Imaging Detectors and Electronics – A View of the Future

Helmuth Spieler
Lawrence Berkeley National Laboratory, Physics Division, Berkeley, CA 94720, U.S.A.

Invited Keynote Talk
Fifth International Workshop on Radiation Imaging Detectors
University of Latvia, Riga, September 7-11, 2003

Abstract

Imaging sensors and readout electronics have made tremendous strides in the past two decades. The application of modern semiconductor fabrication techniques and the introduction of customized monolithic integrated circuits have made large scale imaging systems routine in high energy physics. This technology is now finding its way into other areas, such as space missions, synchrotron light sources, and medical imaging. I review current developments and discuss the promise and limits of new technologies. Several detector systems are described as examples of future trends. The discussion emphasizes semiconductor detector systems, but I also include recent developments for large-scale superconducting detector arrays.

Introduction

Gazing into the future is always risky, so I've backed off a bit by calling my presentation "a view". Views can reach far, or not so far, as shown in Fig. 1. Time will tell how perceptive this view is. Imaging detectors include a wide range of applications, but I will restrict the discussion to devices that emphasize high sensitivity or resolution. To gauge the rate of progress we might expect, it is useful to go back and see where we were 10 to 20 years ago.

Fig. 1 Views can reach far (left), or not so far (right).

1. Where were we two decades ago?

1.1. Sensors

For imaging visible light we had photographic emulsions and CCDs. Emulsions were still the mainstay of astronomical observations. The first CCDs with good quantum efficiency and the very high charge transfer efficiency needed for faint light imaging were just becoming available. Even after the power of these devices had been demonstrated, it took years for astronomers to adopt this new technology. Today, CCDs dominate the field (see [1] for an historical overview).

Imaging x-rays was the domain of wire chambers. Multiwire proportional chambers were common. Low-noise electronics still largely required discrete input stages, so many detectors utilized interpolation techniques, for example resistive or delay line readout, that provided many resolution bins with few electronic channels. Characteristic of these systems is that they used large sensors with a few channels of highly optimized electronics.

Traditionally, semiconductor detectors were used primarily as high-resolution energy measuring devices. In the early 1980s, however, a major change began with the introduction of the silicon strip detector [2]. Although these devices did make use of the high charge yield and fast response of silicon diodes, their prime purpose was position sensing. This development was driven by charm experiments that tagged interesting events by reconstructing displaced vertices, which required micron-scale position resolution. On silicon the electrodes could be patterned at the μm scale, so position resolution was obtained by segmentation, rather than interpolation.

Electronic detectors for position sensing were mostly one-dimensional devices. Two dimensional readout was obtained by combining two one-dimensional detectors oriented at right angles to one another. This projective geometry suffers at high hit densities, as multiple hits within the resolution time yield multiple coordinate combinations ("ghosts") that cannot be resolved without additional track information. CCDs provided unambiguous two-dimensional information and demonstrated their advantages in pattern recognition in NA32 [3]. Time Projection Chambers (TPCs) [4] provided three-dimensional non-projective information, but they were bulky, slow, and not suited for high rates or compact vertex detectors.

The projection of position onto time was also applied in a novel semiconductor sensor introduced in the early 1980s, the Si drift chamber [5]. This clever device still followed the traditional paradigm of minimizing the number of electronic readout channels.

The early 1980s also saw the development of the first concepts for random-access pixel devices [6]. These devices, which mate sensor pixels with individual front-end electronics channels have emerged as the architecture of choice for many applications.

1.2. Electronics

Electronics were still quite traditional, using a combination of discrete transistors and commercial integrated circuits. Both the noise and frequency response of monolithic integrated circuits were still marginal for demanding detector applications, so circuits using discrete components were quite common, although miniaturization reduced the size [7].

Silicon strip detectors initiated a dramatic change. The NA32 experiment at CERN pioneered the use of both silicon strip and pixel detectors in high-energy physics [3]. Strip pitches of 20 to 25 μm were used. Only every fourth strip was read out; the intermediate "floating" strips provided analog interpolation via capacitive division. Each electrode fed a preamplifier and shaper, so in initial implementations using discrete component hybrid technology the front-end electronics dwarfed the sensor. This prompted the development of custom-designed monolithically integrated circuits. The MicroPlex IC [8] pioneered the basic architecture used in many designs to this day. Many parallel analog channels on a chip are combined with an output multiplexer to reduce the number of readout lines. Since MOS fabrication processes typically don't provide high value resistors suitable for *RC* pulse shaping networks, the MicroPlex used switched-capacitor correlated double sampling, which at that time was novel for charged particle detectors. The MicroPlex was a high power design requiring about 20 mW per channel, so the IC could not be operated without special cooling. Low-power designs emerged soon thereafter. The MX1 chip [9] translated the MicroPlex circuit into CMOS using a longer – but fully adequate – shaping time that greatly reduced power requirements. The CAMEX64 was a lower-density chip with added flexibility in pulse shaping [10]. Concurrently, the SVX chip also used CMOS with a switched-capacitor correlated sampler, but added threshold detection and on-chip zero-suppression (sparsification), so only channels with hits were read out [11].

Superficially, the reduction in power in the MX and SVX chips has been attributed to the use of CMOS, rather than NMOS as in the MicroPlex. Although CMOS reduces the power in the digital stages, in analog circuitry the improvement is not as large. The main difference in the next generation ICs was the appropriate choice of design requirements, which is still the most important part of any design.

The Superconducting SuperCollider (SSC) prompted many investigations of detectors and electronics at high event rates. Although designed for the Tevatron, the SVX was the outgrowth of detector R&D for the SSC and served as a test bed for many concepts in use today. However, although SSC detector R&D funding was crucial for many of these developments, the biggest hurdle was convincing established workers in the field that designing a full-custom IC in a physics research environment was practical and that the effort required was comparable to what was needed for any complex system. However, the style of working had to change from the traditional "cut and try" to a more systematic design process with detailed simulations. The introduction of custom ICs in HEP was greatly facilitated by the introduction of foundry "brokerages', for example MOSIS, which provided economical access to IC fabrication through multi-project runs.

It is also worthwhile to remember that not all items on the detector wish lists turned out to be useful. One example is the monolithic integration of large scale sensors with electronics, commonly viewed as the "holy grail" at the time. Clearly, it is an appealing concept to have a 6 x 6 cm^2 detector tile that combines a strip detector and 1200 channels of readout electronics with only the power and data readout as external connections. The problem was perceived to be the incompatibility between IC and detector fabrication processes. Development of an IC-compatible detector process allowed the monolithic integration of high-quality electronics and full depletion silicon sensors without degrading sensor performance [12], including the implementation of full CMOS circuitry [13]. Nevertheless, a simple yield estimate shows that this isn't practical. In the conventional scheme reading out ~1200 channels with a 50 μm readout pitch requires 10 ICs with 128 channels each. These devices are complex, so their yield is not 100%. Even when assuming 90% functional yield per 128-channel array, the probability of ten adjacent arrays on the wafer being functional is prohibitively small. The integration techniques are applicable, however to simpler circuitry and have been utilized in monolithic pixel detectors [14].

1.3. Radiation Damage

Except for the nuclear weapons community, designers of imaging detectors could be blissfully ignorant of radiation damage. In high energy physics, event rates at LEP, the SLC, and the Tevatron were sufficiently low that radiation damage was not a concern. Radiation damage was observed in NA32 and in heavy-ion experiments in nuclear physics, but without leading to designs that systematically mitigated the effects of radiation damage. This changed with proposals for a new generation of high-luminosity colliders. The Superconducting SuperCollider (SSC) chose as high an energy as practical to reduce the required luminosity, but the luminosity of 10^{33} $cm^{-2}s^{-1}$ required for the physics goals still made radiation damage a big problem, especially in the tracking detectors. The LHC adopted a lower energy, but this increased the required luminosity ten-fold, so savings in the accelerator came at the expense of the detectors.

The effect of radiation (specifically displacement damage) on leakage current in semiconductor detectors was well-known and mitigation techniques were recognized [15]. Moderate cooling reduced the leakage current substantially. However, more important for the system was the use of segmentation, which reduces the leakage current per readout channel (and hence the shot noise). It also reduces the capacitance, which provides a higher signal-to-noise ratio and allows greater degradation with radiation.

The effect of displacement damage on doping properties, however, was poorly understood. We knew nothing of type-inversion and anti-annealing. Indeed, the prevailing advice from the experts was that silicon becomes intrinsic. Data on type inversion had been reported [16], but this work was not sufficiently appreciated.

In electronics, radiation-hard CMOS processes allowed operation into the Mrad regime and indicated the feasibility to the 100 Mrad level, but these techniques were controlled by the military and classified as secret.

2. Where are we today?

Strip detectors and custom ICs (now called ASICs) are routine in high-energy physics. They are used in all major experiments. Ambitious new projects are constructing huge arrays. The ATLAS SCT has 60 m^2 of silicon area with $6 \cdot 10^6$ readout channels and CMS has 210 m^2 with nearly 10^7 channels.

Random access pixel systems with unprecedented radiation tolerance are near the construction stage in ATLAS, CMS and LHCb. Radiation damage has proven to be difficult, but manageable. Defect engineering, the introduction of oxygen to reduce the electronic activity of defects, for example, has extended the lifetime of silicon sensors significantly. In electronics the use of industry standard "deep submicron" fabrication processes provides radiation resistance much better than previous classified radiation-hard processes. Pixel readouts have remained usable beyond 100 Mrad and a fluence of 10^{15} cm^{-2} minimum ionizing protons.

TPCs are still going strong, providing unexcelled pattern recognition in relativistic heavy-ion collisions. TPCs using liquid Ar or Xe provide exquisitely detailed images of particle interactions.

Furthermore, we see the technology of complex imaging arrays with customized monolithic electronics moving into other fields.

The intellectual and technical infrastructure required for in-house IC design makes it difficult for small groups to participate in these developments. However, when multiple groups agree on a set of common design requirements, the design can be performed at an appropriately equipped institution, and devices provided to the community. The "Medipix" chip is an excellent example, as demonstrated by papers at this Workshop.

This retrospective indicates the time scale for developments to move from concept to reality. *The things we talked about 20 years ago are now coming to fruition.*

We also see that the technology of imaging detectors is evolutionary, so we can expect future developments to follow a similar path. Many diverse technology developments have contributed, but a few basic trends have emerged.

- Segmentation
  Detectors are subdivided into many small elements. This increases overall rate capability and reduces electronic noise.
- Parallelism
  Many identical electronic channels operate simultaneously.
- Complexity
  Detector ICs today combine many channels of low-noise analog front-ends with digital circuitry, including on-chip analog-to-digital conversion. However, increased functionality requires process control and reliability. This implies that

circuits must be implemented in well-controlled commercial foundry processes (no exotics!).

Our experience also reinforces another lesson, which is by no means new:

Even when you think it will take longer, it will take longer than you think.

3. Where do we want improvements?

Although one might hope for the "miracles of modern technology" to solve all technical problems, this is not what brought us to where we are. Of course, we'd like to have detectors with infinitely good energy resolution and perfect efficiency, but detectors and electronics are subject to some basic constraints. Advances in technology have facilitated what has been accomplished, but much of the progress we have witnessed has more to do with better insight and experience in applying available technology. If the trend to larger and more complex systems continues, there are some core – and sometimes mundane – technical issues to deal with.

3.1. Increased functionality

The increased application of large-scale semiconductor systems in fields beyond high-energy physics will require improvements in precision. High-resolution x-ray spectroscopy requires lower noise and better baseline stability at high random rates. This increases circuit complexity and will require on-chip digitization with greater precision than in current systems. Some applications will also require improved time resolution.

Enhancing the adaptability of readout ICs to different experimental conditions will open their use to multiple experiments. A single design will not meet everyone's needs, but could address classes of experiments and put the technology within reach of small groups. However, increasing the functionality of readout ICs will be limited by power dissipation, unless circuitry becomes more efficient.

3.2. Power dissipation

Power dissipation translates into cooling requirements and cross-section in the power cabling, both to limit voltage drops and power dissipated in the cables. Power dissipated in ICs adjacent to a semiconductor sensor will drive up the leakage current with temperature, and thus the electronic noise. More efficient front-end circuitry can provide lower noise or higher speed. Conversely, one can keep the same level of functionality and apply the benefits of more efficient circuitry to simplified cooling and cabling systems.

3.3. Radiation resistance

Luminosity upgrades for the LHC are already being discussed, but as applications for large scale systems proliferate, so will radiation requirements. The need for a ten-fold improvement is already on the horizon.

3.4. Simplified construction

Although complex integrated circuits capture most of the attention, system integration remains the area where most systems experience major setbacks. Weaknesses in this area often impair system performance, but always incur penalties in construction time and cost. Robustness against cross-talk is crucial and can be addressed by circuit design and on-chip functionality. Reducing the number of mechanical connections also helps. In pixel arrays, eliminating the need for bump bonding would be a great simplification, but may not be practical when the combination of thick sensor layers with complex circuitry is required.

3.5. Cost

Large-scale semiconductor detector systems are chronically underfunded. Increased functionality does not necessarily increase cost, but it does require more preparation and up-front R&D funding.

4. Power-efficient design

4.1. Sensors

The sensor must be considered together with the electronics. The equivalent noise charge

$$Q_n^2 = i_n^2 F_i T + e_n^2 C^2 \frac{F_v}{T} ,$$

where i_n is the spectral noise current density at the input, e_n is the noise voltage density, C the total capacitance at the input, and T the shaping time. F_i and F_v are shaper parameters determined by the pulse shape [17]. If the noise current i_n is made negligible, by cooling the detector, for example,

$$Q_n \approx e_n C \sqrt{\frac{F_v}{T}} \propto e_n C = e_n \varepsilon \frac{A}{d} .$$

Here ε is the dielectric constant, and A and d are the sensor's active area and thickness. The energy resolution is the product of the noise charge Q_n and the energy required to form an electron-hole pair E_i

$$\Delta E = E_i Q_n \propto e_n E_i \varepsilon \frac{A}{d} .$$

Thus, for a given pulse shaper and sensor geometry, constant noise requires that the product of the input noise voltage and the sensor's ionization energy and dielectric constant remains constant. As will be shown below, at best the power dissipation scales

inversely with the square of the required noise charge $P \propto 1/Q_n^2$, so the desire for wide band-gap materials tends to carry a substantial penalty in front-end power. To some degree this can be alleviated by segmentation, i.e. reducing the area electrode area *A* per channel. Nevertheless, the first step towards improving electronic noise is in the detector.

4.2. Electronics

Both the equivalent input noise voltage

$$e_n^2 \approx \frac{4kT}{g_m}$$

and the gain-bandwidth product

$$f_0 = \frac{g_m}{2\pi C_o}$$

depend on the transconductance g_m of the input transistor [17]. The capacitance C_o at the node where voltage gain obtains invariably limits the obtainable circuit rise time, rather than the inherent speed of the transistor. From this we see that increasing transconductance improves both noise and speed. The transconductance depends primarily on device current. In a bipolar transistor

$$g_m = \frac{I_C}{kT/e},$$

where I_C is the collector current, *k* the Boltzmann constant, *T* the absolute temperature and *e* the electronic charge. In a MOSFET in strong inversion

$$g_m = \sqrt{\frac{W}{L} \cdot \frac{\varepsilon_{ox}}{t_{ox}} \mu \cdot I_D},$$

so for a given device width *W*, reducing the channel length *L* or gate oxide thickness t_{ox} should increase the transconductance. The choice of bulk material determines the carrier mobility μ and the gate oxide's dielectric constant ε_{ox}. However, this simple scaling rule only applies in strong inversion, whereas MOSFETs in large detector arrays are best operated in weak or moderate inversion. In weak inversion, the dependence of transconductance on current is the same as for a bipolar transistor, so it depends only on current and not on device geometry. The moderate inversion regime is the transition from weak inversion (low current) to strong inversion (high current) and its dependence is more complicated.

Since transconductance sets both the noise and speed, power efficiency improves when we increase the ratio of transconductance to drain current g_m / I_D. Increasing the device width *W* at constant current density is equivalent to connecting multiple devices operating at the same current in parallel, so to yield a universal curve Fig. 2 shows the normalized transconductance g_m / I_D vs. normalized drain current I_D / W. These data were measured on devices with channel lengths ranging from 0.8 to 25.2 μm, all on the same chip and fabricated in a 0.8 μm process. At low currents g_m / I_D is constant, as predicted for weak inversion. At the upper end of the current scale, in strong inversion, g_m / I_D is much

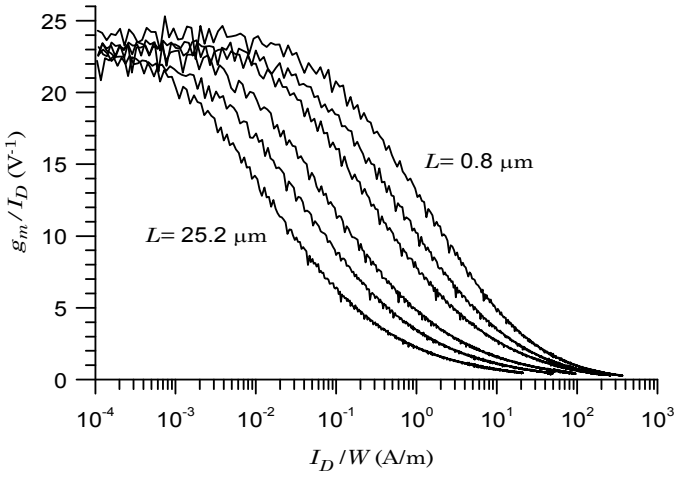

Fig. 2 Normalized transconductance g_m / I_D vs. normalized drain current I_D / W for channel lengths of 0.8, 1.2, 2.0, 5.2, 10.0, and 25.2 μm. All devices were fabricated on the same die in a 0.8 μm CMOS process. The transconductance is determined by differencing the raw measured I_D vs. V_{GS} data, so the irregularities in the curves are due to the differential non-linearity of the digitizer in the measurement system.

smaller. For example, the 0.8 μm channel length device shows $g_m / I_D = 24$ at $I_D / W = 10^{-3}$ and $g_m / I_D \approx 1$ at $I_D / W = 100$. The transition from weak to strong inversion shifts to higher currents as the channel length is reduced. At $I_D / W = 0.1$ the 0.8 μm long device yields $g_m / I_D = 21$, whereas 25 μm long devices yield $g_m / I_D = 6$. Thus, reducing the channel length allows more efficient circuitry, although not as predicted by the strong inversion formula.

The best power efficiency obtains at the highest normalized transconductance g_m / I_D that will provide the desired noise level. Uniquely associated with this value of g_m / I_D is a current density $(I_D / W)_{g_m / I_D}$, which for a given technology depends on the channel length. While keeping the current density constant, one can adjust the width to change the transconductance. As the width is changed the drain current $I_D = W \cdot (I_D / W)_{g_m / I_D}$ changes proportionally. This value of drain current sets the transconductance $g_m(W) = W \cdot (I_D / W)_{g_m / I_D} (g_m / I_D)_{selected}$. Thus, both the drain current and the transconductance scale proportionally to width, as does the FET's input capacitance. As the width is increased the equivalent noise charge decreases until the input capacitance equals the sensor capacitance. With further increases in width the increase in capacitance outweighs the decrease in noise voltage, so the noise charge increases. If the minimum noise is too high, one chooses a lower value of g_m / I_D, which will achieve a given transconductance at a smaller device width, so capacitive matching will occur at a higher transconductance. Thus the minimum noise will be lower, albeit at the expense of power dissipation. This procedure is illustrated in Fig 3.

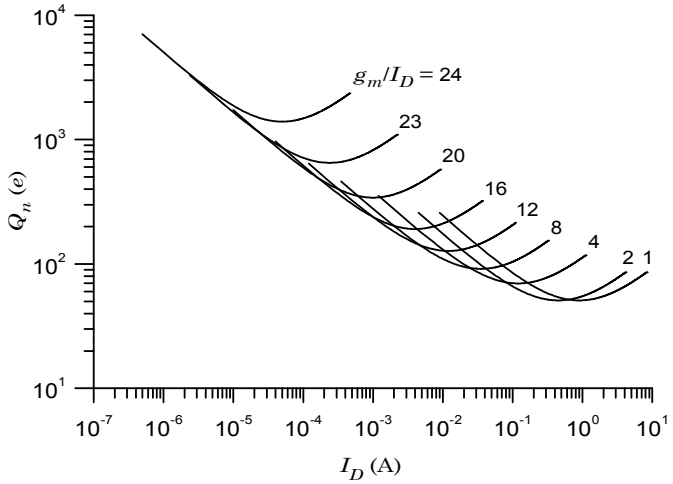

Fig. 3 Equivalent noise charge vs. drain current for various current densities I_D / W . The calculation assumes a detector capacitance of 10 pF, a shaping time of 100 ns, and a transistor input capacitance of 1 fF per μm width. In the low-current regime the asymptote for all curves follows the relationship $I_D \propto 1/Q_n^2$.

For example, assume that the desired noise level is 1000 *e*. A normalized transconductance $g_m / I_D = 24$ (weak inversion) allows a minimum noise of 1400 *e* at a drain current of 50 μA. Increasing the current density so that $g_m / I_D = 20$ shifts the operating mode towards moderate inversion and yields a minimum noise of 340 *e* at a drain current of 1 mA. However, following the $g_m / I_D = 20$ curve to smaller drain currents (device widths) provides the desired 1000 *e* noise level at a drain current of 30 μA, less than the 50 μA needed for 1400 *e* noise at $g_m / I_D = 24$. Going to much smaller values of g_m / I_D yields the desired 1000 *e* noise at higher currents. The choice of g_m / I_D is not very critical; $g_m / I_D = 22$ or 23 gives practically the same result.

This illustrates that capacitive matching is not a good criterion for systems where low power is important. Near capacitive matching the device width (and hence the current) can be reduced significantly without a substantial increase in noise. For example, at $g_m / I_D = 24$ allowing a 10% increase in noise reduces the device current to 40% of the current at capacitive matching. For currents well below the noise minimum all curves follow the relationship $I_D \propto 1/Q_n^2$, so for constant supply voltage the required power increases with the inverse square of the required noise charge, which depends on the signal magnitude provided by the sensor.

To what extent do improvements in device technology improve this situation? Can we simply rely on Moore's Law to meet future needs? The basic scaling rules still apply, but we seek improvements in the normalized transconductance g_m / I . In bipolar transistors this ratio is set by basic physics, so it is unaffected by improved process technology,

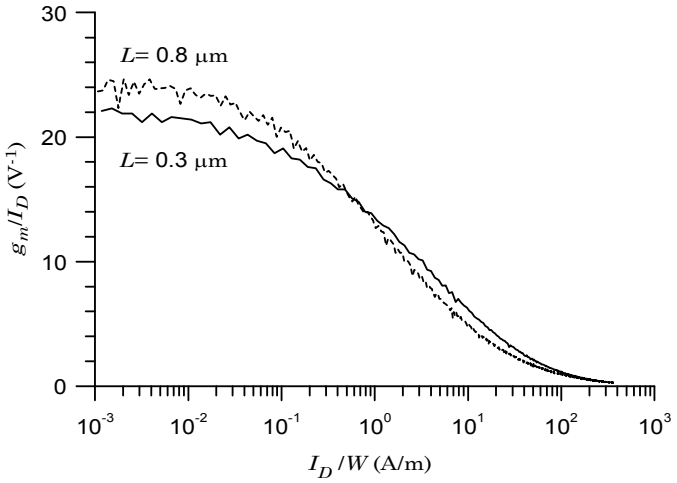

Fig. 4 Comparison of normalized transconductance g_m / I_D vs. normalized drain current I_D / W for NMOS devices with L= 0.8 and 0.3 μm channel lengths, fabricated in 0.8 and 0.25 μm CMOS processes, respectively.

increased device speed, or the use of heterojunction devices, e.g. SiGe devices. However, improved process technology can reduce parasitic noise sources, such as the base or emitter resistance. Perhaps even more important is the reduction of trap density in the base-emitter junction, which improves the DC current gain at low currents and thus extends the usable operating range to lower currents.

In MOSFETs the transition to smaller feature sizes should move the weak inversion regime to higher current densities. Fig. 4 compares measured data for two devices; a MOSFET with 0.3 μm channel length fabricated in a 0.25 μm process and a device with 0.8 μm channel length fabricated in a 0.8 μm process. The transition from weak to moderate inversion in the 0.25 μm MOSFET occurs in the same current range as in the 0.8 μm device and the normalized transconductance in weak inversion is distinctly lower.

Why does the 0.3 μm channel length not show the expected improvement? Scaling to smaller feature size involves more than lateral scaling, i.e. resolution in lithography. The vertical dimensions, i.e. the depth of the source and drain implants must also be reduced to avoid spreading the channel into the bulk, which reduces transconductance. The gate oxide must also be thinned. All this reduces the maximum operating voltage. In digital circuitry this implies smaller logic swings, so threshold control and noise immunity are concerns. In analog circuitry the dynamic range is reduced, as the maximum signal level is reduced while the electronic noise levels remain essentially the same. In some fabrication processes this is addressed by providing two choices of oxide thickness to allow “low-voltage” and “high-voltage” devices. Clearly, this comes at the expense of process complexity.

The point of this discussion is to emphasize the importance of transconductance and its relationship to power dissipation. Low electronic noise levels require sufficient transconductance coupled with acceptable input capacitance. In addition, large detector arrays require that these parameters obtain at low power. These requirements militate against many novel technologies that appear to simplify fabrication and reduce cost. Examples are amorphous silicon transistors or thin film transistors deposited by inkjet printing. All of these devices suffer from low mobilities and, hence, low transconductance. Nanotechnology offers the potential of very small devices, and the notion of "self assembly" will appeal to anyone who has constructed a complex detector, but nano-transistors will require nano-sensors to reduce capacitance to match the small transconductance. Nanometer thin sensors in trackers yield correspondingly small signals, which require lower noise and drive up front-end power. These novel devices will make inroads as switching devices, but their applicability to low-noise analog circuits is dubious. For the applications considered here crystalline devices appear to offer the most realistic prospects for technological improvements.

5. Prospects for Electronics

Improvements in analog performance with reduced feature size are not clear. In MOSFETs reduced gate oxide thickness and shallower source and drain implants could enhance low-current performance, but at the expense of gate leakage current and dynamic range. Reducing the thickness of the gate oxide also increases the input capacitance without a commensurate reduction in equivalent input noise voltage. Reduced feature size does provide substantial benefits in digital circuitry, both in circuit density and power.

Bipolar transistors provide the highest transconductance per unit current, in practice outperforming MOSFETs even in the weak inversion regime. Furthermore, bipolar transistors tend to have substantially lower input capacitance for comparable noise levels, further reducing power requirements. Transconductance per unit current in bipolar transistors is independent of technology, but high-density processes tend to improve contamination control, which improves DC current gain at low currents. Furthermore, faster devices tend to reduce parasitic base and emitter resistances. In the past, bipolar processes have suffered from low circuit density, but the cellular telephone market has promoted mixed technology BiCMOS processes, which combine high frequency SiGe bipolar transistors with high-density CMOS. This is a very attractive option.

Improvements in digital circuit power dissipation and circuit density in future CMOS processes will facilitate enhanced digital circuitry. High-resolution ADCs will still require larger feature sizes in the analog portions, but the logic and readout circuitry will require less die area. On-chip digitization also allows digital signal processing to provide functions that are difficult to implement with analog circuitry. One example, implemented in the readout for the ALICE TPC, is multiple tail cancellation [18]. Reducing the area required for event buffering, control logic, and readout is very beneficial in reducing the material in pixel arrays, as illustrated below.

6. Current projects that exemplify future directions

As noted in the introduction, progress in detector systems depends as much on architectures and design innovation as it does on new technologies, if not more. Next I discuss a few examples that I believe exemplify some future directions.

6.1. Random Access Pixel Detectors

Large-scale random access pixel devices are currently being designed for ATLAS, CMS, and LHCb at the LHC and BTeV at the Tevatron. As an example, Fig. 5 shows the layout of the ATLAS pixel detector [19]. Fig. 6 shows a module. A module consists of a 6 by 1.6 cm^2 silicon sensor wafer, onto which 16 readout ICs are bonded by an array of solder bumps [20, 21]. The output pads of the readout ICs extend beyond the edge of the sensor, to allow wire bonding to a flex-hybrid, which accommodates bypass capacitors, a readout controller IC, and power, control and readout bussing. The overall pixel detector has about 1 m^2 of sensor area and 10^8 channels. The pixels are 50 x 400 μm^2, but the production version of the readout IC implemented in 0.25 μm CMOS could accommodate the electronics in half the size. The sensor design had been "frozen" before the high-density IC was available.

Each pixel cell contains a preamplifier, shaper with 100 ns peaking time, threshold comparator, trim-DAC for pixel-by-pixel fine adjustment of the threshold, time stamp logic, and event buffering. The die is 7.3 x 10.9 mm^2 in size and contains 2880 pixels. Time-over-threshold using the 40 MHz clock provides coarse amplitude digitization. Fig. 7 shows a reticle containing two pixel ICs, a readout controller, and support and test devices. On the pixel ICs the upper 75% are the pixel cells, whereas the lower 25% are readout logic and output drivers. Higher density processes would reduce this area and also allow smaller pixel cells. Tiling is accomplished without dead area by making the

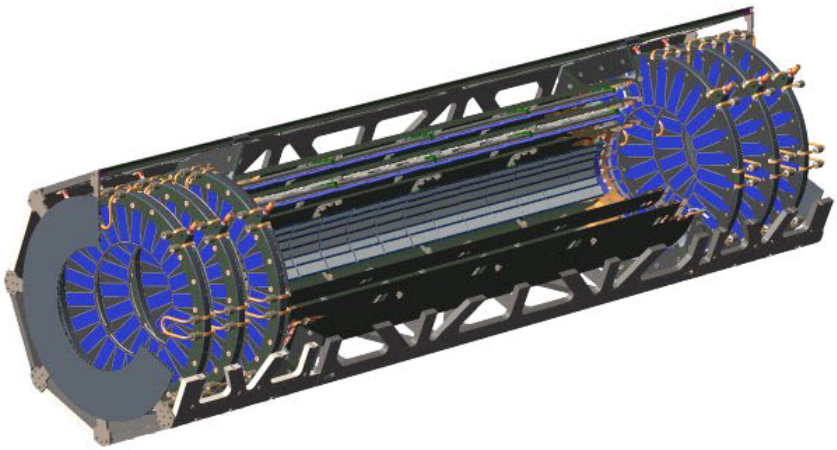

Fig. 5 The ATLAS pixel detector. The length of the detector is 1.4 m and the radius of the outermost pixel layer is 12 cm.

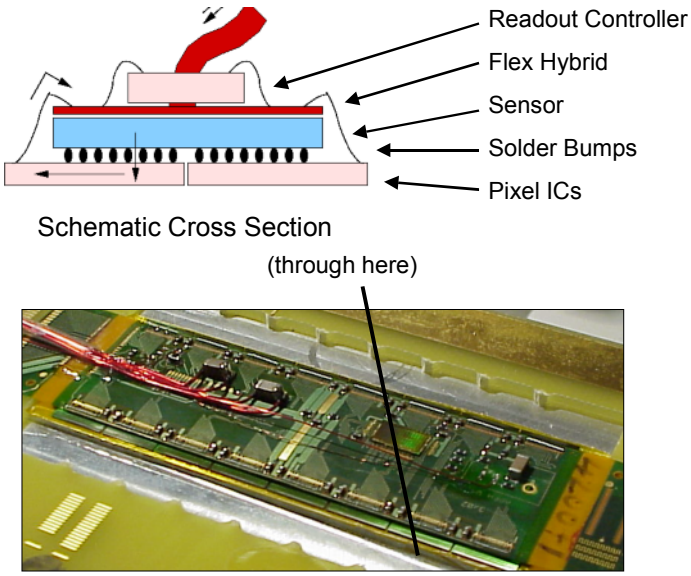

Fig. 6 An ATLAS pixel module. In the cross-section view the pixel ICs are at the bottom, bonded to the sensor above through a two-dimensional array of solder bumps. A polyimide flex-hybrid on top of the sensor has traces for bussing, bypass capacitors, and a readout controller IC. Connections from the pixel ICs to the flex hybrid are by wire bonds. The module includes 16 pixel ICs with 2880 pixels each. The complete detector include about 1000 modules with ~10^8 pixels.

sensor pixels at the readout chip boundaries larger to bridge the gap. The electronic noise is about 150 *e*, obtained during simultaneous readout at a 40 MHz rate.

Design and construction of these large-scale pixel systems pose formidable challenges. The ATLAS pixel IC contains 3.5 million transistors, so simulation and design verification are crucial. As in all large-scale semiconductor detector systems, electro-mechanical integration – combining sensors, electronics, cabling, cooling, and mechanical support systems – is a major part of the project. The complexity of integrating these systems is usually not appreciated by those who haven't done it. Furthermore, these systems are chronically underfunded, as funding agencies, reviewers, and project managers underestimate the required effort.

The hybrid structure has the drawback of requiring bump bonding. Currently, only a few vendors provide this service at the fine pitches required. The cost and technical overhead are barriers for small projects. Industry is moving towards smaller bonding pitches [20,21], so this situation could change. Apart from this, the hybrid array, which combines separate sensor and readout units, has many advantages. It allows the use of non-silicon sensors, which is crucial for many x-ray imaging applications. Tiling is facilitated, as the sensor can "bridge" the gaps between the readout ICs. Furthermore, it places no special requirements on the technology of the readout IC, e.g. the thickness of epi-layers that can be used as sensors [22, 23]. Monolithic pixel devices are in widespread use in optical imaging ("active pixel arrays" or "CMOS imagers"). Their small sensitive depth limits their use in charged particle and x-ray detection. Since they rely on diffusion in unde-

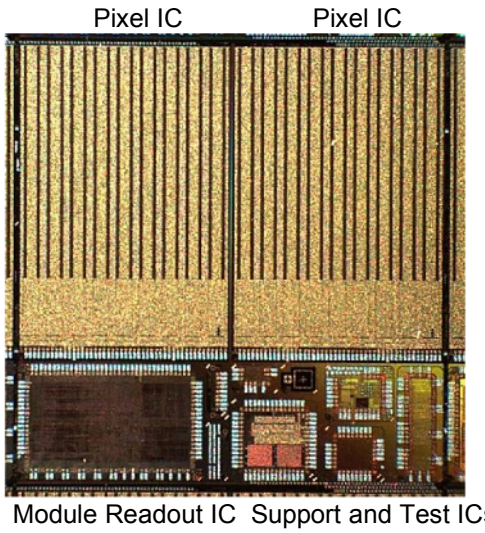

Fig. 7 A reticle of the pixel IC wafer, showing how multiple ICs are accommodated in one reticle. Reticles are copied by a step-and-repeat process to fill the entire 200 mm wafer. Two pixel ICs are at the top, with the readout controller and test/support ICs at the bottom. The pixel IC is 7.3 x 10.9 mm^2 and contains 2880 pixels.

pleted material, they are quite sensitive to minority carrier lifetime, which limits their radiation resistance.

6.2. Fully depleted CCDs

Although the monolithic integration of sensors and electronics on high-resistivity silicon was presented as an evolutionary dead end, it did lead to a novel CCD structure that is being applied to faint-light imaging in astronomy [24]. Front-side illumination of CCDs limits the quantum efficiency, because of absorption in the metallization and charge transfer structures. However, since in conventional CCDs the substrate is field-free, back-

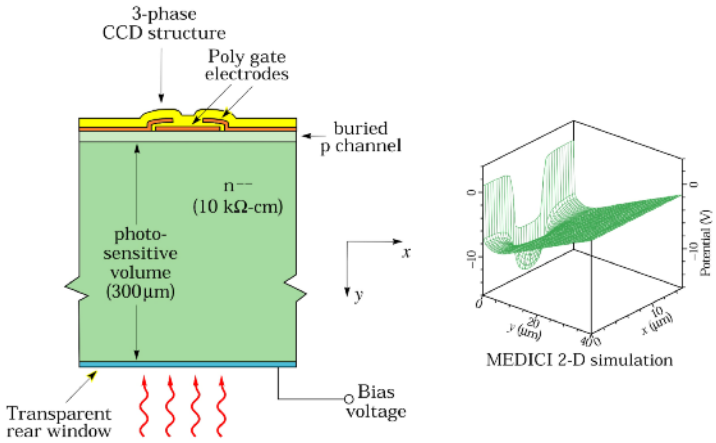

Fig. 8 Fully depleted CCD structure (left) and potential distribution into the bulk (right). (courtesy S.E. Holland)

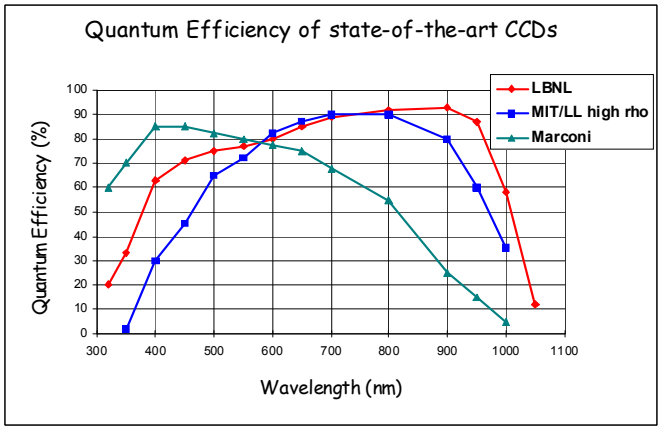

Fig. 9 Quantum efficiency of a thinned CCD, a partially depleted CCD, and a fully depleted CCD with 300 µm sensitive thickness. (courtesy S.E. Holland)

For astronomical observations the 300 µm thickness has the very important advantage of improving the red response, as shown in Fig. 9. Since the interstellar dust absorbs in the blue, the extended red response significantly enhances imaging sensitivity [25]. Radiation resistance is also good; devices have been tested to fluences of 10^{11} cm^{-2} (12 MeV protons) [26].

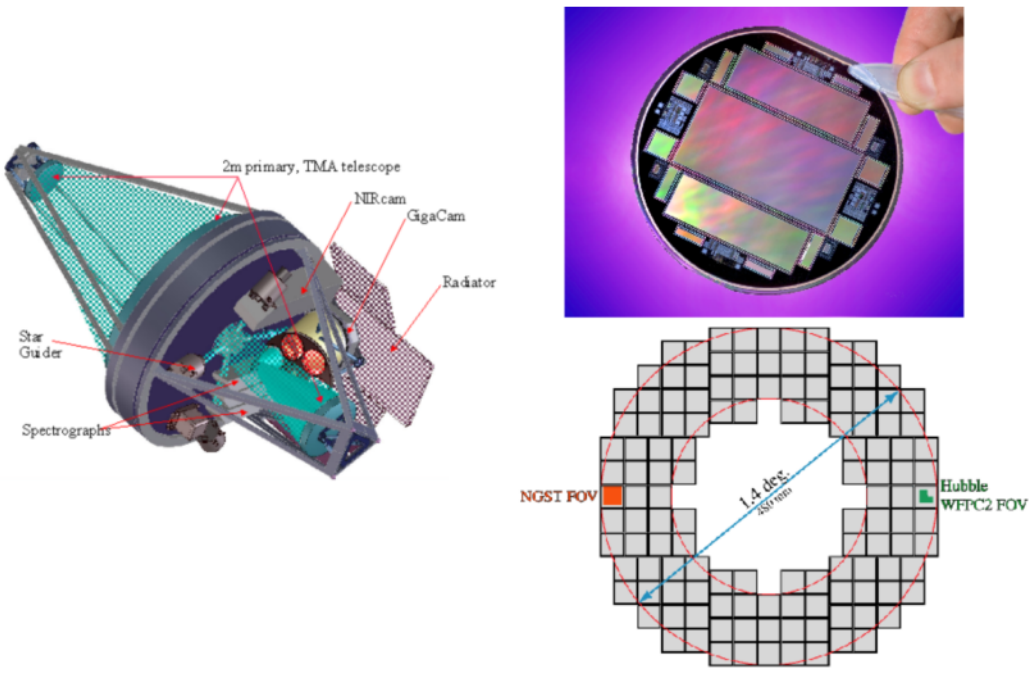

Fig. 10 The proposed SNAP telescope (left). A 100 mm wafer with CCDs and test structures is shown at the top right with the proposed SNAP focal plane mosaic below. (courtesy S.E. Holland)

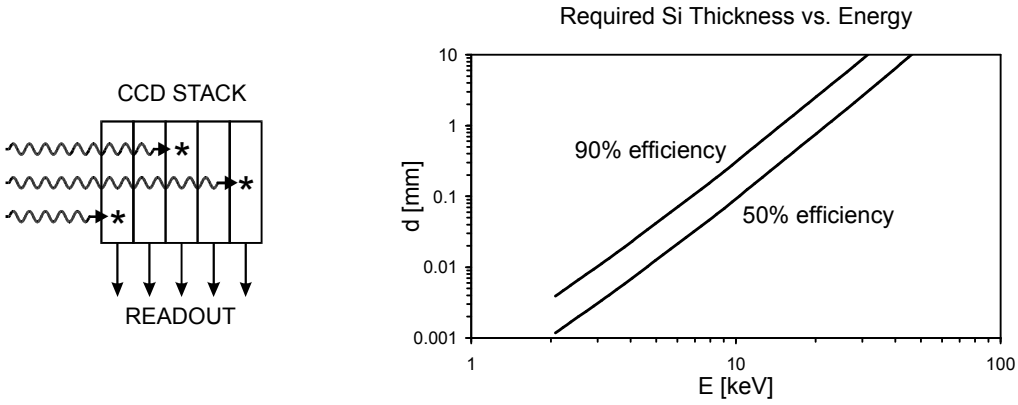

Fig. 11 Fully depleted CCDs can be stacked to improve detection efficiency (left). The required thickness for 50 and 90% photoelectric absorption efficiency is shown at the right.

The combination of technologies developed for high-energy physics [12] and medical imaging [27] have led to the enabling technology for a proposed satellite observatory, the SuperNova Acceleration Probe (SNAP) [28,29]. The conceptual design of the satellite and the "giga-pixel" imaging array are shown in Fig. 10, together with a wafer showing a 2K x 4K CCD fabricated at LBNL. This project utilizes not just the technology spin-off for the CCDs, but also the experience from high-energy physics in custom IC design for the readout and low-mass electro-mechanical integration.

Full depletion CCDs are also excellent for x-ray imaging, but the 300 μm sensitive depth limits the usable energy range to about 10 keV. However, these devices can be stacked to provide a fully active detection volume, as shown in Fig. 11 together with a plot of required thickness vs. energy. A stack of 30 CCDs would provide >50% efficiency up to 40 keV.

6.3. Hybrid detector systems

The technology developed for large-scale silicon detector arrays can enable breakthrough performance in other types of detectors. One example is a high-rate detector for photoelectron spectroscopy at synchrotron light sources [30]. The arrangement is shown in Fig. 12. Photoelectrons emitted from the sample pass through an electrostatic analyzer, which translates energy into position. A position-sensitive detector in the focal plane that registers the count rate provides the energy spectrum. A typical detector uses a microchannel plate (MCP) as an electron multiplier followed by a CCD, which suffers both from long readout times and non-linearities [30].

The new detector replaces the CCD by an array of strip electrodes connected to a fully parallel readout using a front-end IC, the CAFE chip, originally designed as a prototype

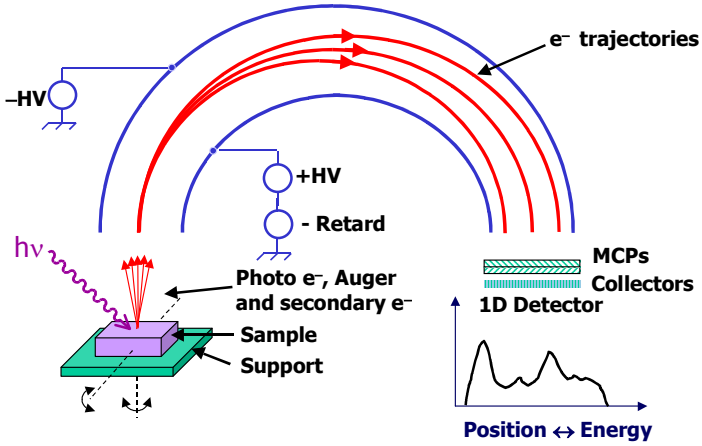

Fig. 12 High-speed photoelectron spectrometer.

for the ATLAS Semiconductor Tracker [31]. Each front-end channel includes a low-noise preamplifier, a shaper with a 25 ns peaking time, and a threshold comparator. This IC, fabricated in bipolar transistor technology, has a noise level <1000 *e* and a double pulse resolution of 50 ns, so each channel can accept random rates $>10^6$ s^{-1}. The second IC includes a 16-bit counter per channel with a double-buffered readout, so data can be read out during data acquisition without incurring dead time. 768 strip electrodes at a pitch of 48 μm provide an energy resolution $\Delta E/E=10^{-4}$ and a maximum total count rate of 2 GHz. The fast readout allows time-resolved measurements on a time scale as low as 150 μs. The focal plane electronics are at voltages up to 1.5 kV, so special precautions are crucial to prevent damage to the ICs in the event of discharges or sudden ramp-downs.

This scheme can be extended to two-dimensional imaging by replacing the strip electrodes and readout ICs by a monolithic pixel array. The collection electrode in each pixel cell can take up but a fraction of the pixel area and still collect all electrons from the MCP. This system would increase the count-rate capability to $\sim 10^{12}$ hits/s, while providing two-dimensional imaging. A 0.13 μm CMOS process allows 50 μm pixels, with about 50 million transistors per chip [32].

Bringing this detector to reliable operation was a very lengthy process, with several flawed iterations. Technology was not the problem; this detector uses technology that was mature and readily available a decade ago. The problem was one of sociology. The synchrotron light source community has little experience with the design of complex electronic detectors. A detector of this type requires an engineering team with the relevant experience and expertise, which requires appropriate project funding. Once this was put in place, the project progressed successfully, at less cost than the preceding low-budget attempts..

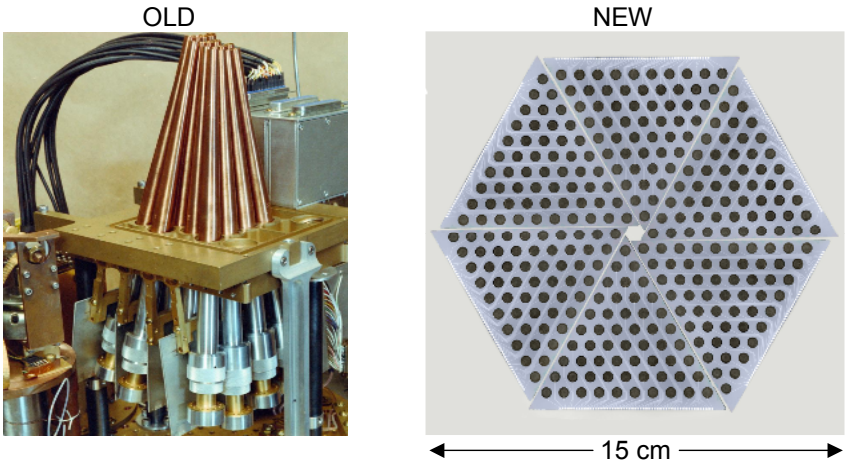

Fig. 13 Comparison of the 16 bolometer array used by MAXIMA and the new 300 bolometer array under construction for APEX-SZ.

6.4. Superconducting detector arrays

The final example uses a different technology in a different field, imaging the Cosmic Microwave Background (CMB) in experimental cosmology. The universe is permeated by background radiation from the Big Bang, today at 2.7 K. . The 2.7 K background radiation shows anisotropies at the 10^{-5} level, which are the seeds of matter formation, in turn leading to stars and galaxies. The angular scale of the temperature distribution shows multiple peaks, which provide information on the geometry of the universe (from the CMB we know that the universe is "flat"), the baryon density, and other fundamental cosmological parameters. Bolometers, i.e. microcalorimeters, are used to map the temperature variations across the sky. Mapping the polarization of the CMB is one of the next goals.

The CMB power peaks at about 200 GHz. The signal can be detected either directly in sensors that convert the absorbed thermal power into an electrical signal, or via antennas coupled though transmission lines to load resistors, which in turn heat the bolometer. Today, bolometers operating at sub-Kelvin temperatures are sufficiently sensitive that signal fluctuations are dominated by the shot noise of the CMB photons. However, future experiments require orders of magnitude improvements in sensitivity. Sensitivity can be increased by extending the measurement time and by performing many measurements simultaneously, which brings us to array technology.

In the past bolometers have been hand-crafted and difficult to operate. However, recent developments have changed this picture and brought large-scale bolometer arrays to the realm of practicality [33]. The bolometers used in the detectors discussed here are superconducting transition-edge sensors (TES), where a thin superconducting film is electrically biased, so its temperature is at the transition from the superconducting to normal state. This operating point provides a large change in resistance for a small change in temperature.

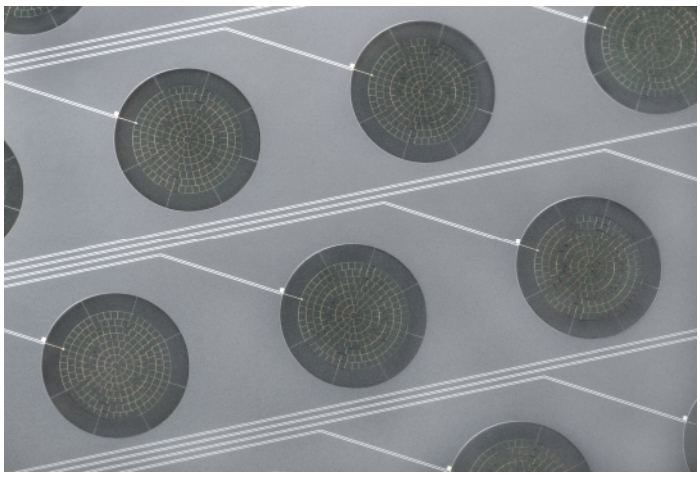

Fig. 14 Close-up of a spiderweb bolometer fabricated with photolithographic techniques. The diameter of each pixel is about 3 mm. The spiderweb intercepts the mm-wave power and heats the bolometer, visible as the small bright dot at the "10 o'clock" position on the outer circle of the spiderweb.

The first development is the insight that biasing the TES at constant voltage, rather than constant current, introduces electro-thermal negative feedback [34]. Analogously to amplifiers, this speeds up the response. stabilizes the operating point, and provides a well-defined responsivity (output signal vs. absorbed power). Stable operating points and well-defined response are both key ingredients for the practical operation of large arrays.

The second development is that bolometers can be fabricated using photolithographic fabrication techniques for silicon ICs and micro-mechanics [35,36]. This allows wafer-scale fabrication of bolometer arrays. Fig. 13 shows a comparison of the 16-bolometer array flown in the balloon experiment MAXIMA [37] and the 300-bolometer array currently under construction in Berkeley for APEX-SZ. The array is made of 6 wedges, each of which fits in a 100 mm wafer. Fig. 14 shows a close-up an individual pixel. The bolometer is suspended in a "spider-web", formed by "beams" of ~5 μm diameter etched from silicon-nitride. The silicon beneath the spider-web is etched away, so the web is suspended only from 8 "spokes" of small thermal conductivity around the periphery. The bolometer itself together with the electrical connections is visible at the edge of the web. A web is used both to reduce the heat capacity and reduce the cross-section for cosmic rays. The meshes of the web are sufficiently small to intercept the mm-wave CMB radiation, which heats up the whole web together with the bolometer at the edge.

The bolometers operate at 500 mK, where the power budget is <1 μW. The heat leak through the bolometer wiring to the next 4 K stage is just acceptable for a few hundred bolometers, but prohibitive for the kilopixel arrays planned for future experiments. Thus, the number of readout lines must be reduced by multiplexing. This can be performed in the time or frequency domain. A time-domain multiplexer using Superconducting Quantum Interference Devices (SQUIDs) has been developed at NIST [38]. At Berkeley we have developed a frequency-domain multiplexer [39,40]. Rather than DC biasing the

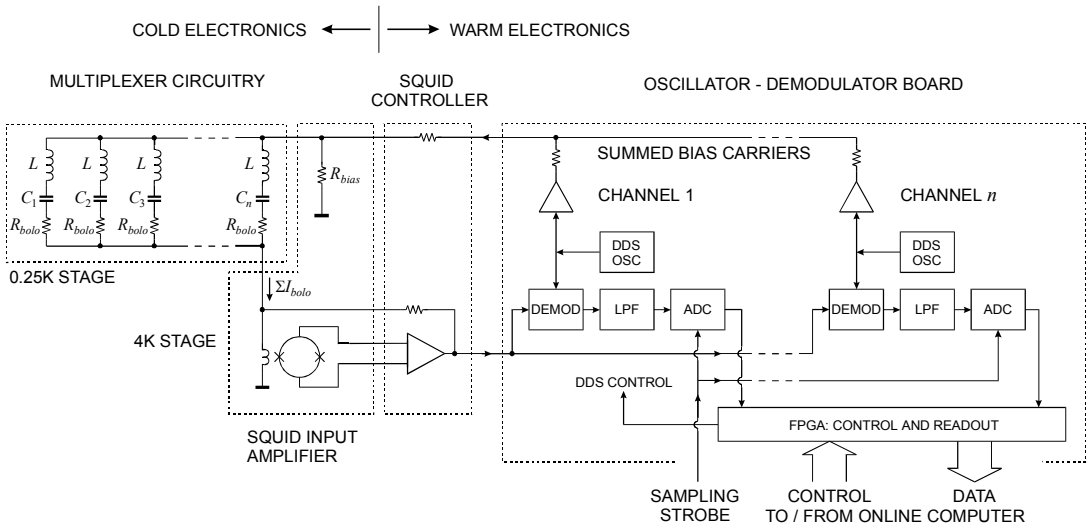

Fig. 15 Block diagram of a cryogenic frequency-domain multiplexer. All bias frequencies are summed and applied as a "comb" through a voltage divider to provide the low source resistance required for constant voltage bias. The bolometers (R_{bolo}) and superconducting *LC* circuits are on a 0.25 K stage; the bolometers are biased to operate at 0.5 K. The SQUID input amplifier and bias resistor R_{bias} are on a 4 K stage. Shunt feedback is applied from the warm SQUID controller to the SQUID input to provide a low input impedance. The readout amplifier output feeds the room-temperature demodulator circuits that extract the individual sensor signals.

TES, we apply a high frequency bias (100 kHz to 1 MHz). When the bolometer absorbs signal power, its resistance changes, thus modulating the current. Amplitude modulation translates the signal into sidebands above and below the bias carrier. As a result, the signal is uniquely placed in frequency space. Each bolometer is biased at a different frequency, so we can sum the currents from all bolometers and read them out through a single readout line. Frequency-selective demodulators separate the signals in the warm electronics. Fig. 15 shows the multiplexer circuit, which also shows that the bias frequencies can be fed through a single line. Tuned circuits associated with each bolometer "steer" each bias frequency to its designated bolometer. The bandwidth of the tuned circuits determines the cross-talk between channels and also limits the noise bandwidth to reduce the contribution of wideband Johnson noise from a given bolometer to the other channels. Multiplexing 16 or 32 bolometers appears to be practical, primarily limited by the SQUID readout amplifier. Only two wires are needed per multiplexed array. The bias frequencies are generated by direct digital synthesis (DDS) ICs, which provide programmable precision frequency control, excellent amplitude stability, and very low sideband noise close to the carrier.

TES frequency-domain multiplexing has also been applied to x-ray detection [41], providing an important ingredient for increasing detection efficiency and rate capability. By their very nature, high-resolution bolometers are small, so they are not very effective detectors. Even with electrothermal feedback their decay times are of order μs, so their rate capability is limited. However, as in silicon strip and pixel arrays at the LHC, dis-

tributing the rate over many detector channels reduces the rate per channel, so large TES arrays can increase both the detection area and rate capability.

This technology of superconducting bolometer arrays is still in its infancy. Although the basic ingredients have been demonstrated, experience with full systems in experiment environments is essential. Arrays with several hundred bolometers are now under construction and designs for the next generation with thousands of pixels are well underway.

7. Conclusion

The past two decades have brought about a major transition in imaging detectors. Highly segmented semiconductor detectors with "massively parallel" IC readout systems have become commonplace in high energy physics and the technology is now moving into other fields. These developments were not brought about by technology alone; detector builders had to rethink traditional design approaches and work in a different style. Although key developments were stimulated and guided by specific experimental goals, dedicated detector R&D programs were essential.

A similar development is now taking place in the field of cryogenic detectors. Large arrays of superconducting detectors are now practical through the application of silicon processing techniques in the fabrication of wafer-scale bolometer arrays. Next generation experiments are now under construction to map the Cosmic Microwave Background with unprecedented precision. Cosmology theory is providing testable predictions and the experimental tools to test them have been developed. Recent cosmological experiments have yielded the following composition of the universe [42]:

Atomic Matter	4%
Dark Matter	23%
Dark Energy	73%

Of the two largest contributions, dark matter and dark energy, all one can say is that we see their effect, but don't know what they are. We do know that they are not the stuff we or the planets are made of. Put differently, all of the physics and chemistry of the past ~400 years has been directed at understanding less then 5% of the universe! Indeed, we may find the "new physics" by looking 13 billion years into the past. However, whatever the nature of these unknown constituents, we can be confident that new imaging detectors will play a major role in solving their mysteries.

Acknowledgements

Many thanks to the organizers for putting together a stimulating workshop in a beautiful city. I also wish to thank the many friends and colleagues whose figures I used.

This work was supported by the Director, Office of Science, Office of High Energy and Nuclear Physics, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098

References

1. J.R. Janesick, Scientific Charge-Coupled Devices, SPIE Press Monograph Vol. PM83, ISBN: 0819436984
2. E. Heijne et al., Nucl. Instr. and Meth. 178 (1980) 331
3. B.D. Hyams et al., Nucl. Instr. and Meth. **205** (1983) 99
4. D.R. Nygren, The Time Projection Chamber: A New 4 PI Detector for Charged Particles, SLAC PEP-0144, in Proceedings, 1974 PEP Summer Study, Berkeley, 1975, 58-78 and D. Fancher, et al., Nucl. Instr. and Meth. **161** (1979) 383
5. E. Gatti et al., IEEE Trans. Nucl. Sci. **NS-32** (1985) 1204
6. Bill Reay and Helmuth Spieler, Working Group Report on Position Sensitive Devices, *Proceedings of the Workshop on New Solid State Devices for High Energy Physics, Berkeley, October 28-31, 1985,* LBL-22778
7. R.C. Jared et al., IEEE Trans. Nucl. Sci. NS-29 (1982) 57-61
8. J.T. Walker et al., Nucl. Instr. and Meth. **A 226** (1984) 200
9. P. Seller, P. Allport, M. Tyndel, IEEE Trans. Nucl. Sci. **NS-35** (1988) 176-180
10. W. Buttler et al. Nucl. Instr. and Meth. **A273** (1988) 778-783
11. S.A. Kleinfelder et al., IEEE Trans. Nucl. Sci. **NS-35** (1988) 171
12. S. Holland and H. Spieler, IEEE Trans. Nucl. Sci. **NS-37** (1990) 463-468
13. S. Holland, IEEE Trans. Nucl. Sci. **NS-39** (1992) 809-813
14. W. Snoeys et al., IEEE Trans. Nucl. Sci. **NS-39** (1992) 1263-1269
15. David Nygren and Helmuth Spieler, Semiconductor Detector Size and Radiation Hardness, in *Report of the Task Force on Radiation Effects at the SSC*, SSC-SR-1035, SSC Central Design Group, Lawrence Berkeley Laboratory, 1988
16. V.I. Kuznetsov et al., Fiz. Tekh. Poluprovodn. **9** (1975) 749-752, Sov. Phys. Semicond. **9/4** (1975) 491-492
17. H. Spieler, Front-End Electronics and Signal Processing, in Proceedings of the First ICFA School at the ICFA Instrumentation Center in Morelia, AIP vol. 674, pp. 76 – 100, eds. L. Villaseñor, V. Villanueva, LBNL-52914, also at http://www-physics.LBL.gov/~spieler
18. Luciano Musa, at http://alice.web.cern.ch/Alice/transparencies/2001/LHC2001/Musa.pdf
19. C. Gemme et al., Nucl. Instr. and Meth. **A 501** (2003) 87-92
20. J. John, presentation at this Workshop
21. J. Salmi, presentation at this Workshop
22. Yu. Gornushkin et al., Nucl. Instr. and Meth. **A 478** (2002) 311-315
23. P. Collins, presentation at this Workshop
24. S.E. Holland et al., IEEE Trans. Electron. Dev. **ED-50** (3) (2003) 225-238 and LBNL-49992
25. D. E. Groom, SPIE **4008** (2000) 634-645 and LBNL-45277
26. C. J. Bebek et al., SPIE **4669** (2002) 161-171 and LBNL-49316
27. S. E. Holland, N. W. Wang, and W. W. Moses, IEEE Trans. Nucl. Sci. **44** (3), 443-447 (June 1997)
28. G. Aldering et al., Supernova / Acceleration Probe: A Satellite Experiment to Study the Nature of the Dark Energy, submitted to Publications of the Astronomical Society of the Pacific (PASP), http://arXiv.org/abs/astro-ph/0405232

29. Eric V. Linder, Dark Energy, Expansion History of the Universe, and SNAP, Proceedings 3rd Topical Workshop on Particle Physics and Cosmology: Neutrinos, Branes and Cosmology, San Juan, Puerto Rico, 19-24 Aug 2002, AIP Conf. Proc. 655 (2003) 193-207, http://arXiv.org/abs/astro-ph/0302038
30. J-M. Bussat et al., A Next Generation, High Speed Detector for Synchrotron Radiation Research, to be published in IEEE Trans. Nucl. Sci.
31. I. Kipnis, H. Spieler and T. Collins, IEEE Trans. Nucl. Sci. **NS-41/4** (1994) 1095-1103 (CAFE-M predecessor with very similar characteristics)
32. P. Denes et al., New 1D and 2D Detectors for Synchrotron Radiation Research, presented at Eighth International Conference on Synchrotron Radiation Research, San Francisco, Aug. 25-29, 2003, http://www.sri2003.lbl.gov/html/presentations.html
33. A.T. Lee et al., Voltage-Biased TES Bolometers for the Far-Infrared to Millimeter Wavelength Range, in Proceedings Far-IR, Sub-mm & mm Detector Technology Workshop, Wolf J., Farhoomand J. and McCreight C.R. (eds.), NASA/CP-211408, 2002 and at http://www.sofia.usra.edu/det_workshop/papers/manuscript_session5.html
34. A.T. Lee et al., Appl. Phys. Lett. 69 (1996) 1801
35. J.M. Gildemeister, Adrian T. Lee, P.L. Richards, Appl. Phys. Lett. **74** (1999) 868
36. J.M. Gildemeister, Adrian T. Lee, P.L. Richards, Appl. Phys. Lett., **77** (2000) 4040
37. B. Rabii et al, MAXIMA: A Balloon-borne Cosmic Microwave Background Anisotropy Experiment, accepted by Astrophysical Journal, http://arXiv.org/abs/astro-ph/0309414
38. J.A. Chervenak et al., Appl. Phys. Lett. **74** (1999) 4043-4045
39. H. Spieler, Frequency Domain Multiplexing for Large-Scale Bolometer Arrays, in Proceedings Far-IR, Sub-mm & mm Detector Technology Workshop, Wolf J., Farhoomand J. and McCreight C.R. (eds.), NASA/CP-211408, 2002, LBNL-49993 and at http://www.sofia.usra.edu/det_workshop/papers/manuscript_session5.html
40. T.M. Lanting et al., IEEE Trans. Appl. Superconductivity **13/2** (2003) 626-629
41. J. N. Ullom et al., IEEE Trans. Appl. Superconductivity **13/2** (2003) 643-648
42. C.L. Bennet et al., Astrophys. J. Suppl. **148** (2003) 1