For linear e^+e^- colliders, microstrip vertex detectors were never ideally suited, due to the high background associated with the single-pass collider operation. (As Witold Kozanecki puts it, "Backgrounds at SLC are similar to those at LEP, *during injection.*") This will also be true at small radii for the future high-energy linear collider. However, as at LHC, there is a good chance that silicon microstrip detectors may be the chosen technology for the outer tracking system at this machine.

5 Pixel-Based Detectors

5.1 Introduction

There are exceptions to every classification scheme. I was delighted to read a paper [41] contributed to the recent European Conference on Semiconductor Detectors which neatly bridged the gap between the one-dimensional microstrip detectors and the two-dimensional pixel-based detectors. How could this be? The authors were interested in detecting hard X-rays for which the attenuation length in silicon is rather long. To achieve a reasonable efficiency, they had the excellent idea to turn a microstrip detector ***edge-on*** to the X-ray direction, so that the strip length (several mm) became the effective detector thickness. In this way, they were able to achieve 80% efficiency for detecting 20 keV X-rays. By sweeping the detector slowly across the image, they were able to build up full two-dimensional images of excellent quality.

More usually, the term "pixel detector" is taken to mean a device equipped with a one- or two-dimensional array of pixels (picture elements). The two-dimensional variety, given the sensitivity of silicon for visible light, is the basis for the huge commercial market for camcorders and other electronic image-capture products. This marks the most important contrast with respect to the previously discussed microstrip detectors; while the strips can provide very precise position information, the fact that they are inherently one-dimensional precludes any application in which the desired output is some form of picture. Hence, pixel devices are of much greater interdisciplinary importance (both in terms of scientific sensors and in commercial terms) than microstrip detectors.

However, for tracking devices such as vertex detectors, how important is it to have this picture-taking capability? Figure 31 demonstrates that a few planes of pixel-based detectors give unambiguous track-finding capability, whereas the same number of planes of ***double-sided*** microstrip detectors do not produce an immediately recognizable pattern of tracks. There are in fact $N!$ patterns possible in the case of a jet of N particles. This is not too bad for the three-particle case shown (six-fold ambiguous), but for a high-energy jet of ten tracks, there are 3.6×10^6 possibilities!

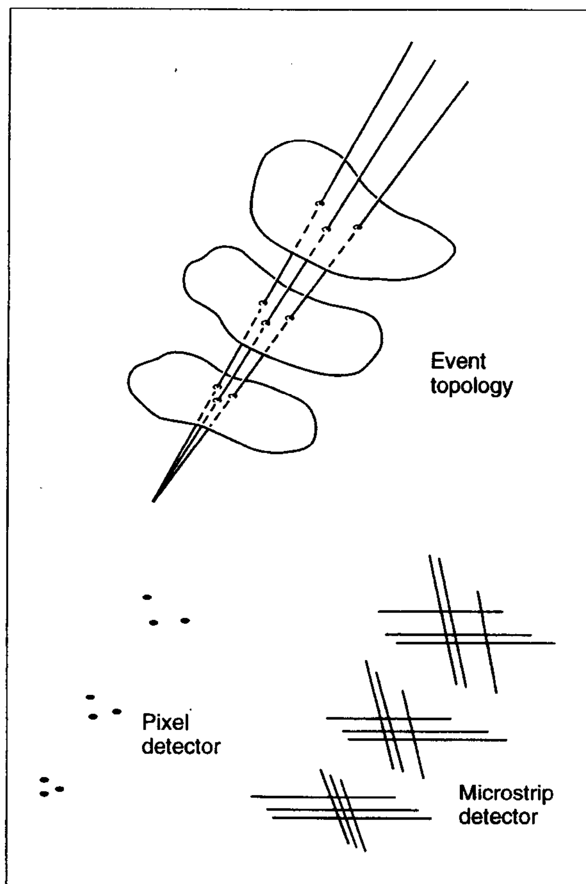


Fig. 31. Upper sketch: a few tracks traversing an unspecified set of three detector plates. Lower sketch: resulting information in case of one- and two-dimensional detector types.

What this means in practice is that such detectors would need to combine information from different planes having strips oriented differently (not necessarily a practical option in a collider detector) or (more usually) rely on the external detectors to perform the pattern recognition. Since there can be a lot going on between the IP and the outer tracking detectors (decays, γ-conversions, secondary interactions), a pixel-based vertex detector capable of stand-alone pattern recognition is manifestly a much more powerful tool for physics.

A second advantage is that of granularity. A single typical microstrip (e.g., of the DELPHI detector) covers $70\ \text{mm} \times 50\ \mu\text{m}$. This area (in a CCD detector) would be covered by 9000 pixels. These four orders of magnitude in granularity make for a huge advantage in tolerable hit density before the problems of cluster-merging start to make life difficult for the track reconstruction algorithm. One can for this reason position a pixel-based detector much closer to the IP, with obvious advantages for impact parameter precision (shorter extrapolation, just as a short focal-length lens makes for a more powerful microscope). Furthermore, there are physics environments where the density of background hits close to the IP is so high that a microstrip detector would be obliged to back away in order to reduce the occupancy to a tolerable level, whereas a pixel-based detector would be perfectly comfortable.

The third advantage is in terms of radiation hardness. We shall address this complex issue in Sec. 6, but in many cases, the limiting parameter is growth of leakage current, with associated shot noise which eventually can overwhelm the signal. In a microstrip detector, the signal on one strip has to be found against the noise background associated with the entire strip. If the "strip" length is reduced by a factor 10^4 (above example), the noise associated with the leakage current is correspondingly reduced. This can make the difference between a detector lifetime of one month and 2,000 years.

There are two other partly connected advantages. Most forms of pixel-based detectors have extremely low capacitance nodes for the charge collection, and hence need much smaller charge signals for satisfactory signal-to-noise. Excellent MIP detection efficiency is achievable with active layers 20 times thinner than microstrips. As we have seen, this has major advantages for tracking precision, both

for normal-incidence particles (minimizing the problem of δ-electrons) and for angled tracks (minimizing the effect of energy-loss fluctuations). Originally, it was customary, in using these devices with thin active layers, to leave them mechanically thick (say 300 μm), but more recently, techniques have been developed for mechanical lapping, chemical etching, and handling so that thinner devices can now be built into HEP detectors, with a further reduction in multiple Coulomb scattering.

Against these advantages, pixel-based detectors have disadvantages which make them impractical or uncompetitive in some situations. In order to appreciate these, however, we need to consider the two important classes of such detectors, for which the characteristics are extremely different and indeed complementary. These classes are the charge-coupled devices and the active pixel sensors.

5.2 Charge-Coupled Devices (CCD's)

An imaging CCD (Fig. 32) consists firstly of a square matrix of potential wells, so that signal charge generated below the silicon surface can be accumulated, building up an image. Secondly, by manipulating clock voltages in the *parallel register* (the I ϕ gates), charge can be transferred in parallel from one row to the next. Charge in the bottom row of the matrix is transferred into the adjacent *linear register*. The stored signals are then transferred one at a time (by manipulating the R ϕ gates) onto the output node, which is connected to the input of an on-chip charge-sensing amplifier. Also on chip is a reset FET to restore the output node to its nominal value, usually after reading the signal from each pixel. Thus, the CCD image is converted from a two-dimensional charge pattern to a serial train of pulses, well-suited to display on a video monitor. The CCD was invented in 1970 (Ref. [42]). Devices of this pioneering design (so-called surface channel CCD's, because the signal charge is stored at the silicon/silicon dioxide interface) are still used in video cameras. However, within two years, the invention of the more sophisticated buried-channel architecture was published [43]. Here, the signal charge is stored in the bulk of the silicon approximately 1 μm below the surface, suitably remote from the interface states that (as we shall see) can trap signal charge. For the small signals usually

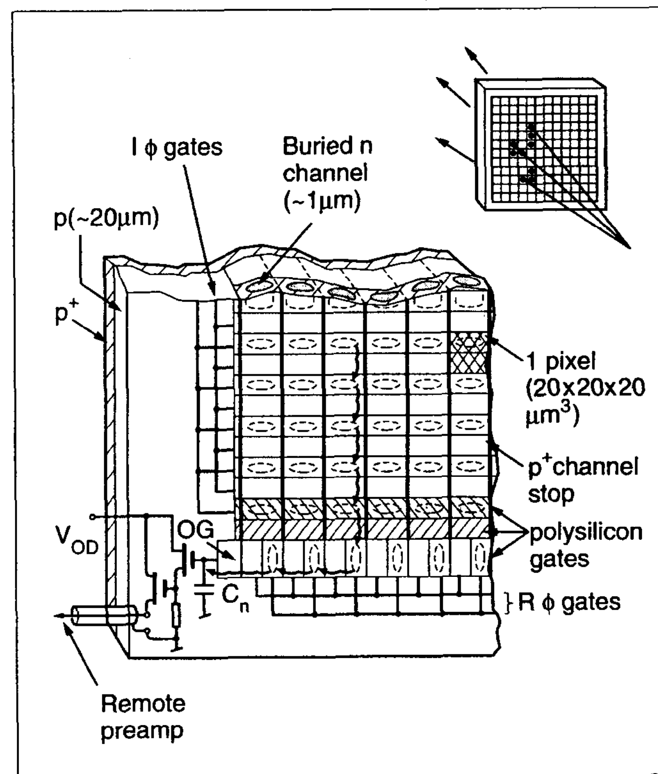


Fig. 32. Upper right: sketch of charge storage in a CCD detector traversed by a number of ionizing particles. Lower left: corner region of CCD showing the principal structural features.

sought in scientific applications, the buried-channel design is much more suitable, so we shall concentrate entirely on this.

Before diving into the details of scientific CCD's, let us take a brief look at the technology push being provided by industry. The largest CCD market is for camcorder sensors. The immediate aim in this market is to increase sensitivity so as to achieve good performance under typical indoor home lighting conditions. The next goal is CCD's for HDTV broadcast cameras (1920 x 1036 pixels, two readout channels, each running at 37 MHz) followed (in about 1998) by the HDTV camcorder. In addition, there is a big push for a high-quality electronic still photography camera, and eventually, an electronic motion picture camera. CCD design rules in the commercial sector are 0.5 μm (and reducing), and wafer sizes are 6" (and increasing). Both of these are currently beyond the reach of the manufacturer of scientific CCD's. The commercial devices use interline transfer and are typically only about 2 μm thick (active layer). This is excellent for sharp color images but makes them inapplicable for most radiation detector applications. The major commercial manufacturers are too busy chasing the frontiers associated with the mass market to be interested in the specialized needs of the scientific CCD users. Fortunately, there are several extremely high-quality manufacturers who serve this particular niche in the market. The possibility of using CCD's as high-precision detectors of MIP's was first evaluated theoretically about 15 years ago [44].

5.2.1 Structure and Basic Operation

We shall concentrate on the frame-transfer MOS CCD family since this is the most commonly used for scientific applications and the only one used to date for vertex detectors in HEP experiments.

Let us examine in some detail, with the aid of the general discussion of Sec. 3, how such a detector can be built. For more detailed information, there are some excellent books on CCD's [45, 46] as well as CCD conference proceedings and hundreds of published papers.

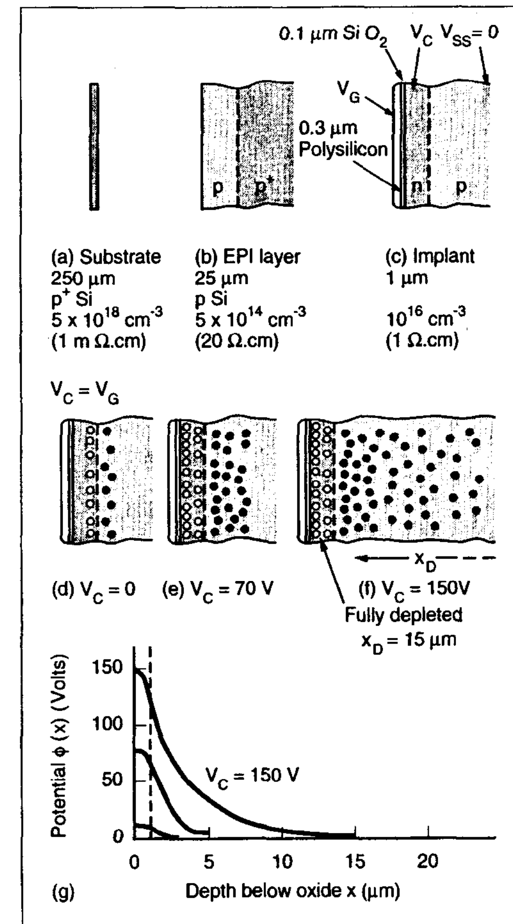


Fig. 33. (a)–(c) The successive stages in making a CCD-like structure (shown with increasing magnification). (d)–(f) The depletion process which would apply if V_C and V_G were increased together. (g) The corresponding potential distributions as a function of depth in the silicon.

Let us first consider the steps in making a device which would have some (but not yet all) of the features of a CCD. Starting with a low-resistivity, suitably inert substrate [Figs. 33(a)–(c)], we proceed to grow an epitaxial layer of higher resistivity silicon with a thickness adequate to contain all the necessary structures and associated field penetration. We next make a *pn* junction by the introduction of a shallow (approximately 1 μm) implant of *n*-type dopant. The surface is oxidized to make an insulating layer, and on top of this is deposited a thin conducting layer. The simplest would be aluminum, but for light detection, a high degree of transparency is important, and about 0.3 μm low resistivity "polysilicon" (amorphous silicon) would commonly be used. By analogy with FET's, the conducting surface layer is termed a "gate."

Let us now put some bias voltage onto the structure, as shown in Figs. 32(d)–(f). Grounding the substrate ($V_{ss} = 0$), we apply V_c to the *n* channel and V_G to the gate. Initially, assume $V_c = V_G$. Even with $V_c = 0$, as we learned in our discussion of the *pn* junction, there will be a thin depletion layer around the interface between the two types of silicon. By increasing V_c, we are able to deplete more of the material as the junction becomes more and more strongly reverse biased. With the parameters chosen in this example, a high voltage would be needed to achieve complete depletion of the *n* channel, at which point we should have depleted about 20 μm of the *p*-type substrate. The potential distributions for increasing values of V_c are shown in Fig. 32(g). For $V_c = 150$ V, such a device when traversed by particles would transport the generated electrons to the surface (silicon/silicon dioxide interface) and dump the holes into the undepleted substrate.

Now [looking at Figs. 34(a) and 34(b)], consider what happens if V_c is increased from zero while V_G is held at zero volts. Here, the situation is entirely different; the large capacitance between the *n* channel and the gate provides a further mechanism for depletion of the channel. The depletion around the *pn* junction proceeds as before, but the voltage across the oxide induces an increasing positive space-charge, starting from the silicon/silicon dioxide surface and growing into the body of the *n* channel. At a very low value of V_c (about eight volts), these depletion regions meet, causing the phenomenon known as *pinch-off.* The corresponding value of V_c

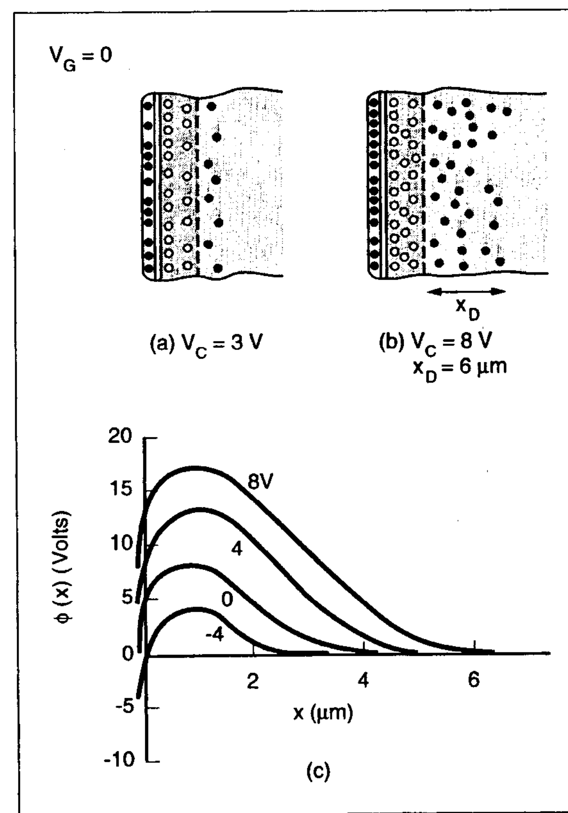


Fig. 34. (a) and (b) The depletion process in normally biased CCD operation with V_G negative with respect to V_c. (c) The corresponding potential distributions after channel pinch-off for various values of V_G.

is called the pinch-off voltage, and when it is reached, further increases of V_c (which can be controlled, say, by an edge connection) have no influence on the potential over the area of the detector. The depletion depth in the *p*-type material is only about 6 μm in this case. What is particularly interesting is the potential distribution in the silicon. This is shown in Fig. 34(c); look initially at the curve for $V_G = 0$. The quadratic form in both types of silicon is, of course, preserved (this is a consequence of Poisson's equation and uniform doping), but there is now a maximum in the electric potential just below the depth of the *pn* junction. This acts as a potential energy minimum for electrons, so (in contrast to the case $V_G = V_c$) the electrons liberated by the passage of a particle would accumulate approximately 1 μm below the silicon surface in the so-called *buried channel* of the device. This is a vital ingredient in the design of CCD's for our application. Tiny charges ($< 10\ e^-$) can be safely stored and transported as long as they are held in the bulk of the silicon. Once they are allowed to make contact with the surface, they encounter numerous traps which cause serious loss of charge. Surface-channel CCD's, while quite commonly used, should be avoided for work with very low signal levels.

Notice that the situation depicted in Fig. 34(c) represents a non-equilibrium condition. Thermally generated electrons would accumulate in the potential energy minimum and drive more and more of the *n* channel out of depletion. CCD operation relies on some procedure for keeping the channel swept clean of electrons at an adequate rate.

Assuming that we avoid this accumulation of electrons, the effect of now varying the gate voltage V_G is to a first approximation simply to vary the depth (in volts) of the potential well, but hardly at all to change its depth (in microns) below the silicon. There is, in fact, a slow variation in the depletion depth with V_G, as can be seen from the figure. The quantitative calculation follows easily from what we have done in Sec. 3; see, for example, Ref. [46] for the details.

The device we have created has all the depth characteristics of an imaging CCD, but it still lacks two important features before it will have the necessary pixel structure over the surface. These are illustrated in Fig. 35. Firstly, at the required pixel granularity (say, 20 μm), p^+ implants are introduced of approximately 1 μm width and

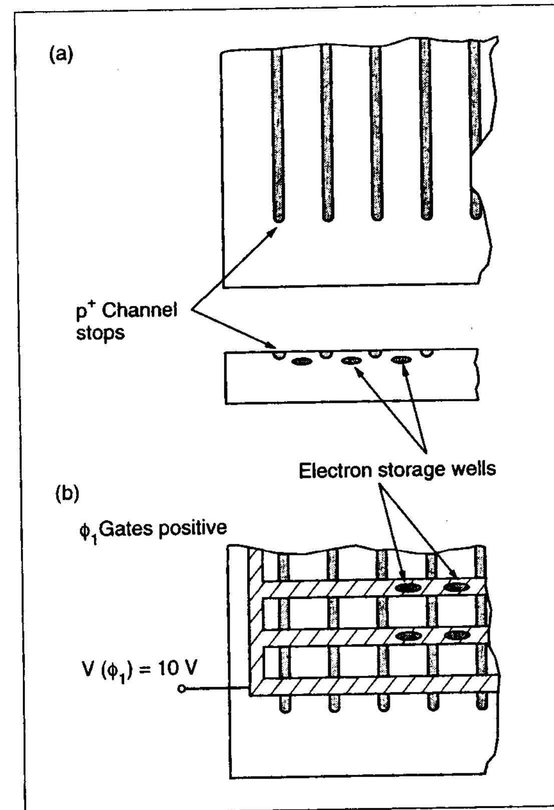


Fig. 35. Establishing the potential well structure: (a) Channel stops create potential barriers running vertically on the device. (b) Gates create horizontal potential barriers. The combined result is a matrix of localized wells, each of which constitutes a pixel.

1 μm depth. These become partly depleted as part of the overall biasing of the CCD, and so provide strips of negative space charge which effectively repel electrons. Thus, the electrons in the buried channel will now be confined to separate storage wells which run from top to bottom of the detector, in the view shown in Fig. 35(a). The typical doping level of the channel stops is $N_a = 10^{18}$ cm^{-3}.

Secondly, the charges are confined in the vertical direction by making a polysilicon gate structure which is not uniform across the surface but which consists of a series of horizontal bars. By biasing these positively [see Fig. 34(c) and Fig. 35(b)], we can achieve potential wells under each of the intersections between these gate electrodes and the regions midway between the channel stops. We now have a matrix of discrete potential wells which may exceed 10^6 in number on a typical CCD (800 channel stops x 2000 gate electrodes).

But still, we do not have a working CCD, since those potential wells are immobile. We can accumulate charge images but cannot read them out. To do this, we make a more complicated gate structure (Fig. 36). We arrange these gates in triplets (ϕ_1, ϕ_2, ϕ_3) in this so-called three-phase CCD structure. The static situation is for one phase (say, ϕ_1) to be high, so that the electrons are stored under this phase. Then by manipulating the voltages between ϕ_1 and ϕ_2 as shown in the figure, the electrons are moved to ϕ_2. Keeping ϕ_3 low throughout this operation ensures that the charges between adjacent pixels cannot be smeared together. The total physical width of $\phi_1 + \phi_2 + \phi_3$ electrodes together constitutes one pixel, e.g., $3 \times 7\ \mu\text{m} = 21\ \mu\text{m}$.

Now we have developed the capability to move all the stored charges down the device (for example) by one pixel at a time. Apart from three-phase CCD's, there exist other varieties (four-phase, two-phase, virtual phase, etc.).

At the bottom of the area array called the "imaging" or "I array" is a linear CCD, the output register or R register into which the charges stored in the bottom row of the I array can be shifted. Once in this register, charges in that row can be transferred sideways so that the charge contained in each pixel is sensed in turn by the on-chip circuit.

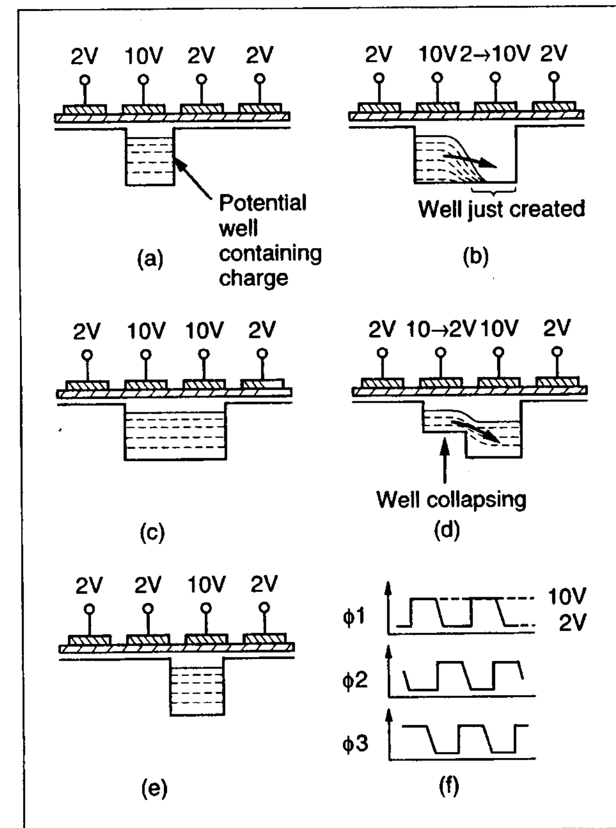


Fig. 36. After Ref. [46]. (a)–(e) Movement of potential well and associated charge packet by clocking of gate electrode voltages. (f) Clocking waveforms for a three-phase CCD.

Referring back to Fig. 32, which shows a two-phase CCD, note that each pixel (shaded area) covers the height of two I gates and is bounded by a pair of channel stops.

The CCD structure shown in this figure is sensitive to light or to particles over the full active area. It should be noted that this is not true of all imaging CCD's. Some, for example, have more complex channel stops, *pnp* structures which can be used for anti-blooming or for fast-clearing the CCD's. Such devices have dead bands between each pixel, a feature which makes them unacceptable for most applications as particle detectors.

In the spirit of Fig. 25 (simplified cross section of a generic microstrip detector), Fig. 37 shows the corresponding case of the MOS CCD. Note the buried channel, a region within the n^+ implant, not crossed by field lines, and the crossover in the electric field at that depth (lower plot). Note that the buried channel depth varies only slightly as the gate voltage is varied. Note also the intrinsic p/p^+ potential barrier created by the narrow depletion region at the back surface interface of the epitaxial silicon. We can correlate this with Fig. 24, which shows how the charge generated by a MIP along its track falls into a number of classes in such a structure. There is a region of typically 5 μm below the surface for which the charge is within the depletion depth and is fully collected into the "central" pixel, i.e., the one traversed by the particle. Next, the charge from the 15 μm of undepleted epitaxial silicon (which generally has a long diffusion length, maybe hundreds of microns) diffuses isotropically. About half of it diffuses into the depletion region and is caught in the central pixel *or* in neighboring ones; the other half gets there after being reflected off the p/p^+ potential barrier.

As has already been noted, the CCD potential wells represent a non-equilibrium condition. Thermal generation of electron-hole pairs in the material provides a source of electrons which accumulate. For TV imaging, these constitute a minor background, but for astronomy, the long integration times and low signal levels necessitate cooling, typically to liquid nitrogen temperatures. For particle detection, the requirements are less stringent and operating temperatures around 200 K may be

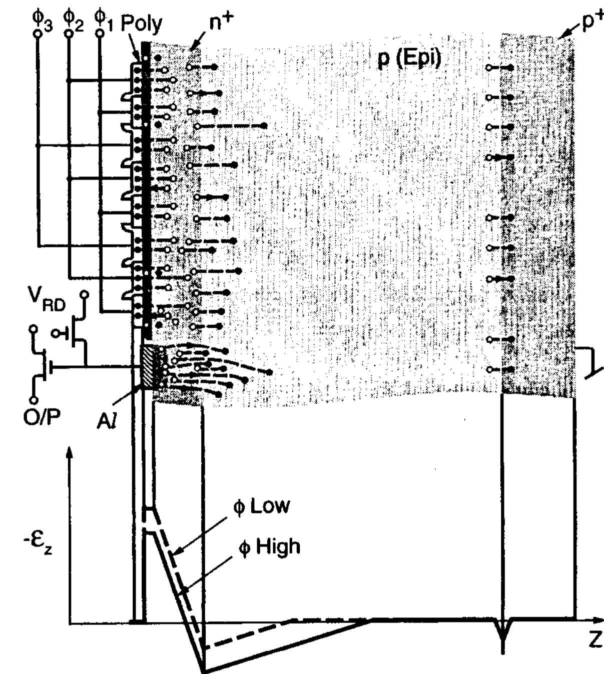


Fig. 37. Sketch of cross section of a generic three-phase MOS CCD. As in Fig. 25, exposed fixed charges are shown by open circles (positive) and filling circles (negative). Also shown is the electric field distribution in regions of high- and low-imaging gate voltage.

entirely adequate, but this depends strongly on the timing of the clearing and readout of the detectors.

It is worth noting that the scientific CCD has in recent years been extended by the development of the *pn* CCD [47]. This is (like many "innovations") far from new, having been developed, then forgotten, soon after the original CCD invention. At that time, it was known as the junction CCD [48]. As shown in Fig. 38, it is very like the MOS CCD except that negatively (reverse) biased p^+ implants are substituted for the MOS gate structure. *pn* CCD's are usually manufactured with a view to high X-ray efficiency, and hence are fabricated on high-resistivity silicon which is fully depleted, as in the microstrip detector. This case is shown in Fig. 38. For X-ray detection, there are recent papers reviewing the relative capabilities of both the MOS [49] and *pn* [50] CCD's. There is a considerable overlap as well as a degree of complementarity in their application areas [51]. For particle physics applications, MOS devices have been exclusively used to date, largely because of their ready availability at competitive prices from a number of manufacturers.

E. Fossum has written an excellent recent review of image sensor technologies (mostly CCD's) and of companies manufacturing these devices for scientific customers [52].

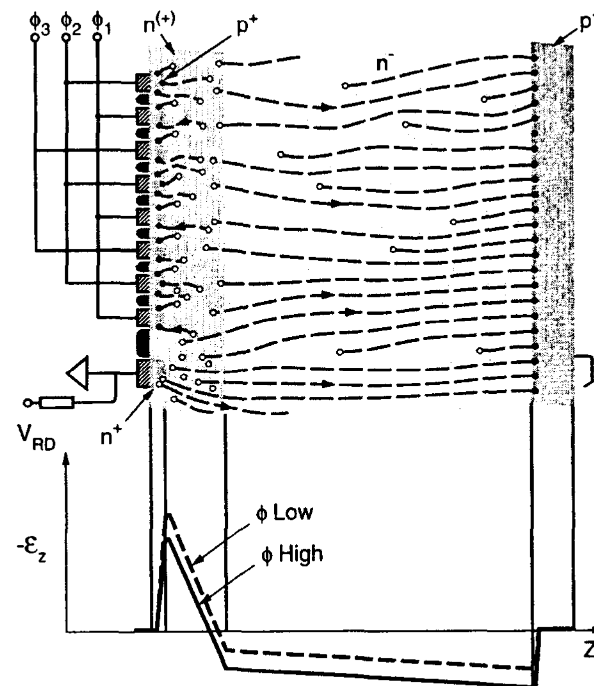


Fig. 38. Sketch of cross section of a generic *pn* CCD manufactured on high-resistivity silicon and depleted over the full thickness.

5.2.2 CCD Charge Transfer and Readout: Detailed Issues

5.2.2.1 Charge Transfer Process

As we have seen, the transfer of signal charge from pixel to pixel is accomplished by changing the voltage levels on the gate electrodes. Since the magnitude of the MIP signals is so small (approximately 2,000 e^- compared with about 10^5 e^- well capacity), one might imagine that very small drive pulse modulations would suffice to achieve good CTE. On the contrary, 5–10 V pulses are needed. Why is this? Firstly, in producing any IC, fixed positive charge is trapped at the silicon/silicon dioxide interface. This is dependent on the processing details, so one would never find perfect equality between (say) the three-phases of a register, which are obviously deposited in separate operations. The uncontrollable differences amount to

several tenths of a volt and result in effective charge storage for MIP signals even in the absence of any applied drive voltages. The idea of getting rid of early hits by letting the charge diffuse along the columns, as enunciated in my group's first paper [44] on the possible application of CCD's for MIP detection, simply does not work, as we found some years later. Small signals in CCD's cannot be eliminated like this; the electrons in the charge packet are, in fact, extremely cohesive.

The most significant factor that determines the minimal drive pulse voltages required for good CTE is the unavoidable presence of shallow traps which tend to pick up signal electrons at every stage of their long journey to the output node. Particularly in a sparse data situation such as one has in a particle detection system, such traps are dangerous. They may emit electrons with a long time constant, then sit empty until the arrival of a signal packet, at which point they capture electrons almost instantaneously. The signal packet moves on, with the trapped electrons being released only much later. As we shall discuss in Sec. 6, radiation damage can cause serious growth in the density of these bulk traps. The operating temperature is a very important parameter in minimizing this problem, since it profoundly affects the trap emission time constants. The problem of bulk traps affecting CTE in CCD's was first treated in a famous paper by Mohsen and Tompsett [53]. The topic has been revisited many times since; for a recent paper dealing specifically with CCD's operated at low temperature, see Ref. [54].

As well as the problem of traps of atomic dimensions, CCD's are also sensitive to more macroscopic potential wells (sometimes referred to as potential pockets) that can swallow part or all of the charge packet within one pixel. There are innumerable processing imperfections liable to cause such potential wells (minor variations in gate oxide thickness, tiny blemishes in gate polysilicon, minor crystal imperfections such as slip lines, and so on). Such manufacturing problems can be very difficult to diagnose; suffice it to say that less than 10% of large area devices made by a top-of-the-range CCD manufacturer are likely to suffer from such effects in more than 1% of the columns. As such, this is not a serious yield issue.

Both as regards atomic-scale bulk traps and as regards potential pockets, high-drive voltages can be extremely effective in releasing electrons from all but the deepest lying bulk traps, by virtue of the Poole-Frenkel effect [55] (lowering of a potential barrier by a potential gradient). Interestingly, the relevant strong electric fields arise not from the horizontal fringing fields, but from the fields developed along the vertical doping profile of the buried channel implant [54]. The device physics may be somewhat subtle, but the experimental observations are unambiguous: for good CTE, drive pulses in excess of 5 V and typically 10 V may be needed. What are the consequences of this?

As regards the parallel register, the capacitance to ground of each of the gates is pretty large, the polysilicon gate electrodes are somewhat resistive, so one may be limited to clock rates of around 100 kHz in order to achieve adequate voltage excursions at the center of a large CCD. The large currents induced in the CCD structure by the voltage excursions in the parallel register (which covers nearly all of the area of the device) generate massive feedthrough signals on the CCD output circuit. Neither the limited clocking frequency nor the feedthrough signals can normally cause any problems, since each parallel transfer is followed by typically 400 serial transfers as that row is read out, so the overall readout time is not seriously affected by the parallel transfer time constants.

For the serial register, equally large drive pulses are required. However, the associated capacitance is much smaller, and there is no problem to clock the serial register with good CTE in excess of 20 MHz. The theoretically maximum clocking rate is a very rapid function of the pixel size (length) [56], 60 MHz for 20 μm but only 4 MHz for 50 μm pixels. Experimentally, 20 μm pixels are easily clocked at 30 MHz.

In a vertex detector application, material in the active detector volume is to be minimized. In an optimized CCD design, the on-chip power dissipation associated with the drive pulses and readout amplifier are similar and extremely modest. A detector of some hundreds of mega-pixels can be cooled by a gentle flow of nitrogen gas. The cooling problems would become approximately a 100 times greater if the drive and readout electronics were contained within the low temperature enclosure. In practice, one always locates these outside the cryostat (using low mass striplines of approximately 30 cm length for the interconnections). Thus, the local electronics

can be run at room temperature, water cooled, and positioned in the small polar-angle region, beyond the coverage of the tracking detectors. Recent developments in electronics design have offered a remarkable opportunity for shrinking *all* the drive electronics into this small space where tracking is not required, at the heart of a collider detector. This allows the cleanest possible drive pulse generation, a major improvement on earlier systems for which these pulses had to be generated in modules on the periphery of the global detector, some tens of meters distant, and carried in on approximately 1000 fine coax cables.

As already noted, because of their low duty cycle, the parallel register drive pulses make only a minor contribution to the detector readout time. This readout time is effectively determined by the duration of the serial register clocks and the analogue signal-sensing electronics. In operating a CCD register, phases are always clocked in opposition, one coming down and another going up as the electron packet is passed on (see Fig. 36). The cleanest arrangement is the two-phase register, where an implant beneath each gate biases the charge packet to be always stored in the "downstream" half of the gate area. Balanced drive pulses to the two gates provide minimal disturbances to the CCD output circuit. But it is a very delicate business. The 10 V pulses are swinging around during the transfer of a MIP signal which (if one is lucky) may give a 1 mV step on the output node. The positive and negative edges of the drive pulses are unlikely to be balanced to better than a few percent.

Even if they were perfect, geometry layout differences on and off CCD (more importantly, the latter) can cause major feedthrough and ringing of the analogue signals by ten to 100 times the 1 mV level. For slow readout systems, one can wait for this to settle down. A major challenge in reading CCD's at 20 MHz or above with low noise (< 100 e$^-$ RMS) is to achieve excellent isolation between the drive and analogue signals in compact systems. Use of miniature coax for the two critical drive lines between the local electronics and the detector is certainly helpful, but there are numerous possible feedthrough paths, all of which need to be extremely carefully controlled.

5.2.2.2 Charge Detection

The most commonly adopted CCD on-chip charge detection circuit is of the general form shown in Fig. 39. It consists of firstly an output diode, the very small n^+ implant seen in Fig. 37, linked to the serial register via the output gate (OG of Fig. 32). Thus, the CCD output node has its potential reset periodically to the reference voltage V_R via the reset transistor, which restores it to an appropriate voltage for collection of signal charge Q_s transferred by clocking from the buried channel of the serial register. This charge transfer causes the node potential to change by $\Delta V = Q_s / C_n$, where the node capacitance C_n is given by

$$C_n = C_d + C_g(1-G).$$

C_d is the node-substrate capacitance, and C_g is the gate-source capacitance of the transistor. G is the voltage gain of the source follower. For optimum signal to noise, these two capacitive components should be approximately matched. See Ref. [57] for a detailed discussion of the optimal transistor design parameters. This implies a small-sized transistor, which consequently has a relatively high impedance at its output source. For optimum noise performance, it is advantageous to use a depletion mode or buried channel MOSFET. This important discovery, made ten years ago [58], is understood in that the drain current in a surface channel FET experiences noise due to the continual random filling and emptying of interface states, which consequently modulate the channel characteristics. For a modern CCD [59], the advantage of a buried channel first-stage MOSFET is indicated in Fig. 40; the $1/f$ noise in the buried channel version is much reduced. There is a penalty in power dissipation in the buried channel device; for the same transconductance, a higher current is needed.

As already explained, for a vertex detector application, the off-chip amplifier and further processing should be external to the cryostat. Thus, the CCD amplifier needs to drive a capacitive load of some tens of picofarads. For slow readout, the first stage source follower alone is adequate, but for a high-speed system, the bandwidth

requirement implies a much larger transistor (lower g_m). Hence, the tendency in such cases will be to use a two- or three-stage output circuit, as shown in Fig. 39. With an optimized design, the noise performance is dominated by the first stage, even in the case where the later stage FET's are enhancement-mode devices.

A most important development in the early days of CCD signal processing was the invention of *correlated double sampling* or CDS [60], a technique which has since been adopted for charge detection circuits for microstrip detectors. The original aim was to reduce *reset noise* in CCD readout systems. The term "reset noise" refers to the unavoidable fluctuations in the node voltage (kTC_n) which arise from thermal fluctuations when the reset transistor is switched in and out of conduction. The procedure consists of sampling the node voltage twice after the reset, once before and once after the transfer of the signal charge. There are various options for filtering the signal preceding each sample; see Ref. [61] for a discussion. The optimal procedure consists of a signal integration for the same fixed period before and after sampling. In this case, the resultant total noise after sampling is given by

$$e_{nT}^2 = \int S_F(f)^2 e_n(f)^2 df ,$$

where $e_n(f)$ is the output circuit noise voltage per $Hz^{1/2}$ at frequency f, and $S_F(f)$ is the Fourier transform of the filter sampling function $S(t)$.

For the case of the dual integration for time τ,

$$S_F(f) = \frac{2\sin^2 \pi f\tau}{\pi f\tau} .$$

Note that this filter function falls to zero both at low and high frequency. Thus, CDS not only eliminates reset noise but also reduces the noise contribution from the output transistor, particularly in the low-frequency $1/f$ region, and in the high-frequency region (though the latter will normally be cut off by a suitable bandwidth limit to the main amplifier). The excellent noise performance of a modern CCD with the benefit of CDS is indicated in Fig. 40.

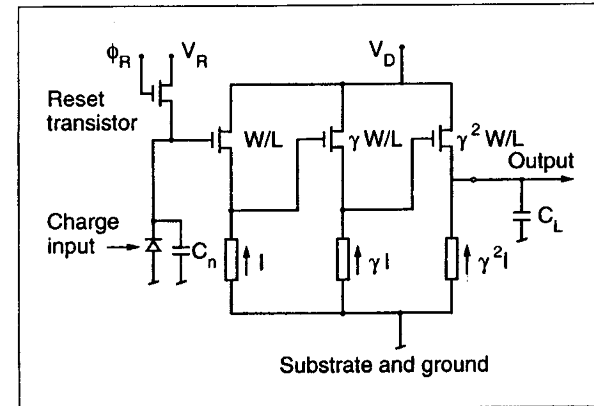


Fig. 39. Schematic diagram of a three-stage output circuit.

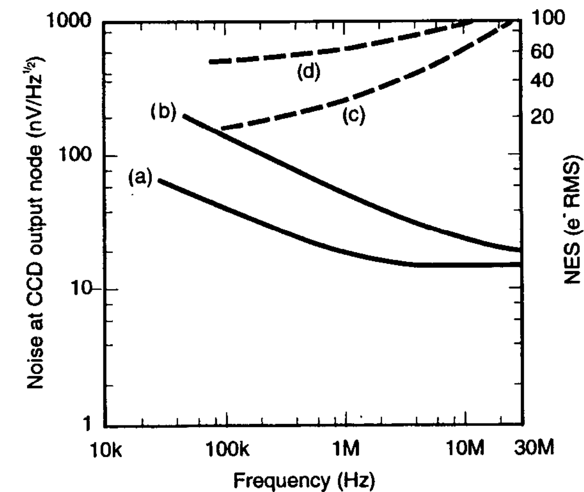


Fig. 40. Noise spectra of (a) buried channel and (b) surface channel first stage MOSFET's in a CCD (left-hand axis). (c) and (d) show the corresponding CDS noise equivalent signals in RMS e⁻ (right-hand axis).

The procedure normally followed in vertex detector readout, where readout speed is to be optimized, is to take advantage of the very small integrated charge to be expected in any row of the image, and hence, to reset the FET only at the end of each row. Thus, the signal charge of each successive pixel is just piled on top of its predecessors, and the CDS processing consists of simply taking successive differences of the filtered signal for pixel N, minus that previously sampled for pixel (N-1). It is not necessary to wait for the clock feedthrough from the linear register to settle; as long as this is constant from one pixel to the next, it is eliminated by the CDS differencing procedure. There is clearly a limit to this, for example, if the feedthrough is so large as to push the amplifier beyond its linear range during the sampling period, or if the sampled signal is swinging too rapidly at the moment of ADC sampling. In a well-controlled system, the readout noise *clocked* will be little greater than the value measured with the CCD unclocked. But achieving this in a system running at 10 MHz or above can be a major challenge for the circuit designer.

5.2.2.3 Vertex Detector Readout Options

Given the many options for CCD architecture and external electronics, the vertex detector designer has the opportunity to adapt the system design to the needs of the experiment, within wide boundaries. This has become particularly apparent as the cost of fully customized CCD design has fallen to the level where it is appropriate to plan on a completely new design for any experiment.

In this discussion, we restrict ourselves to the general architecture of frame transfer CCD's. Interline transfer devices, which can offer (via the variant of gated anti-blooming drains) the option of fast clearing on the microsecond timescale, are not considered. Despite this convenience, such devices are unsuitable for high-precision tracking applications where high detection efficiency is also essential. One cannot afford, in a vertex detector where the overall thickness is critical, to have detector planes which are only 70% efficient; close to 100% MIP efficiency is essential.

As we saw in the previous section, the original idea of disposing of signals from out-of-time tracks by charge diffusion does not work; the only way to get rid of unwanted signals is to clock the charge out via the output node. During the pre-trigger conditions, this can normally be best achieved by running the detector in "fast clear" mode. By synchronously clocking the parallel and serial registers at the upper rate limit for the former (around 100 kHz), one can sweep unwanted signals out in a mean time interval of around 10 ms for a large-area CCD. In a fixed target or rapid-rate collider environment, this implies a certain density of background hits in the CCD at the time of the event trigger. If this density greatly exceeds 1/mm^2, one should consider carefully whether this is an appropriate environment for such a detector. But up to this density (occupancy only approximately 10^{-3} in a detector with 20 μm square pixels), it is no problem to filter out this background.

In a modern experiment, top-level trigger decisions may take a while to arrive, say, 1 ms. During this time, one would be continuing the fast clear operation, in ignorance of the wanted data in the detector. On receipt of the trigger, the clocking would change to readout mode. Valid data from a region of, say,

$$20\mu\text{m} \times \frac{1\,\text{ms}}{10\mu\text{s}} = 2\ \text{mm}$$

at the edge of each CCD would have been lost in the time interval between the event and the trigger. It is no problem to allow for this small reduction in the fiducial region, at the detector design stage.

At this point, one is presented with numerous options depending on conditions. Let us take three examples, a fixed-target experiment, and two e^+e^- linear collider scenarios. These are based on actual experience but should not be taken to mean that CCD vertex detectors are necessarily limited to these environments.

For a fixed-target experiment, there is normally extra space available outside the spectrometer aperture. Therefore, it makes sense to extend the CCD area by at least the size of the fiducial region and to continue fast clearing until the valid data are all in a storage area well away from the high flux beam region (Fig. 41). This was the procedure used in the NA32 experiment. The detector could then be read out at leisure. In fact, to keep conditions even cleaner, a small kicker magnet was used to dump the beam during readout, but this was barely necessary.

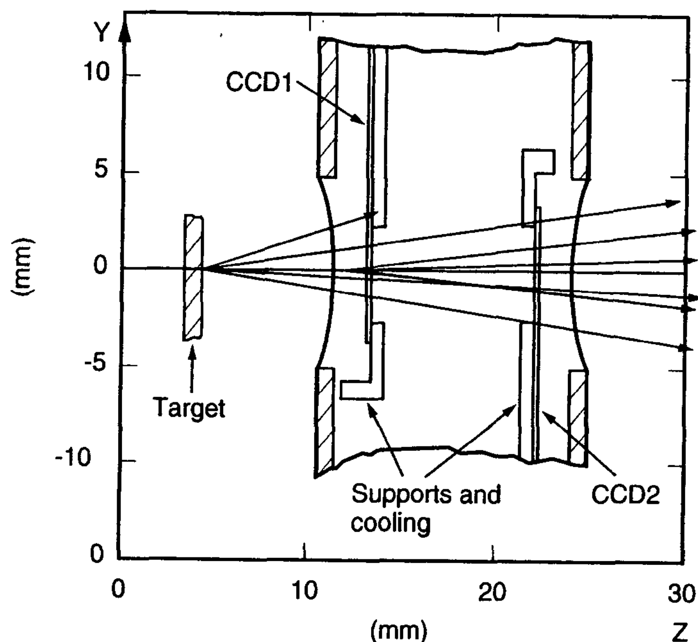


Fig. 41. CCD vertex detector for a fixed-target experiment (NA32). Data are fast-shifted into the quiet regions above and below the spectrometer aperture (CCD's 1 and 2, respectively) prior to readout.

For a linear collider environment such as SLC, the background comes mainly from synchrotron radiation and hence continues to accumulate throughout the readout period. Again, one has the possibility to inhibit this by dumping both beams. This has not been implemented in SLC because the backgrounds are quite tolerable. Furthermore, the trigger rate is sufficiently high that one would experience a significant deadtime loss from this. A CCD detector readout, though slow, can be made inherently deadtimeless; if a second trigger occurs during readout of one event, one just continues reading until data from the second event have been captured completely. Thus, if backgrounds permit, it is more efficient to avoid inhibiting collisions during the detector readout.

For the future linear collider, the SR background can be made negligible, and the small-radius background comes mainly from incoherent e^+e^- pair creation. Here, there are at least two extreme options. Firstly, to use a very small kicker magnet to move one of the beams by about 1 μm, out of collision, during readout. Secondly (if trigger rates are again high so that deadtime losses become an issue), to proceed as in SLC and live with the background. But in this case, one can take advantage of modern CCD design to use a multiport output register (up to 16 ports are commonly available, where in Fig. 32 we have illustrated just one in the corner). This increases the quantity of local readout electronics required, but one can then achieve full detector readout within the period of 5 ms between beam crossings. In practice, once the backgrounds and trigger rates for this environment have been quantitatively evaluated, one will be able to design a CCD vertex detector based on an optimized balance between these extremes.

The purpose of this section has not been to produce specific rules for the design of a CCD vertex detector readout system under specific experimental conditions; both of these are too variable for that. Instead, the hope is to encourage a flexibility of approach and to emphasize the opportunity presented to the vertex detector designer by fully customized CCD design.

5.2.3 Physics Performance and Future Trends

The major *attributes* of a CCD vertex detector are as follows:

1. Two-dimensional space point measurement, hence unsurpassed power for track reconstruction.

2. Two-track resolution. This is approximately 40 μm *in space* (see Fig. 42), compared with about 50 μm *in projection* for a strip detector, some 10^4 times worse.

3. Measurement precision about 3.5 μm for a MIP under typical readout conditions (RMS noise $\approx 50\ e^-$). Note that with less noisy readout (which at present means slower, but other improvements are possible) much higher precision can be achieved. For example, Ref. [62] demonstrates 0.9 μm precision for 15 keV X-rays in a CCD with 6.8 x 6.8 μm^2 pixels.

4. Thin active layer. This implies much lower conversion probability for X-ray background (e.g., synchrotron radiation) than in a thick microstrip detector.

5. Physically thin. Improved performance in terms of multiple scattering.

6. High granularity. Another factor leading to tolerance of high hit density (e.g., in particle jets close to the IP) and to high background. The former quality is demonstrated in Fig. 43, and the latter in Fig. 44.

A striking advantage of the high granularity is the almost total absence of merged clusters. This means that (in contrast to a microstrip-based vertex detector) it is straightforward to write a Monte Carlo program which accurately simulates the detector performance. This is demonstrated in the case of the SLD detector in Fig. 45, which shows the excellent agreement between data and Monte Carlo in the impact parameter distribution projected in orthogonal views. The Monte Carlo program has not needed to be fudged with any empirical smearing function in order to achieve this level of agreement.

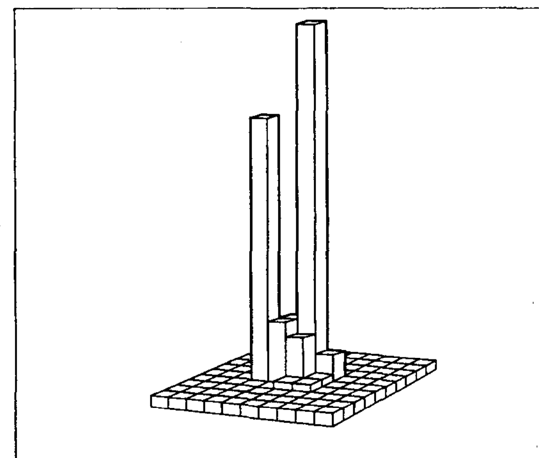


Fig. 42. Two MIP clusters separated in space by 40 μm, well resolved in a single CCD detector plane. Pixel size 20 x 20 μm^2.

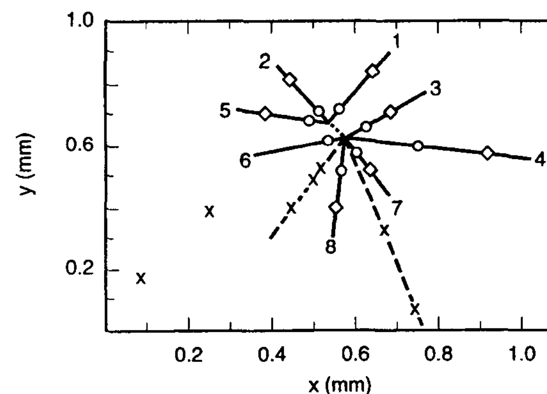


Fig. 43. Tracks from the IP and from a nearby charm decay in the NA32 vertex detector. Frame size 1 x 1 mm^2.

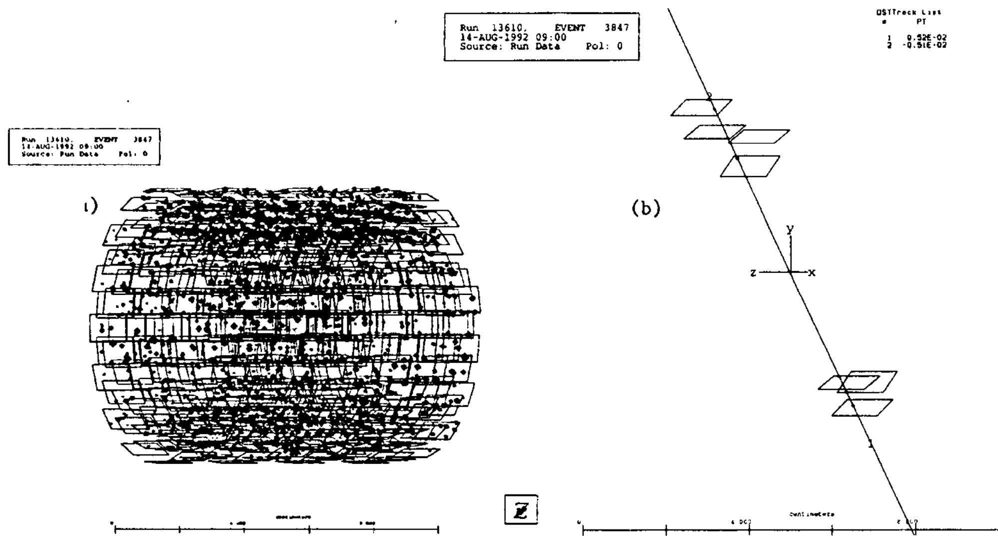


Fig. 44. (a) Raw data (mostly SR X-ray hits) in the SLD vertex detector. (b) The same event, with background filtered out by a drift chamber/vertex detector track-linking algorithm. This proved to be a $Z^0 \rightarrow \mu^+\mu^-$ event.

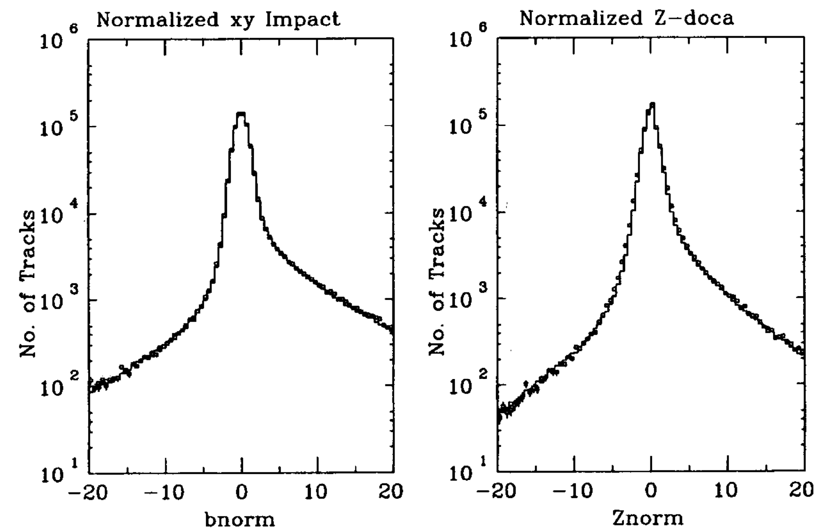


Fig. 45. Data (points) and Monte Carlo (histogram) distributions of impact parameter with respect to the IP in $Z^0 \rightarrow$ hadron decays (SLD experiment). The tails on the positive side are due to heavy flavor decays.

The major *deficiencies* of a CCD vertex detector are as follows:

1. Slow readout. This implies either beam suppression and hence a long dead-time associated with every top-level trigger, or a sufficiently benign background rate.

2. Radiation damage. See Sec. 6. In an environment of high hadronic flux, one either has to exchange CCD's fairly frequently (practicable in a fixed-target experiment) or avoid using them (e.g., at a hadron collider).

Both of these deficiencies can, to a great extent, be overcome with APS's (next section), but one then loses some of the previously listed attributes, as we shall see. Each detector type has its own niche.

The availability of fully customized large-area CCD's has opened the door for very exciting vertex detector developments. For example, Fig. 46 shows the CCD being used in the SLD upgrade detector. Adequate readout time is achieved with four outputs in this case. The devices have wire bonds at each end and are arranged end to end, one on either side of a beryllium motherboard, to build up two-CCD ladders out of which the detector (Figs. 47–49) is constructed. See Ref. [63] for a description of this 307 Mpixel detector.

For the future linear collider, one can be more adventurous. The CCD's can be thinned from 150 to 20 μm and attached to the same side of a beryllium stiffener (Fig. 50). By having outputs at one end only, the material in the active volume can be reduced from 0.35% RL per barrel (SLD upgrade) to 0.13% RL; see Ref. [64]. By a combination of larger and thinner CCD's, leading to higher precision-point measurements with more open geometries, one is seeing a steady evolution in the impact parameter precision achievable in the e^+e^- collider environment. The original SLD vertex detector yielded a measured precision of

$$\sigma^b_{XY} = 13 \oplus \frac{70}{p\sin^{3/2}\theta}\ \mu\text{m}$$

and

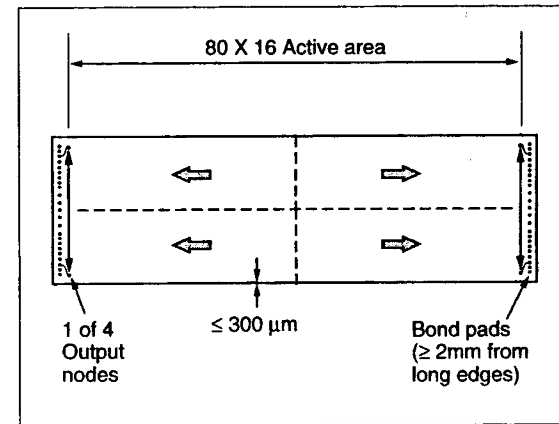


Fig. 46. Four-port CCD developed for the SLD upgrade vertex detector. Chip area = 13 cm^2.

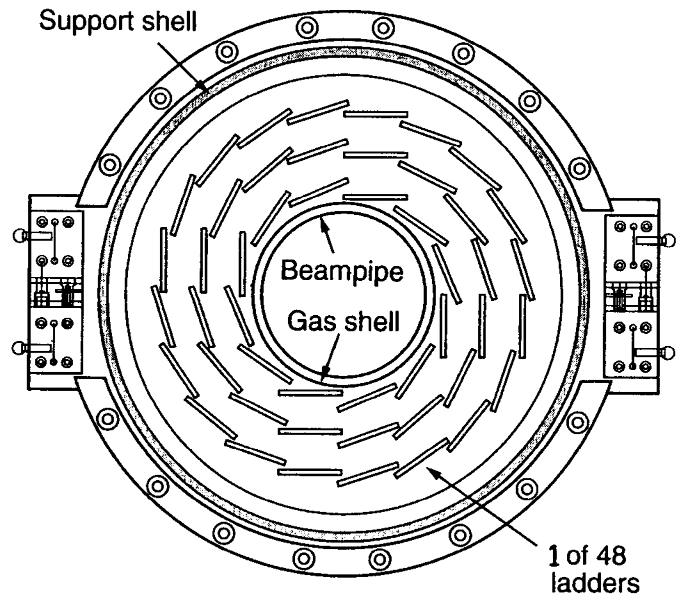


Fig. 47. Cross section (XY view) of SLD upgrade vertex detector.

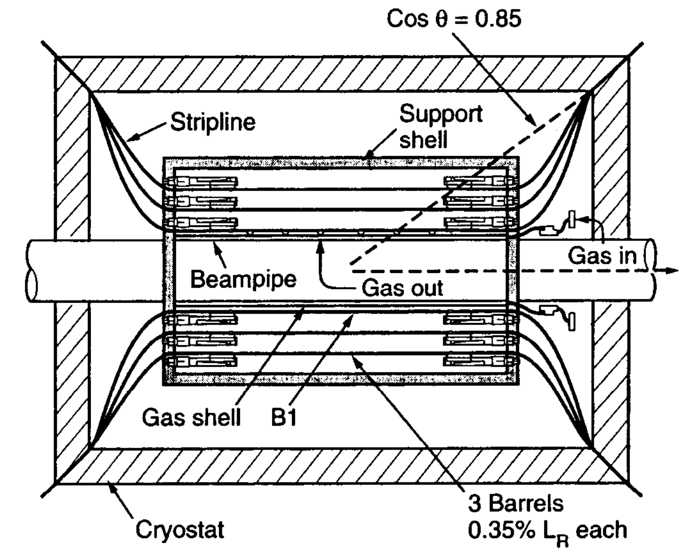


Fig. 48. Cross section (RZ view) of SLD upgrade vertex detector.

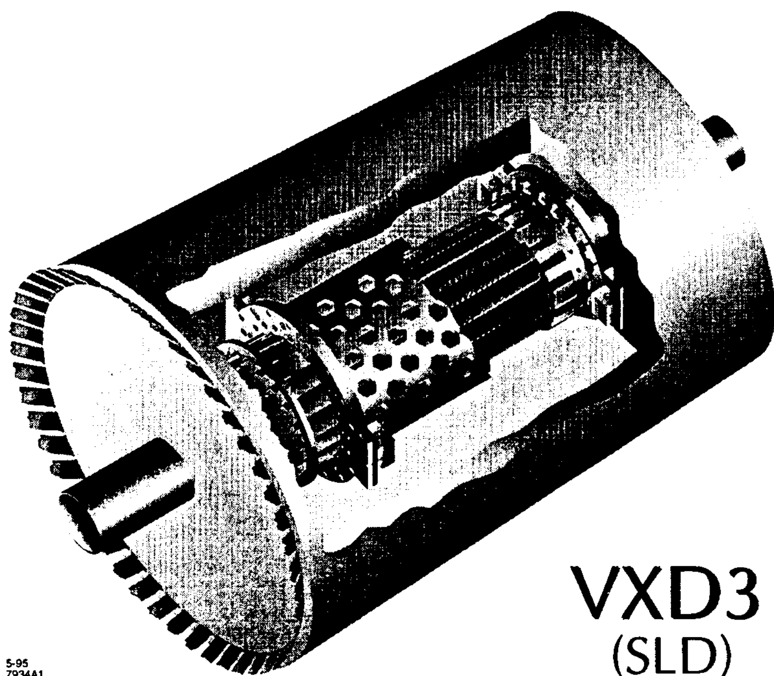


Fig. 49. Isometric drawing of SLD upgrade vertex detector.

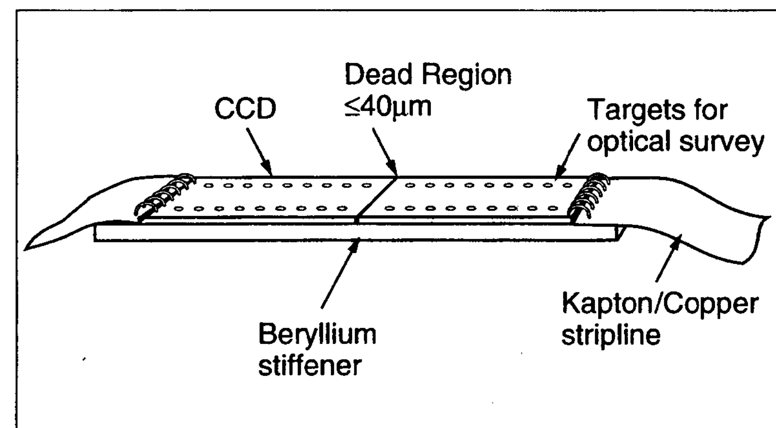


Fig. 50. Advanced two-CCD ladder design for a vertex detector for the future linear e^+e^- collider. Active length ≈ 16 cm.

$$\sigma^b_{RZ} = 35 \oplus \frac{70}{p \sin^{3/2}\theta}\ \mu\text{m}.$$

For the future LC detector, we anticipate

$$\sigma^b_{XY} = \sigma^b_{RZ} = 3 \oplus \frac{5.5}{p \sin^{3/2}\theta}\ \mu\text{m}.$$

Such a detector will be a tracking microscope of unprecedented power, having the capability to open numerous doors for exciting physics discoveries in the realm of Higgs and SUSY particles, as well as exploring the realm of the theoretically totally unexpected.

It should finally be emphasized that the low power dissipation in a well-designed CCD detector (approximately ten watts in the 307 Mpixel SLD upgrade detector) results in very low thermal management overheads. The detector can be cooled with a gentle flow of nitrogen gas, and the cryostat (see Figs. 48 and 49) consists of a low mass (< 1% RL) expanded foam enclosure. The operating temperature of around 200 K is chosen to minimize effects of radiation damage; see Sec. 6.

5.3 Active Pixel Sensors (APS's)

Both in the wider commercial world and in the area of scientific imaging, CCD's have established a dominant role, and as we have seen, are still in the midst of dynamic evolution. Yet they have limitations for vertex detectors, as has been emphasized. In addition, they have limitations for broader applications which have for many years stimulated studies, and more recently, actual devices, constructed according to a completely different architecture, the *active pixel sensor* or APS. The charge collection is as usual to one electrode of a reverse biased diode. But in the APS, these diodes form a discrete matrix over the device area, and each one is connected to its own signal processing circuit within the pixel. These circuits communicate to the outside world via some architecture, usually column-based. The essential point which has taken these devices into the real world has been the continuing shrinkage in feature sizes (and hence transistor sizes) available via the

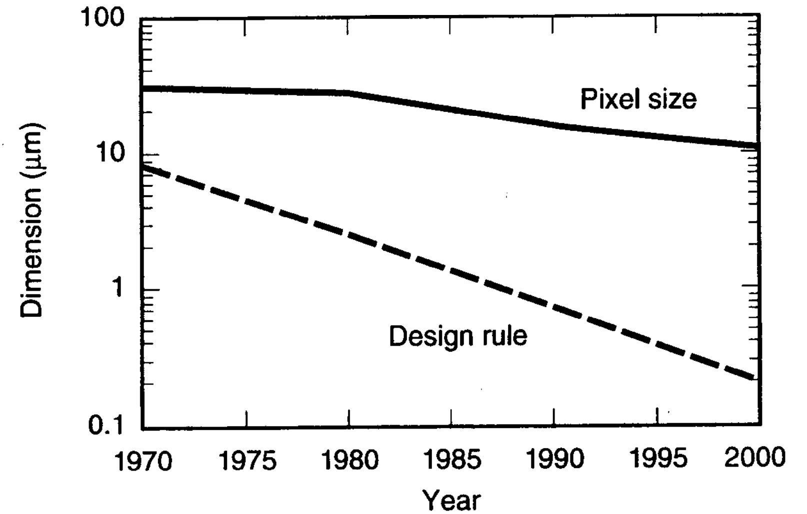


Fig. 51. Evolution of photolithographic feature size versus pixel size.

integrated circuit technology. Figure 51 (after Ref. [65]) illustrates this point. A recent review of developments in this field is to be found in Ref. [66]. Most of the commercial interest has been in the production of inexpensive CMOS chips combining low-quality imagers with processing electronics, for such applications as automatic chrominance control of automobile rear-view mirrors in response to headlights perceived in the field of view. One of the main factors limiting image quality is the separate processing of each channel; it is difficult to match these below 1%, and the eye is very sensitive to such blemishes. In terms of applications such as night vision systems, APS devices do have one interesting advantage over CCD's.

Since the readout can be nondestructive, one can watch on a monitor as the scene gains definition during the exposure time, possibly of advantage for some surveillance applications. However, commercial CMOS sensors made on low-resistivity material are typically limited to 1 or 2 μm detector active thickness, and hence are not useful for MIP detection. In addition, the growth in parasitic capacitance as the area is scaled up leads to escalating power requirements. Devices of area 100 x 100 pixels are relatively easy; beyond that, it becomes difficult. Finally, the spectacular evolution in design rule dimensions is generally associated with *smaller* IC's. Building sensors of area many square centimeters to such rules remains a distant dream. All of these factors do cause problems in the development of APS devices as vertex detectors.

For MIP detection, there are two main options. One of these is to take a high resistivity wafer and manufacture a single-sided microstrip-type detector, but with the strips cut into pads of the desired pixel dimensions, and to bump-bond this detector to a CMOS readout chip. This *hybrid* approach implies the less challenging route of keeping two technologies separate, rather than working to combine them. The second option, the *monolithic* approach, seeks to do the job on one chip. In both cases, the detector goals are similar and can be summarized as follows:

1. High-speed gating. In contrast to CCD's, the aim is to latch signals and associate them with specific beam cross-overs (BCO's) in environments such as LHC (BCO interval 25 ns), where the hit densities from each BCO are so high that one could not afford to integrate signals over more than one.

2. Time stamping. The idea is to transfer the hit information into a pipelined memory clocked at the BCO rate. On receipt of a level-1 trigger, those pixels that were hit at the corresponding BCO will be transferred to an on-chip buffer for readout, in the event that a level-2 or level-3 trigger is asserted.

3. Radiation hardness. Since (unlike the CCD) signal charge is not transported long distances through the silicon, the effects of bulk damage in terms of charge trapping are much reduced.

Leakage current impact is much reduced relative to microstrip detectors, due to the much smaller collection volume per detector element.

However, one does not escape the problems of type inversion and loss of charge collection efficiency (see Sec. 6). Furthermore, one has the same concerns regarding radiation effects in CMOS electronics (now in the active volume of the detector) as we noted in the microstrip environment.

As with microstrip detectors, there are three possible options for the readout electronics (binary, digital, and analogue), all of which are being actively pursued.

A major goal for physics is to be able to operate at relatively small radius (approximately 10 cm) for a reasonable lifetime in LHC at full luminosity. Several European and US groups are actively involved; for a recent review of the European work in this area, see Ref. [67].

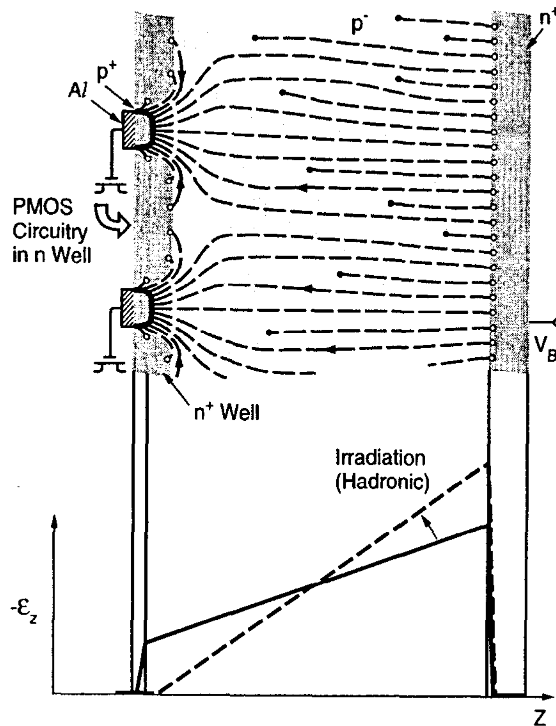


Fig. 52. Pixel structure of the generic monolithic APS. As with the microstrip detector (Fig. 25), hadronic irradiation tends to take the detector out of depletion, losing the signal.

5.3.1 Design Options

Let us consider in turn the two options available for MIP detection systems.

5.3.1.1 Monolithic Detectors

The generic monolithic detector pixel structure is sketched in Fig. 52. Full charge collection over the active area is achievable despite the fact that the p^+ collection implants occupy typically less than 10% of the surface area.

The main hurdle to overcome in moving from the commercial CMOS imager to a MIP-sensitive device was achieving compatibility between the high-temperature processing used for the CMOS activation steps and the preservation of high resistivity of the detector-grade silicon. This was demonstrated by Holland in a pioneering paper [68], in which the process of backside gettering is used for the removal of detrimental impurities from critical device regions. A similar process has been used since the mid-'80s in CCD manufacture, in which the heavily doped substrate is used to getter impurities from the epitaxial region from which the signal charge is collected.

To date, one prototype monolithic detector has been produced and demonstrated its capability for MIP detection [69]. This is an array of 10 x 30 pixels, pixel size 34 x 125 μm^2, overall area 1 mm^2. Ten percent of the chip area around two edges is taken up with CMOS circuitry. The analogue signals are read out sequentially at 1 MHz. Excellent MIP efficiency is achieved, with precision 2.0 μm x 22 μm in the two orthogonal directions. As with the commercial CMOS imagers, a considerable challenge is involved in scaling up the device size, but already a second generation detector of 96 x 128 pixels is under development [70]. European groups are also actively developing monolithic pixel detectors, aiming for the application to LHC vertex detectors.

5.3.1.2 Hybrid Detectors

Hybrid APS devices are being developed by several US and European groups for use in LHC detectors. The detector part consists of essentially a microstrip detector

structure, each strip being subdivided into a series of short strips which constitute the pixels. These are bump-bonded to the collection electrodes of a CMOS readout chip which would be similar in architecture to the monolithic versions. Hybrid detectors have the advantage of relative simplicity (no need to combine the detector and readout functions on one chip), but the complication of millions of interconnections and the disadvantage of extra material in the active volume. The thickness problem is exacerbated for both APS options by the high power dissipation (designers are aiming for about 0.5 W/cm^2, about 100 times higher than a CCD detector). Liquid-filled cooling tubes within the active volume are required.

Already one hybrid detector with 300 kpixels (of size 75 x 500 μm^2, too large for a vertex detector) is in use as a tracking detector in a high-track density, fixed-target environment [71]. This detector produces a binary output from each pixel at a readout rate of 2 MHz and has demonstrated excellent performance as a tracking detector. A second generation detector, shrinking the pixel size somewhat to 50 x 500 μm^2 while increasing the number of transistors per cell from 80 to 350 (using submicron technology), is in design. Zero suppression on-chip will greatly accelerate the speed of readout. These are vitally important steps en route to a viable LHC detector.

5.3.2 Performance and Future Trends

APS detectors for MIP detection are at a relatively early stage of development. They are demonstrating their capability in test beams and in fixed-target experiments as general tracking detectors. Their advancement to the level of an LHC vertex detector (see, for example, Fig. 53), with 100 Mpixels, depends on several challenging developments. Firstly, the functionality referred to earlier needs to be achieved in pixels of a reasonably small size, at least in one dimension (so that precise measurement in the $R\phi$ plane becomes possible). Secondly, the CMOS electronics needs to be sufficiently radiation hard, and finally, the detector needs also to demonstrate adequate radiation hardness. In fact, for the hybrid approach, one has in principle the option of going beyond silicon (see Sec. 7) for the detector, while retaining the rad-hard electronics for the readout. Overall, this is a very dynamic area of detector development, with an assembly of talented groups well-matched to the considerable challenges involved. Furthermore, even though the present prototypes are far from the eventual goals, ideas keep emerging and hold promise for ongoing important developments. An interesting new idea (Ref. [72]) involves the use of a p-channel JFET on a fully depleted high ohmic substrate (DEPJFET) for use as a unit cell for pixel detectors.

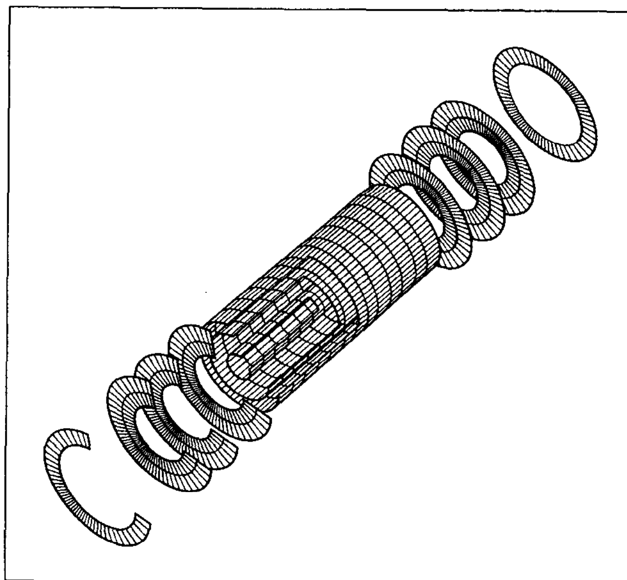


Fig. 53. Conceptual GEANT layout of a pixel vertex detector for ATLAS, consisting of three barrels plus endcaps. The innermost barrel (r = 4 cm) is not expected to survive for long at the full luminosity.

6 Radiation Damage in Silicon Detectors

6.1 Introduction

The subject of radiation damage in silicon devices has been studied intensively for decades, particularly in relation to the effects of nuclear reactors and weapons, both in the form of ionizing radiation and neutrons. References [73] and [74] are very useful books on the subject, Ref. [75] provides a valuable current review, and interesting historical reviews can be found in Refs. [76] and [77]. Yet, far from being exhausted, this is an extremely active area of study in connection with silicon tracking detectors. Why is this?

Firstly, silicon detectors are generally made from high resistivity material having long minority carrier lifetimes (order of magnitude milliseconds). Such material, unfamiliar to the field of electronic devices, behaves in unusual ways when irradiated; in general, it is more sensitive than electronic grade material to radiation effects. Secondly, there is an increasing number of important scientific applications (space-based equipment which spends time in radiation belts, detectors at small radius in LHC, etc.) for which the radiation environment is unusually hostile.

If we start by considering electromagnetic radiation of energy E_γ at long wavelengths (*e.g.*, visible light), the effects in silicon devices (electron-hole pair generation) are entirely transient. Above about 10 eV, electron-hole pairs in silicon dioxide are generated. These nearly all recombine, but as E_γ is increased, the hot carriers have an increasing probability of becoming independent within the oxide layer, leading to some degree of *surface damage*. Once E_γ exceeds approximately 250 keV, the energy is sufficient to start dislodging silicon atoms from their lattice sites; we are entering the realm of *displacement damage*.

For massive charged particles, displacement damage sets in at much lower energy. Low-energy protons are extremely dangerous due to the large cross section for *p* Si Coulomb scattering.

These two mechanisms form the basis of all radiation damage effects that concern us in regard to silicon detectors and the local electronics supporting them. Yet the

possible range of consequences of these effects is rather diverse. Let us consider these in some detail.

6.2 Ionizing Radiation

The band gap in silicon dioxide is 8.8 eV, and on average, 18 eV is needed to release an electron-hole pair. Figure 54 shows the time development of the charge distribution in an irradiated MOS structure.

The radiation generates a charge Q_g in the oxide, where $Q_g \propto t_{ox}$. The magnitude of this charge is totally independent of the nature of the oxide, rad-hard or "soft." A fraction f_c of the charge is trapped at the interface (where f_c can vary from 2% for a hard oxide to 80% for standard oxide), giving a trapped charge $Q_{tr} = f_c Q_g$. This induces a flat-band voltage shift ΔV_{FB}, where

$$\Delta V_{FB} = Q_{tr} / C_{ox}.$$

Now $C_{ox} \propto 1 / t_{ox}$, so

$$\Delta V_{FB} \propto t_{ox}^2.$$

Below 1200 Å, the dependence can be even faster, approximately as t_{ox}^3.

Note that this time development follows from the vastly different room temperature mobilities of electrons and holes in silicon dioxide, 2×10^5 cm^2/Vs and 20 cm^2/Vs, respectively.

As well as contributing a direct *interface charge*, the trapped holes can induce *interface states* in the case that they have been drifted towards the bulk silicon (as in Fig. 54). The interface state charge may be positive (for *n*-type substrates, i.e., *p*-channel MOS devices) or negative (for *p*-type substrates, i.e., *n*-channel MOS devices).

Note that at reduced temperature, the holes are effectively immobilized, so there is no performance difference between soft and hard oxide. This, however, is not a serious

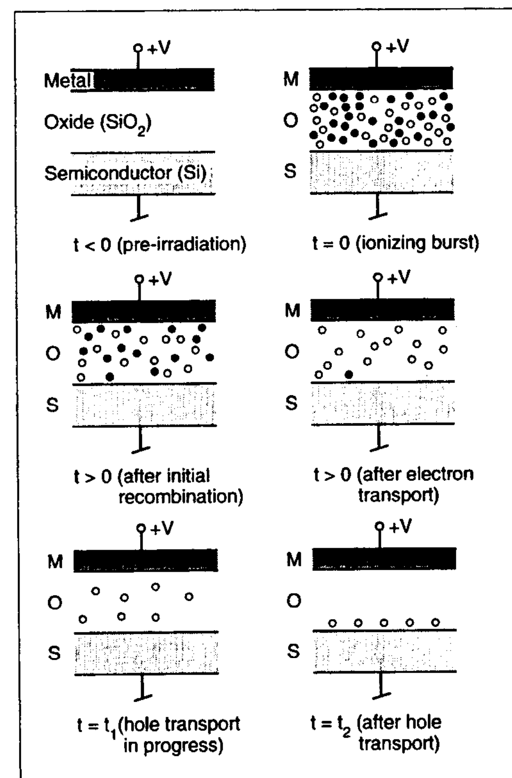


Fig. 54. Time development of charge distributions following a burst of ionizing radiation on a positively biased MOS structure.

concern for detector applications, since the detector can always be cycled up to room temperature for brief periods, restoring the holes to their normal room-temperature evolution.

The induced flat-band voltage shifts can cause various device and detector malfunctions. For nonhardened oxide, the effects are large; for example, 10 krads on a 700 Å oxide induces a 2 V shift. What can be done to reduce this?

> Firstly, the 100 substrate orientation is much preferred (minimal level of dangling bonds).
>
> Secondly, minimize t_{ox}, though not so far as to suffer a serious loss of device yield.
>
> Finally, observe special procedures in post-gate processing (most notably, keeping the temperature below 900 °C).

As well as the *gate oxide*, charge buildup in regions of *field oxide* on the device can be equally significant [78]. Huge voltage shifts are associated with the thick field oxide. In the case of *p* substrates, these induce inversion layers which can short all the *n* implants within the substrate. These effects are common to all device types (JFET's, bipolars, MOS devices, and detectors). Careful design practices (e.g., guard structures) are required to avoid them.

Recent developments may lead to a further breakthrough in the area of radiation hardening. It has been found that the conventional use of *hydrogen* to saturate dangling bonds may not be optimal. The Si-H bond is unstable with respect to X-radiation. To this end, a new process has been developed [79] based on semi-insulating polycrystalline silicon or SIPOS. Possible implications for radiation detectors are being evaluated.

6.3 Displacement Damage

Atomic collisions with high momentum transfer, as well as nuclear interactions, can permanently alter the properties of the bulk material. Such processes are grouped together as the source of *displacement damage*, in which silicon atoms are displaced from their normal lattice locations. These effects may be local single-atom displacements, in which case the damage is classified as a *point defect*; such defects commonly result from high-energy electromagnetic irradiation (X-rays or electrons). Displacement damage may also occur as *damage clusters* which consist of relatively large disturbed regions within the crystal; such defects commonly result from nuclear interactions of (for example) neutrons and protons. The most probable events of this type are elastic Coulomb scattering of silicon nuclei by the incident high-energy (charged) particle. As shown in Fig. 55 (based on Ref. [73]), a 50 keV recoil silicon nucleus can create clusters of damage (with knock-on and stopping of other nuclei) over a volume of several hundred Ångstroms typical dimensions.

The bulk damage due to the passage of high-energy particles can be described by the number of atomic (silicon) displacements per cm of track length. For protons traversing silicon, this rate falls from $\approx 10^4$/cm at 1 MeV to $\approx 10^2$/cm at 1 GeV. This nonionizing energy loss (NIEL) depends both on the particle type and energy, though at high energy (above approximately 1 GeV), it is nearly the same for all hadrons. See Refs. [80] and [81] for pioneering papers on this subject. The NIEL for various particle types is plotted in Fig. 56. To a good approximation, displacement damage effects depend only on the overall nonionizing dose received, except that the effects are much reduced for electromagnetic radiation. In this case, as well as the low specific NIEL value, all momentum transfers are so low as to liberate at most one atom (leading to point defects as opposed to cluster damage). Specifically for 5 MeV particles, an electron, proton, and neutron produce a primary knock-on atom (PKA) which on average generates in total 1.2, 4.2, and 8000 further displacements, respectively.

As far as the primary displacement damage is concerned, the generation of these clusters of vacancies (V) and interstitial silicon atoms (I) is the entire story. Even in low-resistivity material, the concentration of dopant atoms is so low that they play effectively no part in this process. However, the role of dopant and impurity atoms is crucial in understanding the electrical effects, because both vacancies and interstitials are mobile, and can combine stably with atoms other than silicon in the crystal structure.

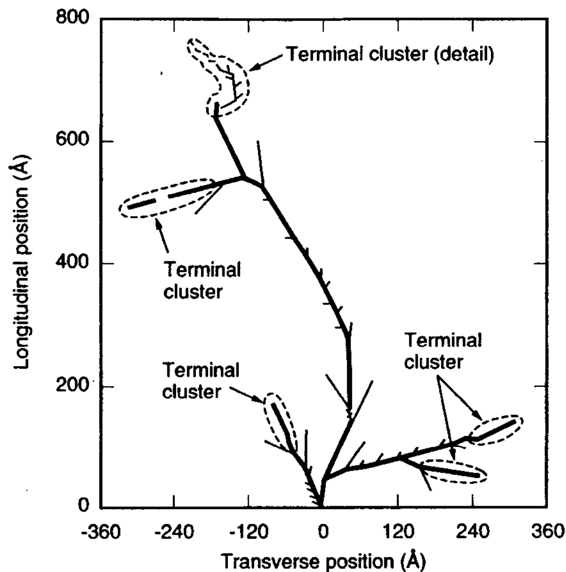


Fig. 55. Development of cluster damage due to a primary knock-on silicon atom of 50 keV, within the bulk material.

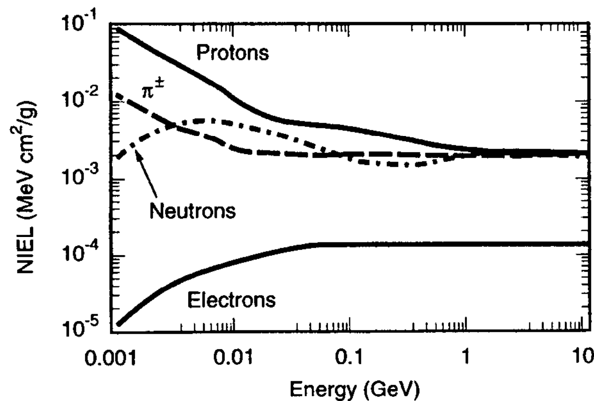


Fig. 56. NIEL for various particle types as a function of energy. A frequently used unit is the NIEL associated with a 1 MeV neutron.

Before considering this, we note that the practical effect is the development of a large number of energy levels within the band gap, some donor-like and some acceptor-like, some being capable of existing in more than two charge states. These levels, depending on their state of occupancy, can act as trapping centers and hence seriously degrade the minority carrier lifetime. In addition, these extraneous generation-recombination centers cause extra leakage current in depleted material and reduction in the carrier mobility. For electronic grade silicon, the description of displacement damage effects in terms of these macroscopic properties is sufficient.

For detector-grade material, the situation is more complex. It is rather like comparing the effects of an earthquake on a steel frame building as opposed to one made with bricks. The basic physics processes are the same, but the effects are very different. Detector-grade material (high resistivity, long minority carrier lifetime) is particularly sensitive to radiation-induced displacement damage. Let us start with an empirical description of what is observed, and then tackle the basic physics processes involved.

Measurements on *undepleted* detector-grade silicon reveal a monotonically increasing *rise* in resistivity with dose. This can be understood in that the disordered material generates a huge number of extra donor and acceptor states, populating the entire band gap. Statistically, the Fermi level drifts to approximately midgap, so the material becomes effectively *compensated.*

However, when one depletes the material, one finds a leakage current which grows linearly with dose (i.e., accumulated NIEL) but which anneals with more than one time constant. One is seeing the global effect of generation current from a number of intergap states which physically evolve with time. *Provided* the detector is designed for low-temperature operation, the leakage current is not a fundamental problem, since one can reduce it to an acceptable level by cooling.

Next, we consider the effective dopant concentration N_{eff}. From the resistivity measurements, we might have expected the material to change from n type to intrinsic, and to stabilize with a low value of N_{eff} as the Fermi level sits around mid-gap. On the contrary, as we saw in Figs. 25 and 52, the depleted material behaves

quite differently from the material in equilibrium. It becomes steadily more p type with fluence, going through *type inversion* at an equivalent fluence of approximately 5×10^{12} neutrons/cm^2, as shown in Fig. 57. As we saw in the case of the leakage current, the material shows a medium-term annealing behavior, which is extremely temperature dependent [82–84]. For highly irradiated samples (well beyond type inversion), N_{eff} falls back over a period of days (at room temperature) or years (at -20°C). However, this is by no means the end of the story. At room temperature, the material now enters a *reverse-annealing* phase; N_{eff} increases. The material becomes ever more p type; even after a year, the trend continues. This behavior can be entirely avoided by cooling. The data taken at -20°C show ongoing annealing to the end of the test period, with no tendency to flatten off; the material just becomes steadily more nearly intrinsic.

So what are the microscopic physics processes during this complex behavior pattern? One could even ask, why do we care? The answer to the second question is that there is a possibility that, once the details are understood, it may be possible by *defect engineering* to improve the radiation hardness of the material, e.g., by staving off the reverse annealing problem even at room temperature. This is a very active area of research. At a recent conference, contributions were varied and somewhat controversial [85]. DLTS measurements backed up by a semiconductor device model have enabled Matheson *et al.* [86] to produce a plausible explanation for some of the most striking of the above observations. Their results can be summarized as follows:

1. Based on photoluminescence and DLTS measurements on high resistivity n-type Wacker material, they find the following concentrations of expected and unexpected impurities:

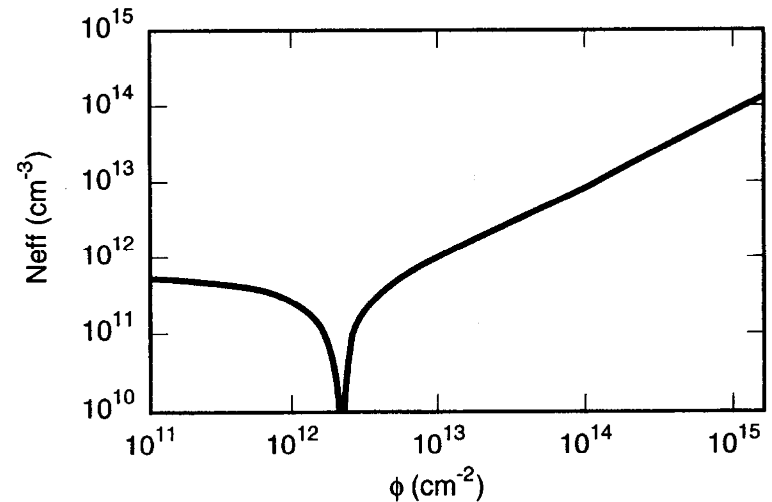


Fig. 57. Dependence of effective dopant concentration N_{eff} on fluence, at room temperature. The material, initially n type, goes through type inversion for $\phi \approx 5 \times 10^{12}$ neutrons / cm^2 equivalent dose.

$[P] \approx 10^{12}\,\mathrm{cm}^{-3}$

$[C_s]$ $1-5\times10^{15}\,\mathrm{cm}^{-3}$ (substitutional carbon)

$[O]$ $5\times10^{15}\,\mathrm{cm}^{-3}$

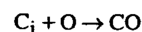
$[H]$ $10^{14}\,\mathrm{cm}^{-3}$.

2. The mobile I and V centers diffuse away from the damage cluster and eventually mostly undergo one of the following reactions:

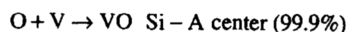
$I + V \rightarrow 0$ (repair)

$I + C_s \rightarrow C_i$ (interstitial carbon, transient)

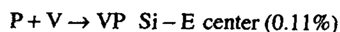
$C_i + C_s \rightarrow CC$



$C_i + O \rightarrow CO$



$O + V \rightarrow VO$ Si – A center (99.9%)



$P + V \rightarrow VP$ Si – E center (0.11%)

$V + V \rightarrow VV$ (but most divacancies emerge from the primary cluster)

$VO + V \rightarrow V_2O$.

These observations rule out some of the almost-established folklore regarding the behavior of detector-grade material. The long-held belief that the resistivity rise was due to donor removal is excluded by the above figures. The phosphorus concentration is simply too low by several orders of magnitude.

3. The authors hypothesize that generation of some deep level acceptor is responsible for the reverse annealing. V_2O is a candidate, suggesting that a less oxygen-rich starting material might be free of this effect.

4. If such a deep-level acceptor is responsible, how does it become filled? The authors hypothesize that this is due to the bulk leakage current, and indeed demonstrate a suggestive correlation between the measured N_{eff} values during the annealing phase and the square root of the leakage current damage constant α. If this were the only effect involved, one would find simple proportionality between these. In fact, there is a nonzero offset, but it seems to me likely that this mechanism is a good part of what is a rather complex picture.

These pioneering studies have led to a concerted effort by LHC physicists to further understand the bulk radiation effects in detector-grade material, possibly leading to more radiation-resistant material in the longer term future.

The final empirical observation relevant to bulk damage effects in detectors is that of loss of *charge-collection efficiency*, CCE. For a 300 μm thick depleted detector, one finds approximately a 10% loss in CCE for a dose of $10^{14}\,\mathrm{n/cm^2}$ equivalent. This is presumably related to the high density of trapping centers generated and probably implies a basic limit to the tolerable radiation dose for such thick detectors, at around the $10^{15}\,\mathrm{n/cm^2}$ level.

6.4 Detector-Specific Effects

6.4.1 Microstrip Detectors and APS Devices

The major challenge which is driving much of the R&D discussed in the previous section is the LHC tracking detectors (vertex region and Central Tracker at larger radii). At small radius, the predominant background comes from pions of energy 100 MeV to 1 GeV, with albedo neutrons playing a relatively larger role at large radii [87]. The overall dose as a function of radius is listed in the following table, for a seven-year run at $\mathcal{L} = 1.7\times10^{34}\,\mathrm{cm^{-2}s^{-1}}$.

R (cm)	Ionizing Dose (Mrads)	Fluence (1 MeV n/cm$^2\times10^{-14}$)
4	117	40
6	55	18.8
10	22	7.4
11.5 L1	17	5.8
14.5 L2	11.4	3.9
30	3.8	1.3
52	1.8	0.6

If detector replacement during this period is assumed, one is entitled to divide by some factor, but there are reasons (beam-filling periods, etc.) to raise the estimate. Overall, these figures probably give a reasonable indication of the requirements.

Discounting, for the purposes of this discussion, the prospect of major progress through defect engineering, what do these figures imply for silicon tracking detectors in such an environment? (While we are discussing this in the context of LHC, the implications for other hadron beam or collider experiments follow directly.)

Within a radius R = 30 cm, one suffers increasingly serious CCE loss. This would be fatal for microstrip detectors. However, silicon pixel devices, with much smaller collection node capacitance, might be able to survive with a considerably smaller signal size, i.e., smaller depletion depth.

Beyond R = 30 cm, the detectors still go far beyond type inversion during their working life. This means one of two things. Either they are made on *p*-type substrates or they must be equipped with guard rings, etc., that allow the junction to move from the *p* side to the *n* side during operation. If one collects signals from the *p* strips (hole signal), one has to beware of loss of signal as the radiation dose increases (remember Fig. 25). This can be avoided by steadily increasing the operating voltage. Alternatively, one may collect the signal from *n* strips (electron signal), in which case the charge collection degrades more gracefully, as the devices fall below depletion. In either case, to prevent the global signal from falling too low, it is necessary to keep the devices at least almost fully depleted. This implies (for R $\geq$ 30 cm) high operating voltage (approximately 1 kV) at the end of the seven-year period, unless the detectors are cooled. Cooling to say -10°C can keep the depletion voltage down to approximately 150 V as well as providing the essential reduction in leakage current. However, if the detector is warmed up for a total of even one month during the seven-year period, the depletion voltage increases by a factor of two, due to rampant reverse annealing during that time.

In conclusion, environments such as LHC with high hadronic background provide a major challenge for silicon detectors. By switching from microstrips to pixels, one can hope to push below R = 30 cm, but within R = 10 cm, the region of interest for a general-purpose vertex detector with good impact parameter resolution, even these devices would not have a useful life expectancy at the full LHC luminosity. The most optimistic current expectation is for an inner layer of pixel detectors on R = 11.5 cm, with an active thickness of 150 μm and (at end-of-life) a depletion voltage of 350 V, 2 nA/pixel leakage current, and 30% ballistic deficit.

The hopes of being able to move into the heat below 10 cm have stimulated a considerable activity in devices made of material beyond silicon, as discussed in Sec. 7.

6.4.2 CCD's

For use as vertex detectors, CCD's have a role mainly in fixed-target experiments (where they are required to cover only a small area and hence can be changed frequently) and in e^+e^- collider experiments, where the hadronic backgrounds are low. Hence, our major concern is their tolerance of ionizing radiation. However, for other applications (notably space-based detectors that suffer from solar flares or spend time in the proton-radiation belts around the earth), the hadronic bulk damage effects can be serious.

Regarding ionizing radiation, the effect to be concerned with in CCD's is the slow shift in the potential of the parts of the device overlaid by gate oxide (the imaging area and output register), in relation to the potential of the output node (nominally fixed). Figure 58 (based on Ref. [88]) shows the flat-band voltage shift after irradiation of a CCD gate oxide at two extreme temperatures. For an *n*-channel CCD, the sign of the electric field is optimal (directed towards the gates, negative in the convention of Figure 58). Thus, at room temperature, the flat-band voltage shift ΔV_{FB} is negligible. However, the situation worsens as the temperature is reduced, and by 77 K, ΔV_{FB} is huge and equally bad for either polarity. Note that even at low temperature, ΔV_{FB} is negligible for an unbiased gate, so CCD's (and, in fact, any MOS devices) in radiation environments should be powered off when not in use. Furthermore, for devices operated cold, an occasional, brief warm-up to room temperature restores ΔV_{FB} to a much reduced level. One can, in addition, tune the output node voltage within limits. Modern standard production CCD's have

$\Delta V_{FB} \approx 20$ mV/krad and can be tuned for operation up to 100 krads. More advanced devices are now proven up to 1 Mrad of ionizing radiation.

In all this, it is extremely important that the polysilicon gate structure completely overlays the oxide layer. Figure 36 is an oversimplification; the actual CCD structure is sketched in Fig. 59.

Regarding bulk damage, we need to consider the effects on dark current, charge collection efficiency, and charge transfer efficiency. Even in heavily irradiated CCD's, the excess dark current can normally be dealt with by modest cooling. Given the thin epitaxial layer, the requirements made on minority carrier lifetime are not severe, and there is essentially no problem with CCE into the potential wells. However, once the electron charge packet starts its long journey to the output node, the situation is far more dangerous. The *n* channel being relatively highly doped, the generation of bulk defects is considerably simpler than was discussed for detector-grade material, being closely similar to that encountered in electronic devices. The mobile vacancies are predominantly captured by the phosphorus dopant atoms, giving an increasing density of Si-E centers (positively charged donor-like defects when empty; with an energy level E_{tr} of 0.44 eV below E_c). These defects have a high probability of capturing signal electrons which come within their electrical sphere of influence. Let us consider this case, a single type of bulk trap which uniformly populates the *n* channel. This situation is a restricted case of the general Shockley-Hall-Read theory of carrier capture and emission from traps, in which only capture and emission of electrons from/to the conduction band plays a part. Hole capture and emission are irrelevant since we are concerned with donor-like traps in depleted material. This situation has been considered by various authors [53, 54, 89, 90].

Let us firstly take a qualitative look at the situation. As the charge packet is transported from gate to gate (within a pixel or between neighboring pixels), *vacant* traps that lie within the storage volume of the charge packet will tend to capture electrons. If the traps are filled (either fortuitously, due to the passage of an earlier signal packet, or deliberately for this purpose by the injection of an earlier "sacrificial" charge packet), they will permit the signal electrons to pass undisturbed.

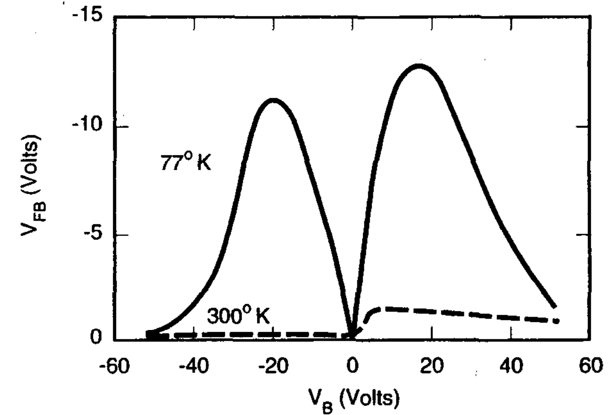


Fig. 58. Flat-band voltage shifts after 100 krads of ionizing radiation across a hardened gate oxide.

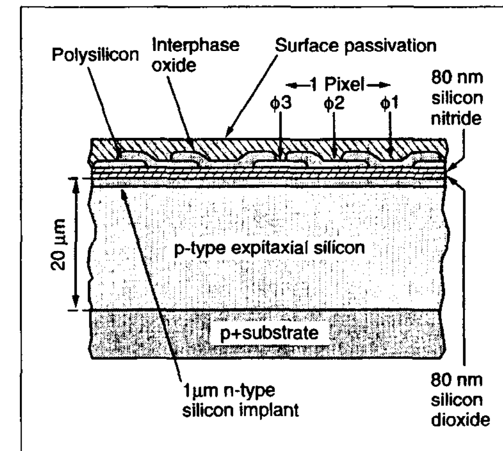


Fig. 59. Gate structure of a modern three-phase CCD register, designed to avoid potential wells due to radiation-induced charge build-up or other spurious charge in the oxide or surface passivation layers.

Also, if the signal packet is transported at a sufficiently high clock rate that the dwell time τ_g under any gate is small compared to the trapping time constant τ_c, the signal electrons will pass. Also, if the trap emission time constant τ_e is small compared with the clock pulse rise/fall time τ_r, the trapped electrons will be re-emitted in time to rejoin their parent charge packet. Only if electrons are *trapped and held long enough* to be redeposited in the next or later potential well, does the process contribute to a loss of CTE. This is evidently a multiparameter problem with some room for maneuver.

Let us now look at the process quantitatively. Assuming all traps are initially empty, the CTI is given by

$$\mathrm{CTI} = \sum_{j=1}^{N_F} F_j \times \frac{N_{tr}}{N_s}\left[1-\exp\left(\frac{\tau_r}{\tau_e}\right)\right].$$

N_F is the number of phases per pixel (three for a three-phase structure).

F_j is the fill-factor for phase j, i.e., the probability that a trap in the charge packet storage volume will become filled during the dwell time.

$$F_j = 1-\exp(-\tau_g / \tau_c).$$

For cases of practical interest, τ_c is of order of magnitude nanoseconds, and F_j may be taken to be unity. N_{tr} is the trap density. N_s, the signal charge density, is a function of the signal size but is effectively constant for charge packets larger than approximately 1000 e^- (Ref. [90]). For smaller charge packets, the effective signal density is reduced, and the CTI is correspondingly degraded. For very small charge packets of N_e electrons, one expects $N_s \propto 1/N_e$ since the signal electrons will occupy a constant volume determined by their thermal energy and the three-dimensional potential well in which they are stored. The volume of this potential well can be reduced (by techniques referred to as narrow channel or supplementary channel processing), so yielding a factor of up to four improvement in CTI, compared with standard channel devices [91].

Now

$$\tau_e = \frac{\exp[(E_c - E_t)/kT]}{\sigma_n X_n v_n N_c}.$$

The terms in the denominator are respectively the electron capture cross section for that trap type, an entropy factor, the electron thermal velocity, and the effective density of states in the conduction band. The numerator tells us that for shallow traps (or high temperature), τ_e is likely to be short, and conversely for deep traps and/or low temperatures, τ_e is likely to be long. In fact, for deep traps and appropriate clock times, by reducing the temperature, one can sweep the CTI through its full range from approximately zero (since the charge is re-emitted into the parent pixel during the drive-pulse risetime) to $3N_{tr} / N_s$ (for a three-phase CCD) and back to zero, as all traps are filled by some long preceding deliberate or accidental charge packets to have been clocked out of the device. Figure 60 (from Ref. [90]) nicely illustrates this point. This demonstrates the growth in CTI due to irradiation of a CCD with a high-energy electron source. The density of Si-E centers increases, but the effect on CTI can be minimized by operating at or below 190 K, where the trap emission time becomes adequately long. The degradation in CTI below 160 K is due to the emission time of a shallower trap becoming significantly long. Eventually (by about 70 K), the phosphorus donor ions can play a role (carrier freeze-out). This sets an effective lower limit to the useful operating temperature of *n*-channel CCD's.

For hadronic irradiation of CCD's, because of the much greater NIEL factor, the damage rates are greatly increased. The CTI effects are qualitatively similar [92], and it is believed that the Si-E center is responsible for 85% of the defects, with 15% due to the VV (divacancy) presumably generated in the initial damage clusters. There are possibly some further discrepancies with respect to the electromagnetic damage data; what is urgently needed are controlled experiments, involving both electromagnetic and hadronic irradiation of the same CCD types under similar clocking conditions, with well-defined injection of "sacrificial" charge packets to (as far as realistically possible) saturate the traps. One should also note the necessity to study the serial *and* parallel register in any test program. One might select a temperature low enough to

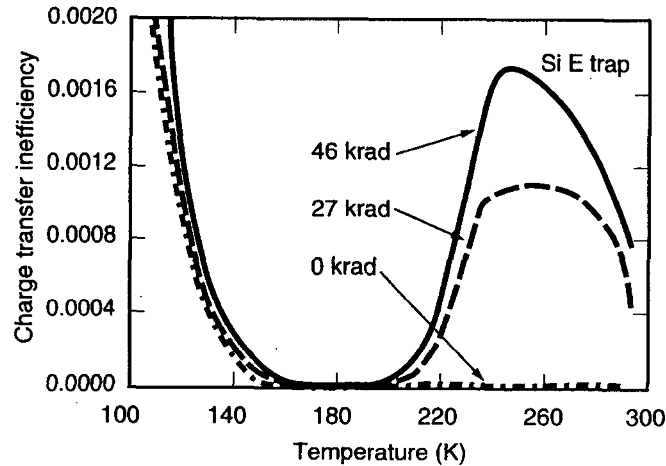


Fig. 60. From Ref. [90], effect of radiation damage on CTE in a CCD, as a function of operating temperature. Irradiated with a Sr^{90} β source.

have good parallel CTE against all traps, but find that this corresponds to long enough emission times for some intermediate depth trap to cause serious CTE loss in the serial register. There is no absolute rule that the serial register CTE exceeds that of the parallel register, though this is often the case.

6.4.3 Local Electronics

The issue of radiation hardness of local electronics for vertex detectors is extremely dependent on the detector type as well (of course) as on the nature of the experiment. In fixed-target experiments, it is no problem to keep the electronics out of the beam, so the issue does not arise. In collider experiments, it has already been mentioned that for CCD-based vertex detectors, it is desirable for thermal management reasons to keep the local electronics outside the cryostat, and due to the analogue multiplexing (by a factor of about 10^6) on the CCD, the number of connections required is small. This allows the electronics to be tucked away behind the tungsten mask used in the small angle region to shield the overall detector, providing a virtually radiation-free environment, even though the detector itself may accumulate as much as 1 Mrad during its working life.

The issue therefore really only arises in the case of nonmultiplexed detectors (microstrip and APS detectors particularly) where the electronics has to be connected by wire bonds or bump bonds, and is therefore inevitably in the same high radiation environment as the small-radius detectors. The worst example is LHC, for which the dose levels tabulated in Sec. 6.4.1 apply equally to the electronics. For the Central Trackers (reaching in to $R \approx 30\,\text{cm}$ but not below), radiation resistance up to around 10 Mrads and $2 \times 10^{14}\,\text{n/cm}^2$ equivalent is required. This is achievable with "standard" rad-hard CMOS and bipolar IC processing. The commercial situation is somewhat unstable. Companies that previously worked closely with the defense industry in the USA and Europe are in some cases looking for new markets and are offering their services to ASIC designers in general, including those at HEP laboratories. Some of these companies, however, have decided that the nondefense markets are inadequate and have ceased to offer facilities for rad-hard electronics. As has been mentioned, the trend towards submicron processing lends itself incidentally to improved radiation resistivity, though care has still to be taken over such issues as

field oxide isolation. In general terms, the needs of the central tracker community for radiation-resistant microstrip electronics are well served; if anything, they have a wider choice than might absolutely be necessary.

For the vertex detector region ($R \leq 10$ cm), the situation is far more challenging (> 100 Mrads and > $5 \times 10^{15} n/cm^2$ at R = 4 cm). Furthermore, hit densities and degradation in the detectors (noise related to leakage current, loss of charge collection efficiency) mandate pixel-based detectors. The precision requirements of a truly general vertex detector would imply precision of a few microns in both views (and hence small pixels). However, this high granularity should not be achieved at the expense of excessive power dissipation, or else the material introduced per layer (including cooling systems) becomes unacceptable. A general aim of not more than 1 W/cm^2 and 1% RL per layer (detector plus electronics) is generally considered reasonable, and the granularity (i.e., the physics capability) is adjusted to suit. This seems to me to be a very reasonable strategy; it has stimulated a huge and diverse effort, and as the technology advances, the physics requirements will become better met. The high particle fluxes at LHC (small radius) mandate a complex circuit for each pixel, and the requirement of radiation hardness, of course, increases the area of that circuit. This is a development area in which it will be necessary to take advantage of the latest developments into and beyond the time of LHC startup ten years from now. Fortunately, vertex detectors are compact and inexpensive in relation to their value for physics, and so can be rebuilt and upgraded pretty much in response to the technological advances.

6.5 Future Prospects

The radiation levels in space-based systems and accelerator environments such as LHC are generating new challenges. Those faced by the vertex detectors at hadron colliders are by far the most difficult. Detectors will necessarily be pixel-based, and the low-and-slow CCD pixel technology must be replaced by APS devices with as-yet unattainable performance. There is a temptation to abandon silicon as being inadequate for these radiation levels, both for the detectors and for the electronics. Yet it is clear that the essential limits to the radiation hardness of silicon, particularly as regards displacement damage in detector-grade material, are far from understood. The role of defects such as carbon and oxygen is only now beginning to be assessed. It therefore seems entirely appropriate to push hard on these developments, and the field of *defect engineering* is being applied to very good effect in elucidating this subject. If sufficient progress is made in radiation hardening, all the other attributes of silicon will give it a tremendous advantage over rival technologies. On the other hand, to have complete confidence that these enormous problems will be solved would be equally naive. It is therefore very important that some groups put their efforts into exploring alternatives, as discussed in the next section.

7 Beyond Silicon

Driven by the fierce radiation levels in future vertex detector environments (notably at LHC), it is natural to ask whether other detector media or IC technologies might be better suited to the task. Given the high probability that the pixel-based detectors to be used in these environments will necessarily be hybrid (as opposed to monolithic), it is even possible that the detector and electronics IC, bump-bonded together, may be made of different materials, either or both of which may be nonsilicon. There is a great deal of R&D under way in a number of technologies; space constraints permit only a glimpse at these in this paper.

7.1 Gallium Arsenide Detectors

Gallium arsenide has long been of interest for high-speed electronics and sensors, due to its high electron mobility (Fig. 16). In addition, the excellent radiation resistance of some heterojunction electronics devices based on gallium arsenide (see Sec. 7.3) has prompted research into its possible use as a detector medium in high-radiation environments. The essential concerns to date have been the lack of technological maturity by comparison with silicon devices, and the slow progress in overcoming these difficult problems.

The most basic material characteristics (high density, high Z, and high fragility), while advantageous for some applications such as X-ray detectors, are all going in the wrong direction for high-precision MIP tracking detectors, particularly vertex detectors. Nevertheless, the potential for high radiation tolerance is a major attraction.

The difficulties begin with the production of detector-grade material. The impressive work going on in this very complex area has been summarized in two excellent recent papers [93, 94]. Three methods of crystal growth and three methods of epitaxial layer deposition have been tried; of these, only one (liquid encapsulation, LEC) has yielded detector-grade material. Even here, resistivities are at present limited to around 100 Ω cm and electron lifetimes to around 10 ns.

The idea of using GaAs for high-speed (GHz) CCD's has great attractions [95], and considerable progress with test devices has been made. This work illustrates the need to extend basic designs with respect to those used with silicon. "Standard" capacitive gates imply processing complications that can be overcome by a resistive gate technology. This, however, gives large leakage current, which can in turn be overcome with a heterostructure design. The overall picture is one of considerable problems but enormous promise.

The use of pixellated GaAs detectors for hard X-rays, bump-bonded to silicon readout IC's, is being pursued by the Leicester University X-ray astronomy group [96, 97].

For tracking detectors in high-radiation environments, possibly including the most challenging vertex detector region, the RD-8 Collaboration at CERN is doing pioneering work [98]. MIP signals are not yet adequate for high-efficiency trackers, but progress in the quality of the starting crystals should improve that. For the present, compensated material (using iron or chromium doping) is used to achieve acceptable depletion depths. Reasonable resistance to neutron irradiation has been observed, but there are recent concerns (unpublished) as to the hardness with respect to protons. There is also the concern that as the carrier lifetime is increased as a result of improved crystal quality, the radiation tolerance may be correspondingly degraded. There is (to my knowledge) nothing to suggest that "detector-grade" GaAs (comparable in its properties to detector-grade silicon) would necessarily be more radiation resistant than silicon. All studies to date relate to material which can only be compared to silicon of resistivity around 100 Ω cm at best, with leakage currents approximately 1000 times greater than those of high grade silicon.

7.2 CVD Diamond

The availability of affordable diamonds grown by the chemical vapor deposition (CVD) process has opened up an exciting possibility for extremely radiation resistant tracking detectors, well-suited to the LHC vertexing environment. A comparison of some of the important parameters with respect to silicon and gallium arsenide is as follows:

Property	Silicon	GaAs	Diamond
Mass density g. cm^{-3}	2.33	5.32	3.5
Radiation length cm	9.4	2.3	12.0
Average e-h pairs per 100 μm	8900	13000	3600
Average e-h pairs per 0.1% RL	8400	3000	4500

Being in a class of its own as regards band gap for detector materials (see Fig. 61), there is no need to create a diode structure. Simply metallizing the insulator surfaces and applying a potential difference results in collection of the generated signal (up to the limit of the electron lifetime in the material) with negligible leakage current. A review paper of the CERN RD42 Collaboration on this subject reports excellent recent progress [99]. The method of crystal growth results in a defect density which diminishes as the thickness is increased (see Fig. 62). Carrier lifetimes have recently increased to the point that collection distances of 100 μm (adequate for an efficient MIP detector) have been achieved (Fig. 63). These properties have been stable with irradiation up to pion fluences of $6 \times 10^{13}\,cm^{-2}$. Of course, for the most challenging vertexing applications, they still need to be checked up to $10^{15}\,cm^{-2}$. Leakage currents are not a problem at any radiation dose.

This technology does appear to offer real hope for a reasonably low-mass detector sitting at the minimal radius ($\approx$ 4 cm) in an LHC experiment for a ten-year lifetime. Due to the track density, it would certainly need to be pixel-based (or very short strips!) so presumably, one is contemplating bump-bonding to appropriately robust electronics. This is the topic of the next section.

7.3 Local Electronics

For the high-radiation vertex detector environments where silicon-based IC's are (probably) ruled out, we are almost certainly in the world of pixels. The basic requirements for the front-end IC's include fast shaping times, low noise at low power, and excellent radiation hardness. The high electron mobility transistor

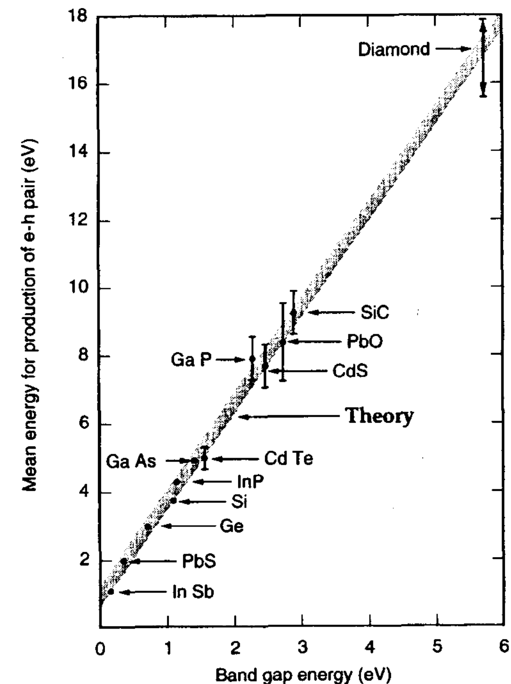


Fig. 61. Band gap and pair-creation energy, for various detector materials.

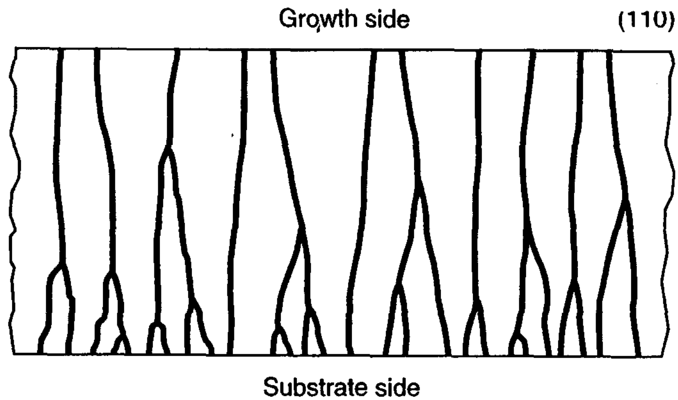


Fig. 62. Evolution of crystalline defects in CVD diamond as a function of thickness of the deposited layer.

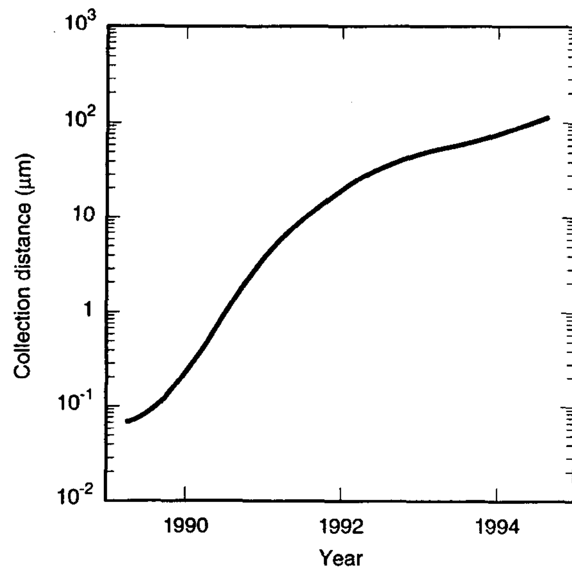


Fig. 63. Time development of collection distance in CVD diamond, from Ref. [99].

(HEMT) based on heterojunctions between different III-IV compounds offers considerable hope of satisfying these requirements. For a recent review paper, see Ref. [100]. The extraordinary radiation hardness of these devices, and indeed their availability as highly engineered structures, stems from the fact that electrons are transported in extremely thin layers (e.g., 10 nm thickness in the typical GaAs/AlGaAs heterostructure). Bulk damage effects are much less severe in such regions of high current density. The gain of both *n*- and *p*-type C-HFET's is stable after irradiation by 100 Mrad gammas and 10^{15} n / cm^2 (Ref. [101]), and these structures readily lend themselves to integrated electronics design (amplifiers, comparators, etc.) as required for APS readout electronics. The prospect of CVD diamond detectors bump-bonded to such readout IC's looks extremely promising. One is, however, still a long way short of demonstrating the LHC functionality at a reasonable pixel size and power dissipation. But there are no seemingly insurmountable obstacles in view.

8 Conclusions

Vertex detectors used in experiments up to the present time come in essentially two varieties, those providing one-dimensional information (microstrip detectors) and those being pixel-based and providing two-dimensional information (charge-coupled devices). The latter, though preferable in principle for several reasons, including superior track reconstruction capability, have restricted applicability in the HEP environment. Both of these detector types found their birthplace in the ACCMOR Collaboration in the early '80s, where they performed with unprecedented precision for charm reconstruction in a fixed-target experiment.

In the move to the collider environment, we experienced, in one sense, a step backward. Due to large beam pipes dictated by background levels at small radius, lower track momenta, and other factors, the enormous effort has been repaid by high-quality b tagging, but only limited charm capability. Fortunately for us, the physics rewards for these restricted technical achievements have been considerable, crowned recently by the discovery of top. The strength of the CDF analysis gained enormously from the b-tagging capability in that experiment.

For the future (B factories, LHC, and the e^+e^- linear collider, among others), the challenges will be still greater. Backgrounds and track densities in the event will in general increase at small radius, due mainly to the higher CM energies giving greater track multiplicities and to the increased luminosity needed to achieve the physics goals. Silicon microstrips, while of increasing value for general tracking, will tend to be pushed out of the small radius region where conditions are too hostile. Regarding the energy frontier (LHC and the future e^+e^- LC), we can expect to see a separation between the vertexing technologies.

For the LHC, one is looking for pixel-based detectors with high timing resolution and phenomenal radiation resistance. This probably leads to the realm beyond silicon, most probably hybrid detectors using GaAs or (more probably) CVD diamond, and hardened silicon or (more probably) heterojunction IC's for the front-end electronics. Some flexibility is gained by the general acceptance of the fact (demonstrated ten years ago in CCD detector systems) that the operating temperature should be considered a tunable parameter. By appropriate mechanical design, it is possible to make very low mass structures of micron-scale mechanical stability that can be repeatedly cycled between room temperature and the optimal cryogenic operating temperature. What is most important, as the overall LHC detectors enter their construction phase, is to preserve adequate funding for the R&D needed to surmount the great challenges associated with vertex detectors in that environment. R&D tends to be squeezed out under pressure of large construction projects, and it is important to remember that the LHC vertex detectors are on a significantly longer timescale than the rest of the system. The optimal detector designs may well continue to evolve through the physics life of the machine, leading to upgrade detectors on several occasions.

For the future e^+e^- Linear Collider, the picture seems to be rather clearer. The main challenge in sitting at small radius is to absorb a very high rate of background MIP hits from incoherent e^+e^- pair background. CCD detectors of unparalleled granularity have this capability, the 307 Mpixel SLD upgrade detector being a useful demonstration model. Ongoing CCD developments hold the promise of vertex detectors for this environment able to operate at $R \approx 10$ mm with space-point precision of approximately 3 μm, and thickness less than 0.2% RL per layer. This combination is unachievable with any APS system conceived to date (thickness of 1% RL per layer is a reasonable goal for such detectors) and the poorer timing information from the CCD detector is not a serious drawback in this environment, given the long beam-crossing interval (≤ 120 Hz bunch crossing frequency).

The *physics* requirements of these detectors operating at the energy frontier are, of course, difficult to define. Hopes of Higgs and SUSY particle decays via bottom provide a clear motivation. However, it is not impossible that even more exciting (i.e., unexpected) discoveries may result from clean recognition of charm jets, or indeed from clean operation in veto mode, recognizing jets which are devoid of heavy flavors. My personal inclination is to be wary of theoretical predictions and to aim to build a general purpose detector which is as powerful as possible within its measurement regime. For vertex detectors, this means aiming to see the full tree of sequential bottom and charm decays with high efficiency. History has taught us the danger of linking experiments too closely to theoretical ideas. One remembers

experiments at the CERN ISR where an intensive effort to discover the *W* boson was mounted. This search was, of course, doomed due to the machine energy being well below the *W* production threshold, but one could easily have discovered the J/ψ which was being prolifically produced in that environment, had experiments been provided with a modest two-muon detection system, rather than a highly sophisticated system geared up to single-muon detection. Such lessons have taught us that future detectors should be made as general as possible in their scope for physics discoveries. In the case of vertex detectors, achieving a good capability for identifying the heaviest long-lived quark of charge +2/3 (charm) as well as the heaviest quark of charge -1/3 (bottom) may pay unpredictable dividends for physics. In this regard, the present generation of collider vertex detectors, if given school grades, would attract comments such as "could do better," "a greater effort is needed in future," etc.

It is perhaps instructive to summarize the time development of the various types of vertex and tracking detectors with respect to some key parameters. Figure 64 shows the area coverage. Microstrip detectors have always been far ahead and seem well-placed to continue their prodigious expansion (to some tens of square meters) at LHC. CCD-based detectors may have peaked in area with the SLD upgrade. For the future LC, the smaller beam pipe leads to no greater an area coverage requirement than has currently been achieved. In this respect, smaller is better. APS systems, not yet used as vertex detectors, need to expand greatly for LHC, but the *performance* increase is a much greater challenge for them, as we have seen.

Figure 65 shows the corresponding situation as regards number of channels. At 300 Mpixels, the SLD vertex detector may have reached some sort of plateau, but the APS system for LHC will need to get close to this in order to meet the initial design specifications. This is an enormous extrapolation from where they are now.

Figure 66 shows a most important parameter, the multiple scattering term in the impact parameter resolution. Microstrip detectors have floated around the 30–60 μm region; however, this will become less significant as their role (at the energy frontier) evolves from vertex detector to general purpose tracker. The APS detector that will fill the vertexing hole at LHC aims for precision at the high end of this

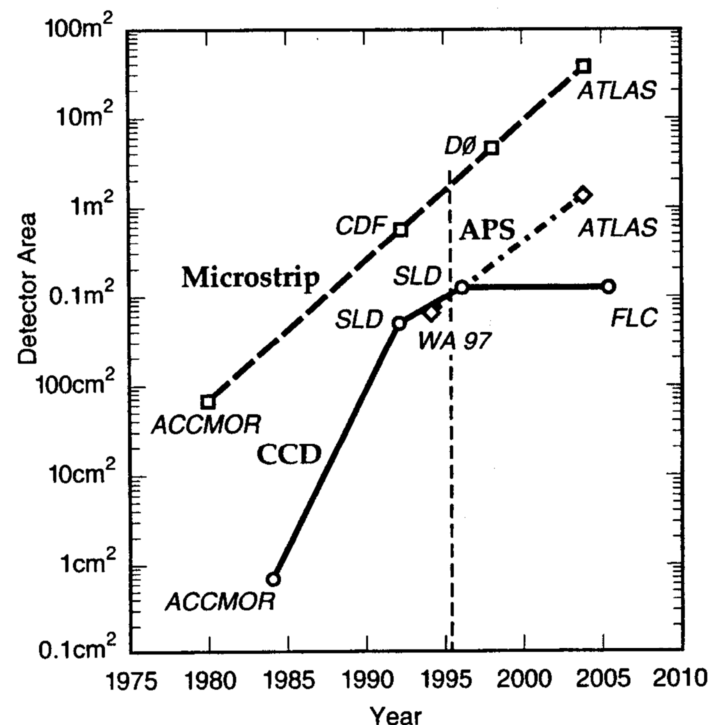


Fig. 64. Time development of area coverage of the leading-edge vertex and tracking detectors according to the main technologies (microstrips and CCD's). The APS is expected to enter the realm of vertex detectors in the LHC environment.

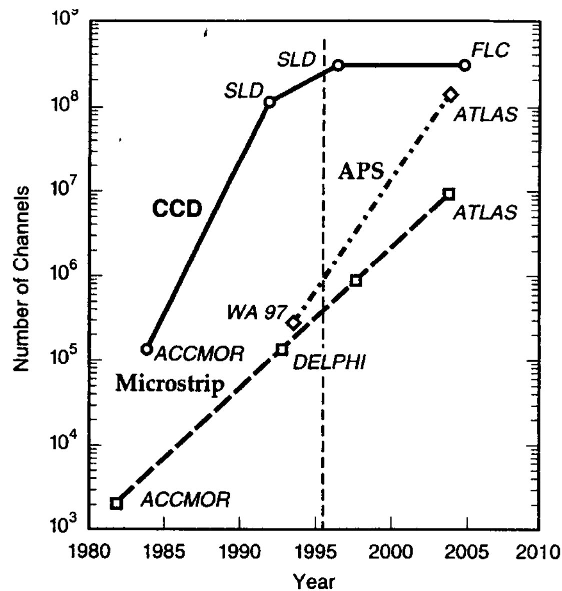


Fig. 65. As Fig. 64, but showing the number of channels in leading-edge detectors as a function of time.

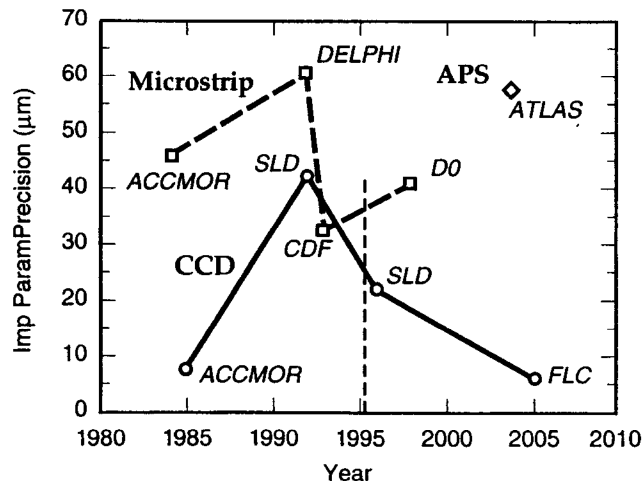


Fig. 66. As Fig. 64, but showing the multiple scattering term in the impact parameter precision as a function of time.

range, due to the fact that nobody yet dreams of going below a radius of 4 cm, and the detectors are intrinsically rather thick. CCD detectors started with marvellous performance in ACCMOR (resulting in some very clean charm physics), degraded badly in the collider environment, are gaining ground with the SLD upgrade detector, and hold promise of their original phenomenal performance (20 years later) at the future LC. The physics rewards on this second round of topologically excellent vertexing could (we hope) be enormous. Incidentally, the ongoing importance of this parameter stems from the increasing particle multiplicities in the events of interest. Despite the increased CM energies, the impact parameter precision for tracks in the 1 to 10 GeV range remains crucial for topological vertexing in the TeV collider regime.

Aside from their applications in particle physics, it is important to remember the very strong interdisciplinary aspects of these detectors. Their use in X-ray detection systems in pure and applied science is enormous, particularly for the pixel-based devices, since the ability to record an image is of rather general interest. Even if the highest aims for vertex detectors are slow in coming (sometimes because of the timescales of the new accelerators), the R&D is proving of great benefit to other areas.

Regarding the specific application to vertex detectors, there is an ongoing need for new ideas. Mostly these will come from young people. I would like to conclude these lectures with a special note of encouragement to these participants. If you get an idea, do not be put off by "the experts." I once attracted a considerable amount of negative expert comment (when I started to push CCD's for vertex detectors in 1980). The established community of experts on silicon radiation detectors was generally extremely skeptical. There were a few exceptions, such as Veljko Radeka and Emilio Gatti, who gave me greatly needed encouragement to carry on. So, if you get an idea, I advise you to pursue it and see where it leads without being too concerned as to the comments of critical bystanders. There is an ancient Chinese proverb that the one who thinks something to be impossible should not interrupt the one who is trying to do it. It would be better for science if some of us middle-aged physicists did more to remember this! I am sure there are wonderful ideas for novel

vertex detectors that nobody has yet thought of, and that some of the participants in this Institute may well discover them.

Acknowledgments

Having worked for 20 years in this field, I am conscious of the vast number of colleagues and collaborators from whom I have learned almost everything I know. In order to make a manageable list, I shall restrict my acknowledgments to those who have contributed most to my ability to prepare these lectures. These include David Burt, David Dorfan, Erik Heijne, Andrew Holland, Gerhard Lutz, Sherwood Parker, Peter Pool, Veljko Radeka, Hartmut Sadrozinski, Paul Seller, Steve Watts, and Peter Weilhammer. I would also like to thank Jacqueline Croft for the careful preparation of this report in record time, John Proch for excellent help with the figures, and David Sankey for digging me out of several PostScript-related traps. Finally, I would like to thank my wife Joan for her great support, and patience regarding the many lost weekends that went into the preparation of the lectures and this written report.

References

1. S. E. Derenzo *et al.*, LBL-1791 (1973).
2. G. Gilder, *Microcosm: The Quantum Revolution in Economics and Technology* (Touchstone, 1989).
3. G. W. Fraser *et al.*, Nucl. Instrum. Methods A **350**, 368 (1994).
4. P. Lechner and L. Strüder, Nucl. Instrum. Methods A **354**, 464 (1995).
5. H. Bichsel, Rev. Mod. Physics **60**, 663 (1988).
6. S. M. Sze, *Physics of Semiconductor Devices* (John Wiley, 1981).
7. W. Shockley, *Electrons and Holes in Semiconductors* (Van Nostrand, 1950).
8. G. K. McKay, Phys. Rev. **84**, 829 (1951).
9. J. B. A. England *et al.*, Nucl. Instrum. Methods **185**, 43 (1981).
10. J. Kemmer, Nucl. Instrum. Methods **169**, 499 (1980).
11. R. Hofmann *et al.*, Nucl. Instrum. Methods **225**, 601 (1984).
12. J. T. Walker *et al.*, Nucl. Instrum. Methods **226**, 200 (1984).
13. P. Holl *et al.*, IEEE Trans. Nucl. Sci., **NS-36**, 251 (1988).
14. P. Allport *et al.*, Nucl. Instrum. Methods A **310**, 155 (1991).
15. N. Bachetta *et al.*, Nucl. Instrum. Methods A **342**, 39 (1994).
16. T. I. Westgaard *et al.*, paper contributed to the 7th European Symposium on Semiconductor Detectors, 1995 (to be published).
17. B. S. Avset *et al.*, IEEE Trans. Nucl. Sci., **NS-37**, 1153 (1990).
18. T. Ohsugi *et al.*, Nucl. Instrum. Methods A **342**, 22 (1994).
19. *Hot electron analyzer, PHEMOS-50*, Hamamatsu Photonics.
20. G. Lutz, private communication.
21. ATLAS and CMS Silicon Central Trackers, technical proposals CERN/LHCC/94-43 (ATLAS) and CERN/LHCC/94-38 (CMS).
22. G. Batignani *et al.*, Nucl. Instrum. Methods A **236**, 183 (1993).
23. P. Weilhammer, Nucl. Instrum. Methods A **342**, 1 (1994).
24. R. Brenner, *Proc. 3rd International Workshop on Vertex Detectors*, Indiana University Report IUHEE-95-1 (1995).
25. E. Gross, *Proc. 3rd International Workshop on Vertex Detectors*, Indiana University Report IUHEE-95-1 (1995).
26. E. Belau *et al.*, Nucl. Instrum. Methods **214**, 253 (1983).
27. B. Hyams *et al.*, Nucl. Instrum. Methods **205**, 99 (1983).
28. L. Hubbeling *et al.*, Nucl. Instrum. Methods A **310**, 197 (1991).
29. H. Tajima *et al.*, Nucl. Instrum. Methods A **310**, 504 (1991).
30. H. Hanai *et al.*, Nucl. Instrum. Methods A **314**, 455 (1992).
31. A. S. Schwarz, Physics Reports **238**, 1 (1994).
32. P. Seller *et al.*, IEEE Trans. Nucl. Sci., **NS-35**, 176 (1988).

33. S. A. Kleinfelder *et al.*, IEEE Trans. Nucl. Sci., **NS-35**, 171 (1989).

34. E. Beauville *et al.*, Nucl. Instrum. Methods A **288**, 157 (1990).

35. D. E. Dorfan, Nucl. Instrum. Methods A **342**, 143 (1994).

36. Z. Y. Chang and W. Sansen, Nucl. Instrum. Methods A **305**, 553 (1991).

37. S. Gadomski *et al.*, Nucl. Instrum. Methods A **320**, 217 (1992).

38. C. Da Via *et al.*, Nucl. Instrum. Methods A **344**, 199 (1994).

39. E. Barberis *et al.*, UC Santa Cruz Preprint, SCIPP 94/29 (1994).

40. J. DeWitt *et al.*, UC Santa Cruz Preprint, SCIPP 94/34 (1994).

41. F. Arfelli *et al.*, paper contributed to the 7th European Symposium on Semiconductor Detectors, 1995 (to be published).

42. W. S. Boyle and G. E. Smith, Bell Syst. Tech. Journal **49**, 587 (1970).

43. R. H. Walden *et al.*, Bell Syst. Tech. Journal **51**, 1635 (1972).

44. C. J. S. Damerell *et al.*, Nucl. Instrum. Methods **185**, 33 (1981).

45. *Charge-Coupled Devices and Systems*, edited by M. J. Howes and D. V. Morgan (Wiley, 1979).

46. J. D. E. Beynon and D. R. Lamb, *Charge-Coupled Devices and Their Applications* (McGraw-Hill, 1982).

47. L. Strüder *et al.*, Nucl. Instrum. Methods A **253**, 386 (1987).

48. M. Kleefstra, *Solid State Electronics* **21**, 1005 (1978).

49. A. D. Holland, paper contributed to the 7th European Symposium on Semiconductor Detectors, 1995 (to be published).

50. H. Soltau, paper contributed to the 7th European Symposium on Semiconductor Detectors, 1995 (to be published).

51. C. J. S. Damerell, summary paper contributed to the 7th European Symposium on Semiconductor Detectors, 1995 (to be published).

52. E. R. Fossum, in *Proceedings of the SPIE International Conference* **2172**, 38 (1994).

53. A. M. Mohsen and M. F. Tompsett, IEEE Trans. Electronic Devices, **ED-21**, 701 (1974).

54. E .K. Bangart *et al.*, IEEE Trans. Electronic Devices, **ED-38**, 1162 (1991).

55. J. Frenkel, Phys. Rev. **54**, 647 (1938).

56. D. J. Burt, Nucl. Instrum. Methods A **305**, 564 (1991).

57. P. Centen, IEEE Trans. Electronic Devices, **ED-38**, 1206 (1991).

58. R. W. Brodersen and S. P. Emmons, IEEE Trans. Electronic Devices, **ED-23**, 215 (1976).

59. D. Burt, private communication.

60. M. H. White *et al.*, IEEE Journal of Solid State Circuits, **SC-9**, 1 (1974).

61. G. R. Hopkinson and D. H. Lumb, J. Phys. E. **15**, 1214 (1982).

62. K. H. Schmidt *et al.*, MPI-PhE/94-30 (1994).

63. SLD Vertex Detector Upgrade Group, SLAC-PUB-95-6950 (1995).

64. C. J. S. Damerell and D. Jackson, paper submitted to the Morioka Workshop on Physics at Future Linear e^+e^- Colliders, 1995 (to be published).

65. E. R. Fossum, transparencies of lecture given at the Univ. of Waterloo Pixel Device Conference, 1993.

66. S. K. Mendis *et al.*, *Proc. SPIE* **2172**, 19 (1994).

67. F. Antinori *et al.*, CERN DRDC/94-51 (1995).

68. S. Holland, Nucl. Instrum. Methods A **275**, 537 (1989).

69. C. J. Kenney *et al.*, Nucl. Instrum. Methods A **342**, 59 (1994).

70. S. Parker, private communication.

71 P. Middelkamp, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

72. G. Cesura *et al.*, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

73. V. A. J. van Lint *et al.*, *Mechanics of Radiation Effects in Electronic Materials* (Wiley, 1980).

74. G. C. Messenger and M. S. Ash, *The Effects of Radiation on Electronic Systems* (Van Nostrand Reinhold, 1986).

75. W. R. Dawes, Nucl. Instrum. Methods A **288**, 54 (1990).

76. V. A. J. van Lint, IEEE Trans. Nucl. Sci., **NS-41**, 2642 (1994).

77. E. E. Conrad, IEEE Trans. Nucl. Sci., **NS-41**, 2648 (1994).

78. J. R. Adams *et al.,* IEEE Trans. Nucl. Sci., **NS-24**, 2099 (1977).

79. W. Füssel, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

80. E. A. Burke, IEEE Trans. Nucl. Sci., **NS-33**, 1276 (1986).

81. G. P. Summers *et al.,* IEEE Trans. Nucl. Sci., **NS-33**, 1282 (1986).

82. H. J. Ziock *et al.,* Nucl. Instrum. Methods A **342**, 96 (1994).

83. E. Fretwurst *et al.,* Nucl. Instrum. Methods A **342**, 119 (1994).

84. F. Lemeilleur *et al.,* CERN-ECP/94-8 (1994).

85. Three papers contributed by G. Lutz, Z. Li, and S. Watts to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

86. J. Matheson *et al.,* CERN report RD20/TN/95/43.

87. M. Huhtinen and P. A. Aarnio, Nucl. Instrum. Methods **A 335**, 580 (1993).

88. J. M. Killiany, "Topics in applied physics," **38**, 147 (1980).

89. N. S. Saks, IEEE Trans. Nucl. Sci., **NS-24**, 2153 (1977).

90. M. S. Robbins, Ph.D. Thesis, Brunel University (1992) and RADECS 91, *IEEE Proceedings*, 368 (1992).

91. A. Holland *et al.,* IEEE Trans. Nucl. Sci., **NS-38**, 1663 (1991).

92. S. Watts *et al.,* ESTEC Report BRUCRD-ESA CCD-95-IR (1995).

93. M Schieber, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

94. E. Bauser, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

95. R. E. Colbeth *et al.,* in *Proceedings of the SPIE International Conference* **1071**, 108 (1989).

96. A. D. Holland *et al.,* Nucl. Instrum. Methods A **346**, 366 (1994).

97. A. D. Holland *et al.,* Leicester University preprint XRA 94/15 (1994).

98. RD-8 Collaboration Report, CERN/DRDC 94-32 (1994).

99. K. T. Knöpfle, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

100. G. Bertuccio *et al.,* paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).

101. W. Karpinski, paper contributed to the Seventh European Symposium on Semiconductor Detectors, 1995 (to be published).