26 [13] sensor in a standard CMOS 0.35 μm OPTO technology. This MAPS, which area occupies a full reticule, features a matrix of 576 rows by 1152 columns of pixels with a 18.4 μm pitch and was fabricated in 2008 and 2009. Operated at a clock frequency of 80 Mhz, the readout time of the full matrix is 112 μs and can deal with a particle rate as high as 10^6 hits/cm^2/s. Several test campaigns estimated the detection efficiency for minimum ionizing particles above 99.5 % for an average fake hit probability below 10^{-4}/pixel and a single point resolution about 3.5 μm. The total power dissipation was measured to amount to about 780 mW (about 1 μW per pixel) when operating continuously with the highest occupancy; the pixels and discriminators dissipation accounts for 68 % of this power. Note that the rolling shutter readout makes this dissipation largely independent of the number of rows.
The MIMOSA 26 detector was produced for the reference planes of the beam telescope included in the EUDET [11] project. The same architecture, with an extension of the pixel matrix size, will equip the next vertex detector of STAR, the Heavy Flavor Tracker (HFT) [14], to start physics data taking in 2013.

The implementation of this architecture in a very deep sub-micrometer processes ($\lesssim 0.18$ μm) will bring various improvements. Indeed, the smaller feature size and the larger number of metal layers provided, help minimizing the area occupied by a given micro-circuit. Consequently the insensitive surface represented by the zero-suppression logic, the memories or steering functionalities will decrease by 25 to 50 %. On top of that, such technologies will speed up the readout (probably at the 10 μs level) and lessen the power dissipation (around or below 100 mW/cm^2) mainly because of a lower capacitance of metal traces and, again, the integration of more micro-circuits in the same area. Finally and as we pointed out earlier, the ionizing radiation tolerance increases with a decreasing feature size.

3.1.2 In-pixel discrimination

The second strategy implements the discrimination and potentially the zero-suppression logic inside the pixel. In this case, the pixel includes at least, a classical charge amplifying stage optimized for capacitive detectors whose output is connected to a threshold discriminator. Additional digital treatments may be implemented inside the pixel or outside the pixel matrix, depending on the readout speed requirement, as it will be further detailed below.
Such an architecture involves obviously many transistors inside the pixel with two main consequences. First, to keep the pixel pitch reasonably small, $\lesssim 50$ μm, it is mandatory to use a deep submicron process. Secondly, for the digital logic, wells with the same doping as the charge collecting diode (usually N-type) cannot be avoided inside the pixel. Thus, they will compete with the main diode to collect charge. This effect may impact the geometrical efficiency of the sensor (or fill factor). It could be alleviated by exploiting a quadruple well technology where an additional very deep P-type implant can mask the necessary N-wells competing with the N-type collecting diodes.

Italian groups (see institution list in [15]) have successfully developed such pixels for the APSEL[4] sensor family [15] in a 0.13 μm triple-well CMOS process. Their two last prototypes target specific operating conditions at the ILC and the SuperB colliders.
For the ILC, the SDR[5] chip includes, in each pixel, a 5-bit time stamp register (offering a time resolution of about 30 μs) and a sparsification logic. The readout, based on a token passing scheme, is delayed with respect to the charge collection, to benefit from the ILC beam time structure (200 ms cycle with a 1 ms colliding duration). This strategy is schematized in figure 2c. The first prototype SDR0 [16] features an impressively small pixel area of 25×25 μm^2 with respect to the intelligence it contains and is design to dissipate 5 μW of power per pixel. Various small size submatrices (16×16 or 8×8 pixels) have validated the design principle up to a clock frequency of 50 MHz. However, the ILC specifications call for an even smaller pitch (about 20 μm) which, in this development, will require a new technology (see the last section and [36]).
For the SuperB collider, the SLIM 5[6] collaboration have produced the APSEL4D [17] prototype in the same 0.13 μm technology. The chip contains a matrix of 128×32 pixels arranged in groups of 4×4 pixels called macro-pixels, the pitch being 50 μm. Each in-pixel discriminator outputs are connected into a macro-pixel which signals itself when hit and freezes the 16 pixel information until they have been been swept toward the periphery of the pixel matrix (data-driven readout mode). Dedicated micro-circuits, implemented in the outside part of the sensor, provide the time-stamping of individual fired pixels and the sparsification logic, see figure 2d. This readout strategy has been designed to cope with an average hit rate of 10^9 hits/s/cm^2 when clocked at 80 MHz, to match the 200 ns time lapse between two bunch crossings. This performance is reached somewhat to the detriment of the power dissipation which amounts to about 30 μW per pixel. Beam test with 12 GeV protons and a sensor clock frequency of 20 MHz, yielded an efficiency of 92 % for a fake hit probability of 2.5×10^{-3}/pixel. The spatial resolution was estimated to a value near the digital resolution of 14.4 μm. Indeed, only a small fraction (2 to 7 %) of the hits contains

[4] Active Pixel Sensor ELectronics
[5] Sparcified Digital Readout
[6] Silicon detectors with Low Interaction with Material

more than one pixel. The pixel optimization is still ongoing to improve the efficiency limited by the presence of competing charge collecting wells. Of course, like the development for the ILC previously discussed, this strategy will greatly benefit from the new vertical integration technology we will discuss later.

3.1.3 Remarks on integration

For the sake of conciseness, we cannot present all the approaches followed by other groups developing MAPS [19, 20, 21] in various CMOS technologies. Nevertheless they are all related to the ones we presented and all aim at decreasing the readout time or, equivalently, at providing a time-stamping of the hit pixels.

The concluding remark of this sub-section should point to the strong anti-correlation between the sensor area, its power dissipation and the readout speed (or the hit rate it can cope with) which stems from the 2D limits of the planar CMOS technology. Indeed, whatever the strategy considered, the intelligence added to the simple sensing element creates an insensitive area and higher clock frequency dissipates more power. Both aspects negatively impacts the integration of MAPS in detectors. Insensitive area increases the effective material budget even if the chip thickness stays at 50 μm. And, at some level, the power dissipation requires a cooling system which contributes as well to the material budget.

However, each strategy, out-of- or in- pixel discrimination, presents a different optimization. The out-of-pixel discrimination combined with the rolling shutter readout mode, offers small pixel size and a low power dissipation. It suits applications where granularity and/or material budget are to be favoured. The in-pixel discrimination allows for the time-stamping of hits and has to be privileged in case of a high hit rate environment. It is yet difficult to quantitatively draw limits since both approaches are still evolving.

3.2 Non-ionizing radiation tolerance

As in many detection technologies, the increase of the radiation tolerance of MAPS is a matter of improvement of the sensitive volume and not of micro-electronics design. In recent years, the CMOS industry, in its quest to increase the low wavelength light detection efficiency, has started to provide a highly resistive epitaxial layer, above 100 Ω.cm, in contrast to the standard layer which offers a resistivity of about 10 Ω.cm. This increase in resistivity results in a deeper depletion area in the sensitive volume, even though the depletion is still incomplete. At the modest biasing voltages for the collecting diode, a few Volts, used in the usual CMOS technology, a sizable proportion of the 10 to 20 μm thick epitaxial layer can be depleted [22].

A partially depleted sensitive volume is expected to shorten the charge collection time and focus the signal on a fewer number of pixels; in short, it increases the signal over noise ratio. This trend has been indeed observed in the MIMOSA 25 prototype [22] fabricated in a 0.6 μm technology with a 1000 Ω.cm epitaxial layer. The signal-to-noise ratio for electrons from a ^{106}Ru source was measured to stay above 30 (most probable value) after an irradiation to a fluence of $3. \times 10^{13}$ 1 MeV n_{eq}/cm^2, for a 20 μm pixel pitch and at a temperature controlled around 20 °C. This insures the detection efficiency to stay very close to 100 % for minimum ionizing particles.

Several groups are developing sensors with high resistivity layers either with a standard voltage level of a few Volts [24], or with a high voltage technology of several tens of Volts [25, 26]. These developments currently bring MAPS tolerance to non-ionizing radiation at a level about 2 orders of magnitude below the fluence of 10^{16} 1 MeV n_{eq}/cm^2 required for the the LHC upgrade (ATLAS, CMS, LHCb inner layers).
However, studies [27] conducted for LHC experiments on the high purity silicon 3D or planar sensors operated at 100s to 1000s of volts bias, indicate that this high non-ionizing fluence decreases the depleted area useful for detection, from a few hundreds to about 20 μm. This is typically the collection thickness for which the in-pixel amplification featured by CMOS sensors is optimized. Consequently we may infer that MAPS will ultimately equal or even surpass hybrid pixel non-ionizing radiation tolerance, provided that the sensitive volume employed in the CMOS technology could approach the hybrid-type sensor characteristics in the future. Reaching such a performance will probably require the vertical integration techniques discussed in the last section.

It is worth to mention here, that the non-ionizing radiation hardness being a matter of charge collection efficiency, the initial performance of the sensor in this matter plays an important role. Since the pixel pitch influences the charge spread over several pixels, it impacts this radiation tolerance [23]. This impact depends certainly on the nature and thickness of the sensitive volume but the quantitative proportion of this dependence is largely unknown today.
Again, we face the correlation between the various performances of a given MAPS, a small pitch sensor will exhibit, with respect to a larger pitch sensor: a better non-ionizing radiation tolerance and a better single point resolution but a smaller readout speed and potentially a larger power dissipation. The monolithic nature of CMOS sensors induces that the part of the system integration usually devoted to the connection of the sensor to the readout electronics, happens at the sensor design phase. As we have seen throughout this section, the design in terms of pixel pitch, material budget, readout strategy and radiation tolerance is constrained by the technology choice which is then of primal importance. This fact explains why each development route associated to a given 2D CMOS technology will lead to different applications.

4. Improvements from integration technologies

Though we already explored some integration aspects through the development of CMOS sensors themselves, we have not yet discussed the following points: mechanical support, cooling and electrical services. Before describing several ongoing integration projects addressing them, we present an idea to improve the overall vertex detector performances. This example intends to demonstrate the integration possibilities opened by very thin sensors.

4.1 Detection performance improvements

The previous section taught us that the design of MAPS in a single 2D CMOS technology is a trade-off between several fundamental characteristics, like granularity and readout speed or even radiation tolerance. However, almost independent of the compromise, the thickness of the sensor stays very low around 50 μm which amounts only to 0.05% of the radiation length (X_0).

This small single sensor material budget opens the possibility to stack two sensors in the same detection layer, with a short spacing at the mm level, without compromising the overall budget. Their proximity insures the possibility to match efficiently their corresponding hits. Each sensor can then be optimized differently. Figure 3 illustrates such a double-layer principle; one sensor being devoted to the spatial resolution with small pixels (and being relatively slow), the other providing a short readout time thanks to larger pixels (and having a somewhat degraded spatial resolution).

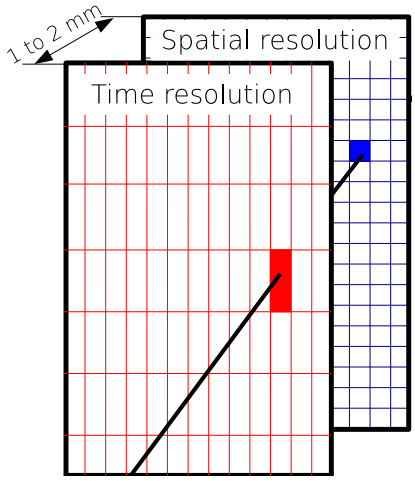

Figure 3: Schematic illustration of the combination in a two-sided layer, of two sensors, with different pixel size and readout speed.

Of course, the choice of the distance separating the two layers is driven by the hit rate which, if high, hinders the hit matching between the layers, especially at large incidence angles. Also, having a double-layer pushes the power dissipation up. Nevertheless, this strategy illustrate how to alleviates the limits encountered with the planar CMOS technology.

4.2 Ongoing integration projects

We already stressed several times that a strong point of monolithic CMOS pixel sensors lies in their small thickness. This stems both from the thin sensitive layer (below 20 μm) and the sensor embedded intelligence which eliminates the need for additional readout electronics. Also, MAPS usually operate at room or modestly low (around 0 °C if needed for radiation tolerance) temperatures. So, they do not require excessively powerful cooling systems which, again, is favorable for the material budget.
However, it has to be demonstrated that structures can actually be built to stably support and bring all services to the ultra-thin sensors while keeping the material budget below the few per mil of X_0 demanded by many projects, see table 1. The mechanical stability along with the ability to align several sensor layers together at all times are all the more important regarding the spatial resolution which can be required at the few μm level.

Figure 4 provides a schematic of various integration projects which are briefly described below with an emphasis on their specificities.

4.2.1 STAR

The STAR experiment will be the first (in 2013), in a collider setup, to equip its vertex detector with CMOS sensors. We already mentioned several times this project, lead by the Relativistic Nuclear Collisions group at LBNL; details can be found in [14]. The detector concept emphasizes both the material budget and the possibility to swap detector copies inside the experiment within several hours. The two 20 cm long single-layers are supported together by a unique carbon composite part for one angular sector. The sensing layer itself is arranged in a stave of 10 sensors, each 2 cm long. The electrical services are brought to the 10 chips by a single micro-cable which is a multi-layer sandwich with interleaved metal and kapton layers. An air flow circulating on top of each ladder evacuates the heat generated by the power dissipation which stays below 200 mW/cm^2. In the current development phase, the global material budget for one layer amounts to 0.37 % X_0 of which about 60 % are accounted by the mechanical support. Indeed, the need for swapping easily detector copies allows the support to be fixed on the detector frame only from one side and hence involve a stiff structure.

The STAR design, depicted in figure 4a, presents the simplest integration variant with three basic elements: a support, an electrical cable and the sensor. We shall note that all design variants[7] will have a similar part: the micro-cable; whereas the other parts will change significantly depending on the priority driving the design.

4.2.2 SuperB

For the SuperB vertex detector, italian groups, within the SLIM5 collaboration [17], face the challenge of a very high hit rate of several 10 MHz/cm^2. The sensor intelligence to cope with this rate induces a power dissipation foreseen in the few W/cm^2 range which, in turn, requires an active cooling. Consequently the SuperB concept departs from the STAR one regarding the supporting structure, though it is also a single-layer design.
A specific development was conducted in the recent years to produce a light supporting structure made of carbon fiber with incorporated microtubes which serve to circulate the coolant, see figure 4b. MAPS will be directly in contact with the tubes and connected to the outside world with a micro-cable on top. Recent prototypes of micro-tubes structure with 0.15 % X_0 have been produced [18]. The support is 700 μm thick and can extent to 30 cm (SuperB layer0 has a 10 cm sensitive length).

4.2.3 PLUME

The approach pursued by the PLUME collaboration [28] (Bristol University, DESY-Hamburg, IPHC-Strasbourg, Oxford University) focuses on double-sided layers, as the one introduced in the previous subsection. The goal is to fabricate by 2012 a double-sided ladder with a total material budget around 0.3 % X_0 with the ILD dimensions (roughly 12×2 cm^2 active area).

[7] Except, perhaps, with the stitching variants of figure 4e.

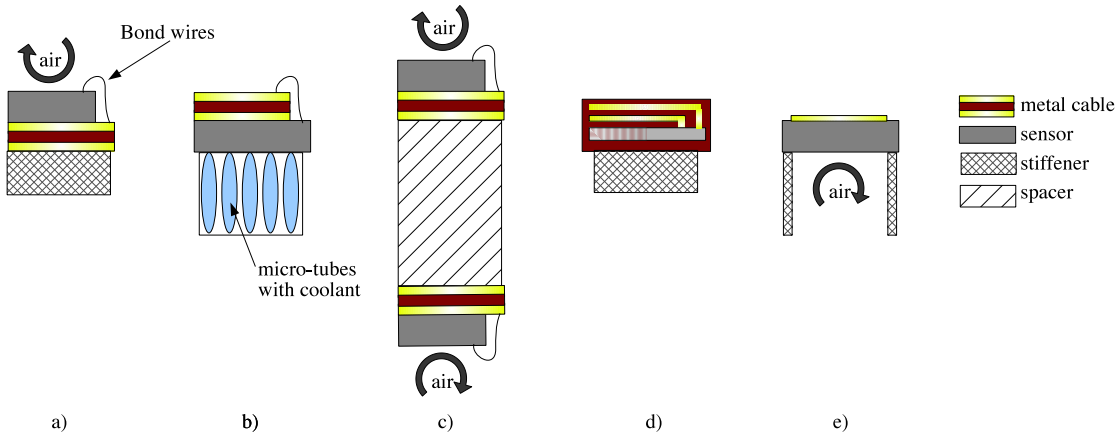

Figure 4: Several MAPS integration variants, a) STAR, b) SuperB, c) PLUME, d) SERWIETE, e) stitching. Note the vertical scale is increased by two orders of magnitude with respect to the horizontal scale.

This ladder is expected to operate at room temperature with air cooling and power pulsing, at the 5 Hz frequency matching the ILC beam time structure. Here, a multi-layer kapton-metal micro-cable is equipped with six 50 μm thick MAPS (currently MIMOSA 26 sensors) to make a module. Two modules are then glued on a 2 mm thick silicon carbide foam, one on each side, see figure 4c. This sandwich structure benefits to the stiffness of the ensemble. So it allows to use a light foam, with a relative density much below 1 %, which serves essentially as a spacer.
The prototype in fabrication for 2010 reaches a material budget of 0.6 % X_0. The excess of material with respect to the goal is mainly due to the micro-cable characteristics which have been optimized for safe electrical functionalities and not yet for material budget. The final prototype with the expected thickness is in preparation for 2011.

The double-sided layer approach is particularly sensitive to the micro-cable material budget since there are two of them. Many geometrical and electrical constraints apply to these cables. Experience from many projects taught us that a few years of development may be required to reach the proper specifications and production quality for such parts. Currently the technology supplies cables made of a stack of metal layers (copper or aluminium with a range of thickness from 10 to 20 μm) spaced by polyimide layers (thickness from 20 to 50 μm). The cable typically features a 150 to 200 μm total thickness and a material budget of 0.05 % (agressive design) to 0.2 % (when many traces are needed) of X_0.
The two last projects we present, offer alternatives to the use of such standard micro-cables.

4.2.4 SERWIETE

The technology provided by the IMEC company and the CMST laboratory in Belgium [29, 30] allows to view the micro-cable as the sensor support itself. Here the micro-cable fabrication steps integrate the embedding and the thinning of the sensor inside a very thin polyimide film (< 25 μm), see figure 4d. Since sensors are thinned after they have been embedded, a minimal thickness for

them is reachable, around 20 μm. Indeed, if the chips were thinned individually to 20 μm before their integration, their handling would be extremely risky. That is why, in practice, thickness above 40 *mu*m are preferred in usual approaches like the ones described before. The electrical connection of traces to the sensor pads happens directly when the metal is laid over the insulating polyimide layer.
A collaboration between IK-Frankfurt and IPHC-Strasbourg investigates the fabrication of a multi-sensors embedded cable for an overall material budget below 0.1 % X_0 by 2011. An original and potentially very interesting point is that this ensemble may be bent to take a curved shape. New geometries of vertex detector, to cover beam pipes for instance, can thus be explored.
Also this technology may open the road to integrate some services directly into the micro-cable, like optocouplers. If successful, this approach would allow to minimize the material requested to connect the ladder of sensors at its end and will thus benefit to forward tracking devices for instance.

4.2.5 Stitching

To complete our tour of integration techniques, we mention the possibility to stitch MAPS at the fabrication level. This technology has already been demonstrated for other sensor types, like CCD [31] and DEPFET [32]. The former was envisioned for the ILC perspective while the latter is currently studied for the BELLE-II detector. Stitching consists in butting several times the same sensor or pixel matrix on a single silicon wafer which is then diced around the group of sensors or matrices. When sensors are replicated at the foundry level, the dead area between them are minimized much better than when they are placed individually. When the pixel matrix is replicated and surrounded by the steering and treatment micro-circuits, the dead area simply reaches zero. In both cases, a single large object is produced with the silicon mechanical stiffness. Hence, depending on the stave length, less additional stiffening material is required for the mechanical stability than in the case of a multi-chips stave. Figure 4e depicts such a structure with just two thin supporting walls.
Stitching being provided by the silicon industry, there is no reason why it could not be of benefit to CMOS pixel sensors. Indeed, the Rutherford Appleton Laboratory CMOS sensor group has already produced a 5.4×5.4 cm^2 prototype [33]. If the stitching is used to form a ladder of continuous pixel sensors (typically 10×2 cm^2), which can be thinned down to a few tens of micrometer, then one can get partially or completely rid of the micro-cable, see figure 4e.
The overall expected material budget with a ladder featuring stitched sensors will reach the level of 0.1 % of X_0. Of course, one of the difficulty comes from the fabrication yield of such a sensitive staves made of 5 to 10 stitched chips. Even if, the yield of individual sensor with a few cm^2, ranges usually from 60 to 90 %.

We are not aware of projects presently focussed on producing sensors stitched in a stave geometry over a length of 10 cm. However, even in more classical approaches, stitching two or three sensors would improve both the heat dissipation and the stiffness. The PLUME collaboration may consider this possibility beyond 2012.
The stitching of four sensors together in a square geometry, like realized in [33], is more interesting for forward tracking in collider or fixed target experiments. In the framework of the Electron Ion

Collider project [34], a collaboration between BNL, the Columbia University and IPHC intends to exploit stitching for producing a sensor featuring a 5×5 cm^2 sensitive area, a pixel pitch about 20 μm and a readout time of 300 μs by 2013. The chip will serve for a proof-of-principle study of forward disks with minimal material budget, especially important for the low momentum particles emitted in the electron-ion collisions.

4.2.6 Concluding remark on integration techniques

We have briefly overviewed five integration projects with different goals and exploiting various techniques. Though their current level of achievement are different, all of them will produce ladder prototypes in the three coming years. These will have to demonstrate they actually offer the support and provide the services required to benefit from the MAPS high spatial resolution.
Starting from February 2011 and for 3 years, the European Union is financing within its Frame Program 7, an international project named AIDA [35] where a work-package is devoted to the complete evaluation of such ladders. Especially a test beam infrastructure [28] will allow to test several ladders arranged geometrically like a vertex detector sector and operated in real conditions: (air-)cooling, power pulsing, high magnetic field. Demonstration of the alignment of such a device is one of the important expectation of the project.

5. Improvements from 3D CMOS process

Up to now, we discussed essentially relatively well established technologies, even if their application in High Energy Physics is under development. We also pointed out that a strong point of MAPS is that they rely on the CMOS industry, which technological progresses continuously move at a good pace. One of the ongoing revolution in this industry is the 3D integrated circuit (3DIC) technology. It consists in stacking several very thin ($O(20)$ μm) circuits and inter-connecting them all over their surface, with a connection pitch as low as a few micrometers. Of course this interconnection pitch is fully relevant for pixel sensors, since it permits to connect a sensitive pixel on one tier (one circuit of the stack) to a signal treatment pixel on another tier. Also, the final thickness of a 3-tiers sensor would reach only about 60 μm compared to the current 50 μm in the 2D technology.

As depicted in figure 5, this strategy clearly enhances the possibilities for intelligence embedded in a sensor, because more area can be devoted to signal treatment. On top of that, each tier may be designed in the most suited CMOS process with respect to its functionality. For instance, the sensing tier will use a high resistive epitaxial layer for non-ionizing radiation tolerance, while the digital treatment tier will choose a very deep submicrometer process allowing for complex algorithm implementation and low power dissipation.
This approach is a major step forward because it almost suppresses the limitations we have encountered with 2D processes, where performances (granularity, radiation tolerance, material budget and readout speed) were anti-correlated. It will bring mega-pixel sensors with high granularity (< 20 μm pitch) and readout speed at the micro-second level or time-stamping with a 10 ns precision to reality.
Integration perspectives are also valuable with 3DIC. Indeed, the last tier can be optimized for ex-

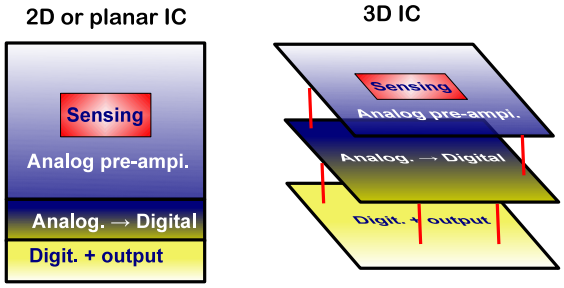

Figure 5: Moving from the 2D or planar integrated circuit technology (on the left) to the 3D technology (on the right) allows to benefit from more area to integrate more functionalities and to potentially use the best process for each of this functionalities.

ternal connections with, for instance, direct optic fiber connection or power supply filtering.

A consortium of several HEP institutes led by the Fermi National Laboratory has submitted many different prototypes in 2009 for an exploration of this technology. Answers are expected by 2011 and more details on this project and the technology can be found elsewhere [36, 37] in these proceedings.

6. Conclusion

The CMOS pixel sensor technology for charged particles was pioneered in the late 90's. It is providing now its first applications in high energy physics: the EUDET beam telescope since 2008 or the STAR-HFT in 2013; and in nuclear physics: the FIRST vertex detector in 2011. These first applications mainly rely on the granularity ($\ll 50$ μm pixel pitch) and small material budget ($0.3-0.4$ % of X_0) almost natively offered by the technology. They will demonstrate the maturity and suitability of MAPS for vertex detectors.
We have shown that the numerous 2D CMOS process variants and the intelligence embedded in these monolithic sensors can be further exploited to enter the realm of the $O(10)\mu$s readout time and the tolerance to fluences of $O(10^{14})$ n_{eq}/cm^2. These performances will match specifications for vertex detectors at the ILC or CBM around the middle of this decade. However, 2D processes limit further improvements into one direction at a time, either readout speed or radiation hardness but not altogether.
The integration projects reviewed, have demonstrated that MAPS offer a wide palette of system optimizations. By the second half of the 2010 decade, long staves or disks equipped with CMOS sensors and featuring a material budget below 0.3% X_0 will have been prototyped. Potentially, stacking in the same layer several thin sensors, with different optimization, will combine their performances to cope with the very high hit rate and still feature a high granularity.
Finally, the awaited 3D integration technology will potentially bring MAPS at the forefront performances on almost all characteristics: granularity, material budget, speed, radiation tolerance and power dissipation. But these are promises for the next decade.

Acknowledgments

The author thanks Marc Winter for inspring advice in the preparation of this talk and help in writing this article. Many thanks go also to all colleagues in the CMOS pixel development community for the discussions and the materials provided.

References

[1] J. Bouchet for the STAR collaboration, *Heavy Flavor Tracker (HFT): A New inner tracking device at STAR*, Nucl. hys. **A 830** (2009) 636C-637C.

[2] A.Dainese, *ALICE perspectives for the study of charm and beauty energy loss at the LHC*, Eur. Phys. J. **C 49** (2007) 135-141.

[3] M. Deveaux *et al.*, *Challenges with decay vertex detection in CBM using an ultra-thin pixel detector system linked with the silicon tracker*, in proceedings of *VERTEX 2008*, PoS(VERTEX 2008)028.

[4] *International Linear Collider Reference Design Report: Detectors*, ILC-Report-2007-001, editors T. Behnke, C. Damerell, J. Jaros, A.Miyamoto, September 2007.

[5] E. Van der Kraaij, *R & D Challenges of a CLIC Vertex Detector*, in proceedings of *VERTEX 2010*, PoS(VERTEX 2010)041.

[6] M. Gersabeck, *LHCb Alignment Tracking and Physics Performance*, in proceedings of *VERTEX 2010*, PoS(VERTEX 2010)014.

[7] *Belle II Technical Design Report*, KEK Report 2010-1 and arXiv:1011.0352v1, edited by Z. Doležal and S. Uno, Tsukuba, November 2010.

[8] E. Grauges *et al.*, *SuperB Progress Reports – Detector*, [arxiv:1007.4241], July 2010.

[9] R. Turchetta *et al.*, *A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology*, Nucl. Instr. and Meth. **A 458** (2001) 677-689.

[10] P. Jarron *et al.*, *Deep submicron CMOS technologies for the LHC experiments*, Nucl. Phys. **B 78** Proc. Suppl. (1999) 625-634.

[11] I. Gregor on behalf of the EUDET Consortium, *The EUDET Infrastructures for Detector R&D*, in proceedings of the IEEE Nuclear Science Symposium, 2009, October 25-31, Orlando, N45-2.

[12] E. Spiriti, *The FIRST (Fragmentation of Ions Relevant for Space and Therapy) experiment: interaction region and MAPS vertex detector*, in proceedings of the 12th Topical Seminar on Innovative Particle and Radiation Detectors, IPRD10, 2010, June 7 - 10, Siena, to appear in Nucl.Phys. B.

[13] C. Hu *et al.*, *First reticule size MAPS with digital output and integrated zero suppression for the EUDET-JRA1 beam telescope*, Nucl. Instr. and Meth. **A 623** (2010) 480-482.

[14] M. Szelezniak, *MAPS based vertex detector at STAR*, in proceedings of *VERTEX 2010*, PoS(VERTEX 2010)043.

[15] S. Bettarini *et al.*, *Development of deep N-well monolithic active pixel sensors in a 0.13 μm CMOS technology*, Nucl. Instr. and Meth. **A 572** (2007) 277-280.

[16] G. Traversi *et al.*, *First generation of deep n-well CMOS MAPS within-pixel sparsification for the ILC vertex detector*, Nucl. Instr. and Meth. **A 604** (2009) 390-392.

[17] S. Bettarini *et al.*, *The SLIM5 low mass silicon tracker demonstrator*, Nucl. Instr. and Meth **A 623** (2010) 942-953.

[18] F. Bosi *et al, Development of light prototypes support for silicon pixel detectors cooling based on micro-channel technology*, in proceedings of the 2010 International Workshop on Semiconductor Pixel Detectors for Particles and Imaging, PIXEL 2010, September 6 - 10, Grindelwald, to appear in Nucl.Instr.Meth. A.

[19] A. Bevan *et al.*, *Design Study for a SuperB Pixel Vertex Detector*, arXiv:1008.1938, August 2010, 8pp.

[20] G. Varner *et al.*, *Development of the Continuous Acquisition Pixel (CAP) sensor for high luminosity lepton colliders*, Nucl. Instr. and Meth. A 565 (2006)126.

[21] J.Brau *et al.*, *Development of an ILC Vertex Detector Sensor with Single Bunch Crossing Tagging*, in proceedings of the 2007 International Linear Collider Workshop (LCWS07 and ILC07), Hamburg, 30 May - 3 Jun 2007, pp TRK15.

[22] A. Dorokhov *et al.*, *Improved radiation tolerance of MAPS using a depleted epitaxial layer*, Nucl. Instr. and Meth. **A** in press, [doi:10.1016/j.nima.2010.03.124].

[23] M. Deveaux. *et al.*, *Charge collection properties of Monolithic Active Pixel Sensors (MAPS) irradiated with non-ionising radiation*, Nucl. Instr. and Meth. A 583 (2007) 134-138.

[24] J. Ballin *et al.*, *Monolithic Active Pixel Sensors (MAPS) in a Quadruple Well Technology for Nearly 100% Fill Factor and Full CMOS Pixels*, Sensors **8** (2008) 5336-5351.

[25] A. Rivetti *et al.*, *LEPIX: monolithic detectors for particle tracking in standard very deep submicron CMOS tehcnologies* in proceedings of 11th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications, Como, 5-9 October 2009.

[26] I. Peric *et al.*, *The first beam test of a monolithic particle pixel detector in high-voltage CMOS technology*, Nucl. Instr. and Meth. **A** in press, [doi:10.1016/j.nima.2010.06.337].

[27] G. Kramberger, *Edge TCT Measurements*, in proceedings of *VERTEX 2010*, PoS(VERTEX 2010)021.

[28] N. Chon-Sen *et al.*, *Development of ultra-light pixelated ladders for an ILC vertex detector*, in proceedings of the 2010 International Linear Collider Workshop (LCWS10 and LC10), Beijing, 26-30 March 2010.

[29] W. Christiaens *et al.*, *Embedding and assembly of ultra-thin chips in multilayer flex boards*, Circuit World **34** (2008) 3 - 8.

[30] J. Govaerts *et al.*, *Fabrication processes for embedding thin chips in flat flexible substrates*, IEEE Trans. Adv. Pack. **32** (2009) 77-83.

[31] K. Stefanov, *Progress with the CPCCD and the ISIS*, in proceedings of the 2007 International Linear Collider Workshop (LCWS07 and ILC07), Hamburg, 30 May - 3 Jun 2007, pp TRK07.

[32] H.G. Moser *et al., R & D on Interconnection Technology for Belle II and SLHC,* in proceedings of *VERTEX 2010,* PoS(VERTEX 2010)036.

[33] H.M. Zin *et al.*, *Characterisation of regional variations in a stitched CMOS active pixel sensor*, Nucl. Instr. and Meth. A 620 (2010) 540-548.

[34] C. Aidala *et al.*, *A High Luminosity, High Energy Electron-Ion-Collider*, White Paper Prepared for the NSAC LRP 2007 April 24, 2007; available at [http://www.eic.bnl.gov/docs.html].

[35] Consult the AIDA web page [https://espace.cern.ch/aida].

[36] G. Traversi *et al.*, *3D consortium*, in proceedings of *VERTEX 2010*, PoS(VERTEX 2010)032.

[37] R. Yarema, *The Via Revolution*, in proceedings of *VERTEX 2010*, PoS(VERTEX 2010)035.