IC/HEP/04-7
August 2004

Recent progress in front end ASICs for High Energy Physics

G. Hall
Blackett Laboratory, Imperial College, London SW7 2AZ, UK

email: g.hall@imperial.ac.uk

Abstract

Developments of Application Specific Integrated Circuits (ASICs) for applications in the CMS experiment are briefly described, along with the motivations for the choice of technology, focussing especially on silicon strip readout of the CMS tracker. The major change in the last few years has been the widespread adoption in CMS of a commercial deep sub-micron CMOS technology in preference to specific radiation hardened processes which seemed to be the only solution meeting the LHC requirements only a few years ago. The reasons for this are described and the performance of representative chips and the technology presented. The implications for future developments are outlined.

Invited talk presented at Fifth International Symposium on the Development and Applications of Semiconductor Tracking Detectors, Hiroshima Japan, June 2004

Introduction

Over the last almost fifteen years, a considerable effort has been invested in developing Application Specific Integrated Circuits (ASICs) for the Large Hadron Collider experiments at CERN, which will start operating in 2007. The last five years saw a drastic change of direction by the Compact Muon Solenoid (CMS) experiment in adopting a standard commercial process for its silicon microstrip tracker readout circuit using a deep sub-micron CMOS technology in preference to the hardened technologies which had been used prior to 1999. This was originally motivated by difficulties in achieving the required level of guaranteed radiation hardness for the circuits required for the tracking detectors, which will be exposed to radiation levels of ~10Mrad (100kGy) over the experiment lifetime. The consequences of this change have been quite striking and have significant implications for future electronic developments for particle physics, and other scientific, experiments.

Early history of HEP ASICs

The major early motivation for silicon detector development was for direct searches for decays of charmed particles [1]. Once silicon sensors had been successfully demonstrated and the power of high resolution tracking had become evident, it was natural to attempt to exploit them in the next generation of experiments, which were intended for colliding beam experiments. It quickly became clear that much higher density electronics was required than had been necessary in fixed target experiments (see, e.g. [2-4]) and a few daring souls began the construction of integrated circuit amplifiers, learning from first principles how to build transistors and circuit elements in MOS technology.

The first example [5] was the Microplex circuit used in the Mark-II silicon vertex detector at SLAC in the early 1980s. It contained 128 channels of amplification, sample-and-hold, double correlated signal processing and multiplexing in an area of 34mm^2 using only NMOS transistors in a 5µm feature size laboratory integrated circuit process. The circuit worked well but was rather power hungry, consuming 14mW/channel, and it was followed in the subsequent few years [6, 7] by a series of chips, based on the same concept, which were exploited in the silicon tracking systems of the LEP experiments in CERN. These all used relatively up-to-date commercial CMOS processes with somewhat smaller geometries, initially ~3µm and later 1.5µm, where the use of both NMOS and PMOS transistors permitted power reductions to 1-2mW/channel depending on the noise target and load capacitance in each case.

In each of these early circuits, the amplifiers were constructed as integrators with switched capacitor filters which, combined with the natural bandwidth limits imposed by the technology, optimised the signal to noise ratio for time constants in the µsec range, perfectly matched to LEP interaction and readout rates. Switching noise injected during the amplifier reset could be subtracted due to its very reproducible behaviour. Later circuits, such as the Amplex [8, 9], began to avoid this source of undesirable noise by incorporating accurate resistors in the form of FETs into the circuits so that more conventional passive-type RC filters could be included in the chips, which otherwise would not have been possible using polysilicon resistors only. This was where many of the developments for the LHC experiments started in the late 1980s, at which time several groups also began to examine the possibilities provided by bipolar technologies which appeared to be well adapted to the fast shaping times which were essential for readout of LHC (and SSC) experiments with the originally proposed clock speeds of ~66MHz.

At the time of writing, in mid-2004, the CMS experiment [10] has constructed all ASICs used in the innermost tracking and electromagnetic calorimetry of the detector, where the radiation levels are most intense, in a modern but standard commercial 0.25µm CMOS process. Many ASICs used in the outer regions of the detector, which had not completed their development before the 0.25µm advantages were recognised, are also implemented in it, as are many chips which have been developed for ALICE, ATLAS and LHCb. At first sight, without knowledge of the history, this might surprise those who had followed the first decade of developments of LHC electronics, where much attention was focussed on the importance of obtaining access to, and understanding the performance of, formally radiation-hardened technologies which were believed to be essential to ensure long term, reliable operation of LHC experiments. The CMS experience is not unique but, from implementing several large scale applications, sheds much light on why this came about.

LHC electronic developments

In the late 1980s and early 1990s, when a significant R&D effort began under the auspices of the US Superconducting SuperCollider (SSC) R&D programme and the CERN LHC-driven Detector Research and Development Committee (DRDC), major obstacles identified to successful operation of LHC experiments were the integrated circuits, which by then were recognised as essential for readout of silicon trackers and inner calorimeters. The SSC and LHC radiation levels were unprecedented, and experimenters had already observed examples of radiation damage in many types of detector, including silicon, as well as become aware of electronic failure modes, such as latch-up, which could be prompted by radiation. It was clear that the methodology for implementing space-borne experiments was unfeasible for particle physics, since it was based on shielding components from the radiation source in conjunction with careful selection and validation of individual components in lengthy qualification programmes. Since LHC experiments were aiming to instrument millions of channels with hundreds of thousands of chips, this was an unthinkable proposition.

Several hardened processes were known to exist, and others came to light, as investigations began [11, 12]. They were mainly of military origin and included less known technologies such as gallium arsenide or silicon-on-sapphire, some of which were still being developed for potential future applications. They were provided by a range of companies mainly, but not only, based in the USA and were subject to significant constraints by governments concerned about proliferation of strategic technologies. Apart from these issues, the costs of utilising the technologies appeared to be daunting and, from experience gained over the subsequent few years, were found to be relatively hard to design in, with limited support for common design tools, and with lengthy manufacture cycles, at least for new customers of limited commercial importance. The processes also, with the possible exception of bipolar technologies, were relatively antiquated compared to those used for consumer electronics, with CMOS feature sizes in the 1-2μm range compared to 0.35-0.5μm for commercial parts.

One significant problem which emerged was availability. Between 1992 and 1997 the number of accessible manufacturers had diminished greatly, for several reasons. Among them were straightforward commercial considerations; the evolution of the world political situation meant that companies were far less interested in military electronics than rapidly growing consumer product markets [13] and had probably realised that, in comparison, the economic value of LHC business was also modest. In addition serious technical challenges had emerged since few processes tested showed noise performance adequate to meet signal processing requirements for sensitive detectors, especially within the demanding power constraints. Finally, sufficient radiation hardness was not certain since most processes required for space applications could formally guarantee only 300krad-1Mrad even though some seemed substantially more robust. However, variations in process parameters meant this would remain questionable without careful monitoring of delivered chips.

It this appeared at this point that possibly only a single technology might be available for LHC detector readout, and this (DMILL [14]) was a process which had been specially developed in France with LHC as one potential market and then had been transferred to a commercial vendor. There remained doubts about the commercial viability of this approach, and the process was not fully optimised for large scale production or had completely satisfied all users that its radiation tolerance would be adequate [15, 16].

Deep sub-micron CMOS technology

The breakthrough which changed the picture came in mid-1997 when it was demonstrated that transistors constructed in several CMOS processes with feature sizes from 0.25-0.5μm showed much more promising results after irradiation than had been expected. Previously, tests on commercial parts from different manufacturers had shown large variations in hardness, even across wafers [13, 17, 18]. What was pointed out in a study led by the CERN Microelectronics Group [19] was that 0.25μm CMOS transistors showed especially good behaviour compared to coarser technologies in leakage current, threshold voltage shifts and transconductance behaviour at Mrad levels and the results suggested that higher doses could be tolerated. It was also observed that construction of “edgeless” NMOS transistors, where the gate was laid out to enclose the transistor drain, combined with guard rings, could prevent problems of radiation-induced leakage currents between devices due to changes in field oxides.

By the following year, further investigations [20] had demonstrated that radiation tolerant layout practices allowed 0.25μm circuits to withstand total doses to 30Mrad and that low noise analogue design looked feasible, based on transistor noise spectra. Tolerance to radiation-induced latch-up was shown to be

good and Single Event Upset (SEU) immunity could be enhanced by circuit architecture. A pixel detector readout prototype developed for ALICE and LHCb using the same technology [21] also showed excellent post-irradiation performance, confirming the transistor studies.

The detailed explanations for this can be found elsewhere [22-24] but, briefly, electron tunnelling into the gate oxide is believed to neutralise trapped positive charges which accumulate under irradiation. The thickness of CMOS transistor gate oxides is proportional to minimum lateral feature size, and is typically ~5nm in 0.25μm CMOS. Tunnelling is considerably more effective in the thin oxides and consequently one of the main total ionising dose effects, namely threshold voltage shifts, diminishes much more rapidly with feature size than the quadratic scaling expected from simple arguments.

These results had already encouraged the CMS Tracker to investigate a 0.25μm version of the APV circuit to be used for silicon strip readout, even though the time remaining for module construction was fast decreasing. From early 1998 to mid-1999, libraries and components were developed and the existing APV6 [25] was translated into the new process. Only minor improvements and changes, such as extended pipeline memory length, which was possible without increasing the overall chip dimensions because of the finer feature size, were incorporated to minimise risk. Some adaptations were essential in the new technology since, for example, power supply voltages were only 0-2.5V, compared to 5V previously. Full advantage was taken of the favourable aspect ratio possible for the input transistor to maximise transconductance and improve signal to noise margins in the system, while maintaining or reducing the power budget. At the same time a series of ancillary chips used in the tracker control system were laid out for submission in the same shared processing run.

The benefits of using this technology appeared to be substantial, to set against the risks to the construction schedule. In addition to the anticipated radiation tolerance, good analogue performance seemed probable, with reduced noise at lower power than coarser technology chips. The fabrication turn-round was only a couple of months and the eventual production costs were estimated to be considerably lower. Several companies had been asked to tender for 0.25μm foundry services and the winning bid was submitted by IBM Microelectronics. This had an additional benefit: there was already in place a good working relationship between the company and CERN from a few long-standing contacts at an individual level.

The outcome was even better than had been hoped. The APV25, which is a large, complex chip, worked extremely well in the first iteration and needed only minor changes to fully meet CMS requirements [25, 26]. The control system chips were also successful, as were several other non-CMS circuits on the same wafers. This was encouraging, as it implied the process was well parameterised despite being primarily intended for digital circuits. The yield of good chips was also high although statistics were obviously limited.

LHC ASICs compared to LEP

It is interesting to compare the present generation of front-end readout ASICs with those produced for LEP. Of course, differences in the machine and experimental conditions, such as beam crossing rates and collision cross-sections and trigger rates, imply different parameters. However, despite the much higher LHC clock speed, power constraints on the amplifiers, where most of the power is consumed, have turned out to be similar simply because of the compromise between the minimum power needed to achieve low noise and the maximum heat removal which can be achieved by the cooling system. The original strong interest in bipolar amplifiers was indeed motivated by the belief that they would offer a more favourable noise at high speed and low power, while CMOS looked marginal for amplifiers until deep sub-micron processes were proven.

A big difference compared to the past was that LHC front-end circuits for tracking needed deep pipeline memories to store data until a first-level trigger initiated readout; it was a major innovation at the inception of the R&D phase to demonstrate that ring-buffer memories with corresponding control logic could be built [27]. These pipeline memories required CMOS technology for logic and storage capacitors which, if bipolar amplifiers were used, implied either two-chip solutions or a BiCMOS technology, which did not appear to be available in radiation hard form. The length of the pipeline was constrained by the desirable overall size of the chip and the storage capacitance which could be achieved in a small area, where thinner oxides were needed to maximise capacitance. The 3-4μsec expected by the LHC trigger developers was a major challenge for the front-end ASICs.

Although deep buffers allowed more events to be stored until the trigger arrived, it was also essential to aim for virtually deadtime free operation at 100kHz trigger rate since small losses accumulated in each detector sub-system would combine to produce much larger effects.

Many additional features have since been added to chips, as designers have gained more confidence. Among them are means to calibrate and program the circuits, thereby avoiding many off-chip components which take space in densely packed systems, as well as potentially increasing construction difficulties and contributing to unreliability or low yield. Internal control of bias currents, voltages and other parameters is possible, which allows alternative shaping times, changes of signal polarity, gains, readout modes or thresholds and can be switched on and off rapidly during automated testing. The deep sub-micron technologies are of tremendous importance in permitting this by providing both higher density and increased numbers of metal layers.

Another benefit is in the trend to standardisation by, for example, re-using circuit building blocks. This has, potentially, a number of advantages in addition to the obvious one of saving effort and cost. By using proven sub-circuit designs, risk can be reduced since, as experience is accumulated from building and operating detector systems, small features are often identified which sometimes need correction or alteration of the circuit design. These are rarely design faults but often involve factors which are outside the control or expectation of the original designer, usually unspecified in the original requirement. Some examples which have been experienced in recent years are subtle timing changes due to transistor temperature dependence (LHC silicon detectors operate well below room temperature), small differences between control signal responses due to minor improvements in later implementations of a standard protocol, speed effects due to different capacitive loading of control lines, currents drawn via protection diodes as control circuits are powered up before readout chips, and so on.

ASIC evaluation and testing

Given the scale of the LHC projects, chip evaluation has evolved considerably. Statistical parameter variations in production are inevitable and must be tolerated. Demonstration of a working circuit on the bench is no longer sufficient, even if usually supplemented by experience in beam tests, which was often the case in the past. In small systems, “work-arounds” might be found to mitigate unexpected minor features, by customisation of circuit boards or selection to allow use of chips with a few defective channels. For the LHC experiments, with their long required lifetime in a hostile and essentially inaccessible environment, much higher standards of chip acceptance are ideally required. This has been provided by the deep sub-micron technologies where it has proven possible to procure large numbers of very uniform chips of extremely high quality, allowing rejection of all chips outside narrow tolerances.

A good example is provided by the CMS silicon tracker readout chip, the APV25. An in-house testing station was developed over several years, eventually permitting two 200mm wafers containing 360 die sites to be tested in a (long) working day, spending about one minute per site [28, 29]. The lengthy development time was determined by the different technologies used prior to the APV25, the availability and procurement of equipment, and the experience gained from larger scale testing, involving feedback from users to refine further test criteria, rather than the development of software.

Of course, this probe testing was motivated by the need to screen large volumes of CMS die. It was complementary to many detailed chip evaluations carried out in customised test systems, mainly in one laboratory at Imperial College, which scrupulously investigated all aspects of the design, the operational parameter space, and issues which emerged from studies of silicon modules in labs and beams. It was followed by extensive irradiation studies to beyond LHC levels, and other reliability tests.

Many observers apparently do not appreciate the rigour which is applied to such studies as it is clear that “testing” is often considered to be merely a demonstration that a chip functions and the design is “correct”, not appreciating the complexity of the task and attention to detail needed. In view of this widespread attitude, it is essential to reduce this evaluation task by building where possible on well proven designs.

Process-dependent factors

The need for attention to detail was proven during APV25 production by a lengthy postscript to the chip development. It was observed during early large scale production that unexpected variations in yield between wafer lots occurred (fig. 1). Initially it seemed that something had gone wrong in manufacturing the first

production lots (each of 24 wafers) but after investigation, replacement of wafers and resumption of production, over a one year period it became clear that the yield was not stable. This was reinforced, although with limited statistics, by variations observed in other HEP projects, including multi-project wafer (MPW) submissions. Yet, despite accumulation of large amounts of data, plenty of ideas, and additional processing of test structures intended to validate hypotheses, no clear link could be demonstrated between the designs and the failure rate and no weak point in the processing could be identified. Nor were similar failures apparently common in the large volumes of commercial products being delivered to other foundry customers.

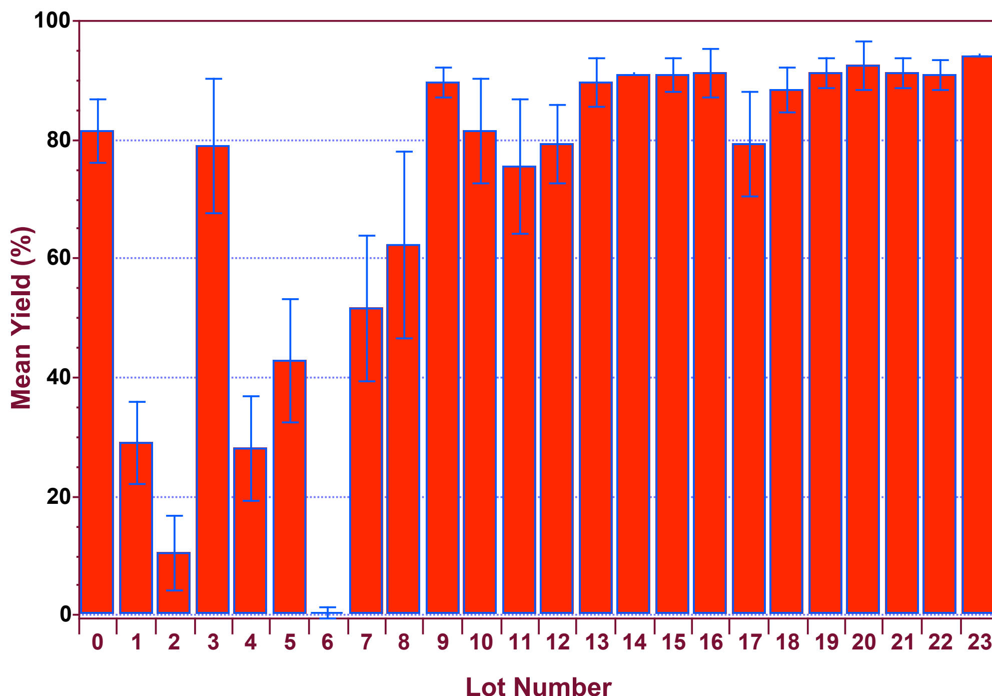

Fig. 1. *The yield of the APV25 as a function of lot number. Lot 0 is an engineering run of 10 wafers, and lots 1-3 are the first production lots. Lots 4-8 were used for investigations of the parameters affecting yield. Lots 9 onwards are those produced in the optimised process. Error bars represent statistical uncertainties. A few "lots" comprise a single wafer only.*

A series of regular intercontinental conference calls was set up between designers, evaluation team, foundry specialists and failure analysis experts from within IBM. A few points in some chips were located where clear failures could be isolated, even though it was far from evident that these were representative of the many different reject patterns observed. IBM used some sophisticated microscopy techniques to section the chips and examine their internal structure. They quite rapidly narrowed down the plausible hypotheses and finally proposed an explanation.

The problem was related to intensive use of one metal layer in HEP designs. The presence of this metal affected the duration required for a chemical-metal polishing (CMP) step responsible for wafer planarisation and, consequently, the minor variations in thickness could give rise to small gaps, e.g. ~0.1µm, between vias and metal layers which they were intended to connect.

Following a series of wafer lots in which CMP was varied, it was proven that the process could be better optimised for the APV25 and other designs. This was both a pleasure and surprise, since it had been believed that a process with the high volume throughput of the IBM foundries (typically ~40,000 wafers per month) would never be tuned for the small volumes required by HEP users (typically a few thousand wafers in total). Fig. 1 shows how successful this optimisation has been, with yields ~80% or more regularly observed.

The moral of the story appears to be that a good and close working relationship between foundry vendor and user is highly desirable, even mandatory. It would have been impossible in a reasonable period of time for HEP users to have replicated alone the IBM failure analysis expertise and to have diagnosed the fundamental problem, not to mention imposing the process modifications. Yet yield variations and unexpected process changes have been a feature of many past experiences with smaller scale ASIC production, usually without any final convincing explanation.

Further CMS developments

As a consequence of growing confidence in the 0.25µm process and, more importantly, ability to design in it as proven by the tracker developments, by 2002 CMS was confident enough to redesign another sub-system electronics, which was considered to be at risk of possible failure to meet demanding technical specifications while exceeding expected costs. The CMS electromagnetic calorimeter (ECAL) is a high resolution crystal scintillator system, read out with silicon avalanche photodiodes in the barrel and vacuum phototriodes in the endcap region where higher radiation levels are experienced. The target is to achieve ~0.5% energy resolution at high energies, over a 16 bit dynamic range. Resolution is dominated by electronic noise and ADC quantisation for photon energies up to 20-30GeV, while the constant term, determined by a multitude of minor effects such as temperature stability and calibration, dominates at all other energies. It is most important to ensure excellent resolution in the range where a low mass Higgs decaying into two photons is expected.

To save money and maintain the very stringent performance it was decided to develop three new 0.25µm chips, with a target of about one year from design to first results. This was considered ambitious by everyone, not least the team responsible for the new system, and it was considered a backup solution in case the original system did not evolve as hoped. The designs began in the second quarter of 2002 and testing began in May 2003. Exploitation of existing 0.25µm control ASICs developed for the tracker and a CERN Gbit/s 0.25µm optical link driver completed the system.

The front-end readout circuit is the multi-gain amplifier MGPA [30], followed by a 12-bit ADC [31], designed with the aid of a specialist design house. The digital processing and data handling ASIC (FENIX) is a configurable multi-function digital chip derived by VHDL translation into an ASIC[32]. All the chips worked successfully on the first iteration with impressive, close-to-specification performance, which was fine-tuned in a rapid second iteration. However, the initial performance of each component and the system was already impressive enough that, motivated also by the large cost savings, the "backup" solution was adopted as the baseline readout system in July 2003. One year later all the chips are in full scale production.

The future and relevant technology trends

Electronics technology is still advancing rapidly and our community is forced to follow technology trends, if only for simple reasons of availability. It is mandatory to maintain access to ASIC technology which is now vital for any accelerator-based HEP experiment. There are huge potential benefits from doing so, although one concern is the requirement to maintain enough specialist capability both in design **and** evaluation.

Some of the benefits are likely to be in lower power dissipation, even if this is partly offset by increased channel counts (power is the biggest single limitation to tracking systems where the material budget is dominated by electronics-related services) and improved performance, where higher speed and more functionality are expected. The future is also likely to bring new requirements; one important desire for an upgraded LHC is to include tracker information in the first-level trigger, requiring new readout and trigger architectures to be considered. Finally there will continue to be unexplored new possibilities, such as further developments of high speed opto-electronic technology, potential use of wireless, etc.

To understand the potential, it is also important to consider the constraints and necessity to investigate new technologies, especially given the expected investments required. Process evolution means that, although the present 0.25µm CMOS technology will probably remain available until about 2009, it is already being supplanted by 0.18µm and 0.13µm processes. Design and simulation software will continue to become increasingly complex and expensive given increased circuit density, numbers of metal layers and overall process complexity. Encouragingly, radiation hardness looks very promising, even without some of the design features used with 0.25µm [33, 34] although single event effects will increase in importance, but this can be offset by fault tolerant logic. 300mm wafers are becoming the next standard, and are already in use, with implications for equipment, such as probers, and bump-bonding.

Supply voltages reduce in range with each technology generation, and 0.13µm processes use 1.2-1.5V compared to 2.5V in 0.25µm. This is challenging for amplifier designs, where the dynamic range could be inhibited. It is also a challenge for detector design since reduced supply voltages means increased current for the same total front end power dissipation. Increased currents in cables imply larger voltage drops and

increased power losses through heating of cables. This will be hard to accommodate since space is already at a premium in the congested routes through the outer shells of experiments like CMS and heat removal by cable cooling is mandatory due to limits on air conditioning. It will probably be essential to develop new means of powering front-end systems, e.g. via switched mode circuits on detector modules and distributed local voltage regulation, with implications for noise avoidance.

Commercial circuits exhibit trends to higher speed and lower power, driven as they are by portable computing, automotive applications and communications, although such systems do not necessarily achieve both at the same time. Higher density means more digital logic is possible in smaller areas, which has implications for fault-tolerant circuits as well as size reductions. Often commercial circuits reduce the risk engendered by higher complexity by utilizing programmable features to tune, self-test, or even correct faults in the design. Since costs of engineering runs in deeper sub-micron technologies appear to increase dramatically, it may be essential to replicate such innovations.

CMOS Scaling

Although the implications of transistor size reduction are of most relevance to digital logic, it is useful to consider the consequences of reducing transistor feature sizes by a factor α. The main transistor parameters in circuit designs are length, width, oxide thickness, and voltage (L, W, t_{ox}, V) which are assumed to scale to L/α, W/α, t_{ox}/α, V/α.

If so, the number of transistors per unit area increases by a factor α^2 but power per transistor reduces by $1/\alpha^2$ so that power per chip remains approximately constant, assuming the same clock speed. Similarly, capacitance/area increases by a factor α, for the same dielectric, while interconnection resistance also scales as α, for the same metal, with the consequence that RC time constants scale as α^2. Not only does this provide a limit to increased chip speed but it motivates engineers to introduce new materials, such as copper and fluorinated glasses, to overcome this constraint. Then new issues could arise, for radiation tolerance for example, which need to be verified.

In the past it was thought that overall chip size reduction, influencing yield, would result but it now seems that larger chips are no longer as daunting, so more functions per chip, and increased integration of systems at chip level are prevailing trends.

Other technologies

Since commercial trends have been used to motivate the ASIC developments needed for HEP applications, should one also consider other technologies than CMOS, which remains dominant at present? One important technology which has been evolving rapidly in recent years is based on silicon-germanium (Si-Ge) heterostructure bipolar transistors [35]. The technology appears to be intrinsically very radiation hard and to have very high speed response for low power. Since no serious developments have yet been carried out for HEP applications it is hard to judge quantitatively the importance of these factors but one can at least ask whether this technology has other drawbacks compared to standard CMOS.

It is certainly already accessible, being available via Multi-Project Wafer runs from commercial vendors, so cost need not necessarily be the main obstacle to its use and perhaps sharing of non-recurrent engineering (NRE) costs may be controlled. However, in contrast, the LHC experience appears to point to significant benefits obtained from widespread use of single technology. Development of libraries and sharing of successful components and re-use of designs have already been shown to be advantageous, as well as detailed evaluation by a large community.

In the LHC context, "in-house" 0.25µm CMOS MPW submissions were organised through CERN and have been highly successful [36]. As of mid-2003, 184 designs involving 20 institutes, and 11 major wafer submissions had been launched, in addition to production orders. A single point of contact is maintained to the foundry, with regular technical exchanges, and contracts are organised in CERN on behalf of everyone. The possible down-side is that submissions are demand-driven so schedules can be less predictable than commercial MPWs. A design which is ready might be delayed by others which are not completed in time, which can be frustrating, even if there are cost advantages. It is also possible that chips with very small volume requirements can not only be developed but produced by means of such submissions, although this can encourage less cautious designers to iterate extensively to complete their chip design, somewhat

offsetting cost savings. To judge the cost-benefit balance requires a more comprehensive review of many projects than this author can offer at present.

Active pixel sensors

One other commercial trend which may be worth emulating is the displacement of some high performance, low volume technologies by low cost, high volume alternatives. Perhaps the most notable example is in digital imaging, where charge coupled device sensors have given way to active pixel sensors required for the proliferation of high quality, inexpensive digital cameras.

Active pixel sensors (APS) [37] are simply arrays where the light sensitivity of silicon is exploited to integrate a photodiode coupled to transistors in some part of the surface of an array of identical cells. This allows to integrate more sophisticated electronics into the sensor and fabricate a more complex device in a single processing line, with important reductions in assembly and handling costs. Such a concept is appealing to some of those who have been exposed to the rigours assembling large LHC systems and where CMS has already successfully pioneered automated assembly techniques for its silicon microstrip tracker for the same reason.

This is not yet a panacea for overcoming all difficulties in building large systems as HEP experiments inevitably involve significant customisation in ensuring large area coverage or maximal hermiticity by designing modules of various shapes to match the overall system geometry. The fact that CMOS pixel sensors are now commonplace does not mean that an immense amount of effort and resources has not been dedicated to developing and debugging the technology. Nevertheless, the trend to reduce HEP technical support means that more imaginative use of scarce resources must be found if our community is to continue to contemplate large scale systems. Labour intensive tasks must be reduced to the minimum in future, especially those which tie up rare specialists in critical points of the system. Examples have been seen in the assembly of detector modules, front-end hybrids and opto-electronic modules during CMS development and construction.

The concept of an integrated electronics and sensor has been a dream of many silicon detector specialists for a couple of decades [38], but until recently this was generally seen to be attainable by significant customisation of sensor fabrication processes to integrate readout circuits [39, 40]. In the APS approach, the opposite viewpoint is adopted by inserting a sensor into the readout circuit and exploiting a standard process, which usually in any case admits a range of variants within it, such as epitaxial layers which can be exploited for sensors. Minimum ionising signal detection has been demonstrated and several types of sensor for visible light, x-ray, and charged particle detection have been rapidly and reliably built by a few teams [41-43]. It is not yet clear if the inherent limitations for ionising particle detection are an obstacle to devising solutions for an environment like LHC. Certainly APS appears to be an attractive option for linear collider-type applications.

On grounds of cost, this approach seems also to be well motivated. The cost per unit area of production 0.25μm CMOS electronics is very similar to that of the relatively simple, but much larger area, silicon sensors used in CMS. However, the cost of assembling a complete detector module is considerably higher, being dominated by labour and tooling costs. The prospect of removing both cost and effort from the complicated module assembly procedure is remarkably enticing!

Conclusions

At least in accelerator-based experiments, HEP detectors would now be impossible without ASICs and an LHC upgrade is inconceivable without further evolution and development of ASIC technology. It is most likely that their use will be extended to many other experimental environments in the coming years. It is therefore vital that our community maintains this essential resource by continuing to invest in the engineers and specialists who support the physics, and developing more expertise among physicists themselves, as has already been the case with computing technology, precision mechanical engineering and assembly technology.

Acknowledgements

I warmly acknowledge all those individuals who have contributed to the deep sub-micron revolution of the last few years. In particular, I would particularly like to thank Sandro Marchioro for his seminal contributions and Mark Raymond, Marcus French and Federico Faccio for their strenuous efforts in design and evaluation. Peter Sharp has been a constant stimulus. Without the efforts and generous support of IBM Microelectronics and the ALTIS foundry this would not have been possible. CERN and PPARC (UK) have contributed greatly to the support of these activities.

References

[1] C. Damerell, these proceedings.
[2] E. Heijne and P. Jarron. Nucl. Instr. And Meths 226 (1984) 12
[3] R. Bailey et al. Nucl. Instr. And Meths 226 (1984) 56
[4] G. Barber et al. Nucl. Instr. And Meths 253 (1987) 530
[5] J. Walker et al. Nucl. Instr. And Meths 226 (1984) 200
[6] G. Hall. Reports on Progress in Physics 57 (1994) 481
[7] A. Schwarz. Physics Reports 238 (1994) 1
[8] E. Beuville et al. Nucl. Instr. And Meths 288 (1990) 157
[9] E. Beuville et al. IEEE Trans. on Nucl. Sci. 39 (1991) 766
[10] CMS Technical Proposal CERN/LHCC 94-38, LHCC/P1 (1994)
[11] Proc. First Workshop on Electronics for LHC Experiments. CERN/LHCC/95-56 (1995)
[12] Proc. Second Workshop on Electronics for LHC Experiments. CERN/LHCC/96-39 (1996)
[13] P. Winokur et al. Proc. Third Workshop on Electronics for LHC Experiments. CERN/LHCC/97-60 (1997) 48.
[14] M. Dentan et al. Proc. Third Workshop on Electronics for LHC Experiments. CERN/LHCC/97-60 (1997) 134.
[15] M. Dentan et al. Proc. Fourth Workshop on Electronics for LHC Experiments. CERN/LHCC/98-36 (1998) 79.
[16] T. Corbiere. Talk given at Fifth Workshop on Electronics for LHC Experiments (unpublished), Cracow, September 1999.
[17] M. Shaneyfelt et al. IEEE Trans. on Nucl. Sci. 41 (1994) 2536
[18] R. L. Pease. IEEE Trans. on Nucl. Sci. 43 (1996) 442
[19] G. Anelli et al. Proc. Third Workshop on Electronics for LHC Experiments. CERN/LHCC/97-60 (1997) 139.
[20] F. Faccio et al. Proc. Fourth Workshop on Electronics for LHC Experiments. CERN/LHCC/98-36 (1998) 105.
[21] W. Snoeys et al. Proc. Fourth Workshop on Electronics for LHC Experiments. CERN/LHCC/98-36 (1998) 114.
[22] N. Saks et al. IEEE Trans. on Nucl. Sci. 31 (1984) 1245
[23] N. Saks et al. IEEE Trans. on Nucl. Sci. 33 (1986) 1185
[24] W. Snoeys et al. IEEE Journal Solid State Circuits 35 (2000) 2018
[25] M. Raymond et al. Proc. Third Workshop on Electronics for LHC Experiments. CERN/LHCC/97-60 (1997) 158.
[25] M. Raymond et al. Proc. Sixth Workshop on Electronics for LHC Experiments. CERN/LHCC/2000-041 (2000) 130
[26] G. Hall. Nucl. Instr. And Meths A453 (2000) 353
[27] E. Heijne et al. Proc. Symposium on Det. R&D for the SSC, ed. T. Dombeck et al. World Scientific (1991) 465
[28] J. Fulcher et al. Proc. Fourth Workshop on Electronics for LHC Experiments. CERN/LHCC/98-36 (1998) 196.
[29] P.Barrillon et al. CMS Note *** (2004) , to be published.
[30] M. Raymond et al. Proc. Ninth Workshop on Electronics for LHC Experiments. CERN-LHCC-2003-055 (2003) 83.
[31] G. Minderico et al. Proc. Ninth Workshop on Electronics for LHC Experiments. CERN-LHCC-2003-055 (2003) 88.

[32] M. Hansen et al. Proc. Ninth Workshop on Electronics for LHC Experiments. CERN-LHCC-2003-055 (2003) 78.
[33] K Hansler et al. Proc. Ninth Workshop on Electronics for LHC Experiments. CERN-LHCC-2003-055 (2003) 71.
[34] F. Faccio et al. To be published in proceedings of RADECS 2003.
[35] J.D. Cressler and G. Niu. *Silicon-Germanium Heterojunction Bipolar Transistors*, Artech House, Boston 2003
[36] A. Marchioro. Proc. Ninth Workshop on Electronics for LHC Experiments. CERN-LHCC-2003-055 (2003) 10.
[37] E. Fossum. IEEE Trans. on Elec. Dev. 44 (1997) 1689.
[38] G. Zimmer. Nucl. Instr. And Meths 226 (1984) 175
[39] V. Radeka et al. Nucl. Instr. And Meths 275 (1989) 536
[40] S. Holland. Nucl. Instr. And Meths 275 (1989) 537
[41] R. Turchetta et al. Nucl. Instr. And Meths A458 (2001) 677
[42] R. Turchetta. Proc. Ninth Workshop on Electronics for LHC Experiments. CERN-LHCC-2003-055 (2003) 28.
[43] G. Deptuch et al. IEEE Trans. on Nucl. Sci. 49 (2002) 601