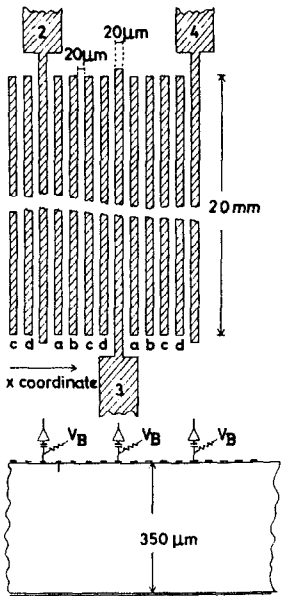

Figure 28 From Reference 26. Layout of strips on one face of the detector, with one strip in 5 connected to the bias voltage and to the readout system via a locally connected preamplifier.

One limitation with the interpolation approach is the inevitable degradation in 2-track resolution, which gives rise to a major loss in precision of pairs of tracks which are merged at the level of the granularity of the readout strips (even though their charge distributions may be well separated on the full set of strips on the detector). This problem is exacerbated by the fact that the number of tracks giving cluster sizes > 2 goes well beyond the expectations based on δ-electron production. In fact, with the 1 in 3 readout, 30% of tracks produce these fat clusters, which can be understood on the basis of capacitances between the strip holding the signal charge and non-neighbouring readout strips.

In order to take advantage of the precision of a detector in measuring impact parameters, it is important to have the first detector plane as close to the vertex as possible, for reasons of multiple scattering as well as geometry. Thus the 2-track resolution is in practice as important, in evaluating detector performance, as the precision. For this reason, the trend in microstrip detectors is to go to readout of every strip. The reason for not doing this in the first place is that with conventional electronics it is just impossible. As an example, consider the microstrip detectors of the NA32 experiment. To cover 24 x 36 mm with detectors equipped with 1 in 3 readout in the central $\frac{1}{3}$ of the detector, and 1 in 6 readout in the outer $\frac{2}{3}$ of the detector, involves fanouts, preamplifiers and connectors which extend outwards over a radius of 60 cm and a thickness of about 2 cm, which sets limits on the stacking density of the detectors. So the ratio of areas of the local readout to the detector is greater than 1000. Connecting readout to every strip would increase this ratio to over 5000. On top of this, one has to make room for the tonnage of cabling which leads to the racks of remote readout electronics.

In spite of these complications, the MPI Group within the ACCMOR Collaboration have succeeded in connecting all strips to external readout in some special detectors used as elements of an active

target. These detectors needed only to have 1 mm of active width (50 strips) so that the total number of channels could be held down to a reasonable level. With these detectors, the precision of measurement improved to $\sigma = 3.0$ μm and the number of clusters more than 2 strips wide was reduced to a very low level, confirming the impression that the problems with the earlier detectors are a peculiarity of the charge division readout. The main remaining problems are:

(a) for angles $\gtrsim$ 100 mrad a rapid loss of precision for the reasons discussed in Section 3, which are common to all detectors of the thickness needed for microstrip readout; and

(b) the density of off-chip (but necessarily local) electronics.

The solution to (b) is under development by two groups. B.Hyams, S.Parker, T.Walker and others at CERN, Hawaii and Stanford are developing a special readout integrated circuit which they call the microplex chip;[28] and G.Lutz, G.Zimner and others at MPI, Munich, and University of Dortmund are working on a similar chip in CMOS rather than nMOS technology.[29] In both cases, the aims are similar, namely:

(a) to avoid fanouts from the detectors. The electronics is fabricated to provide preamp inputs on a pitch of about 50 μm so that by connecting alternate strips at each end of the detector one can deal with all strips with a pitch of 25 μm or greater.

(b) to provide analogue storage for 128 signals on a chip, and multiplexed analogue outputs, so that the number of output cables is reduced by a factor of 100.

Figure 29 shows the general layout of the microplex chip. The chip would butt against the detector at its left-hand edge. The staggering of input pads (IP) is necessary because of the lack of wire bonders capable of bonding at 50 μm pitch. The method of making the connections from the detector to the readout chip is far from simple. Figure 30 shows one option which has been demonstrated to be workable in practice. The quartz rods are needed to maintain

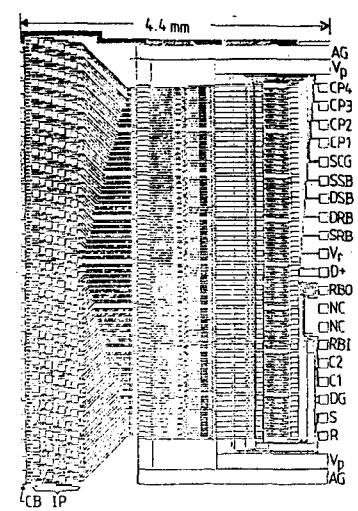

Figure 29 From Reference 28. Microplex chip layout showing the input pads (IP) on a pitch of 50 μm.

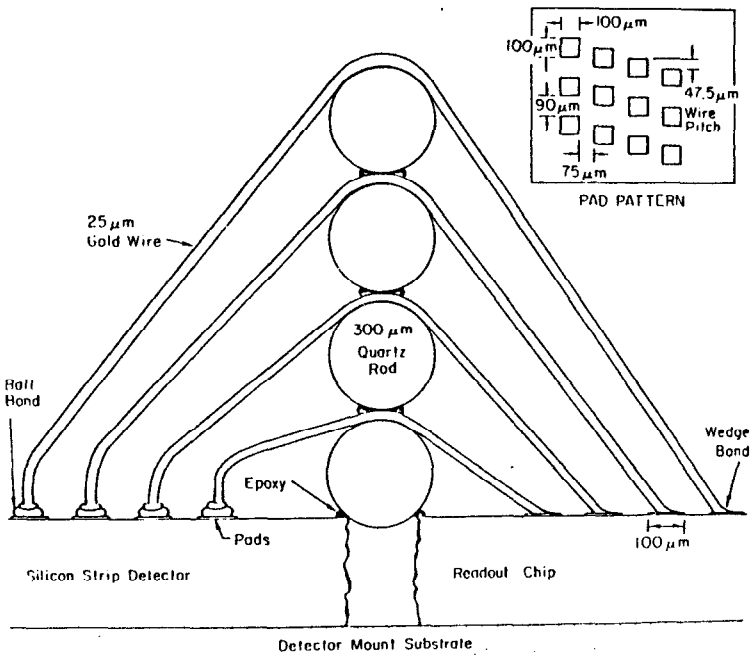

Figure 30 From Reference 28. Method of wire bonding the strip detector to the readout chip.

electrical isolation between the different layers of bond wires. Possible solutions in future include various schemes for bump bonding, and the integration of the readout on the same chip as the detector.[30]

Figure 31 is a block diagram showing the main elements of the microplex chip. The signals are fed through a charge sensitive preamlifier of gain 500 onto a storage capacitor. In collider operation, the storage capacitors would be reset between beam

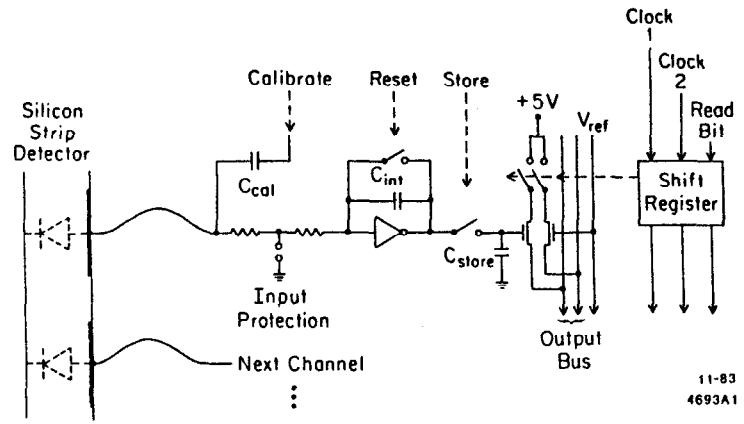

Figure 31 From Reference 28. Block diagram of readout electronics contained on the microplex chip.

crossings, and in fixed target applications they are reset frequently during the spill so that the data are adequately free of out-of-time track information. The stored information is read out at ~ 1 MHz (~ 150 μs for readout of the complete chip). The analogue signal from each capacitor in turn is connected to a line driver for remote digitization.

One of the problems with this very elegant concept is to achieve low enough noise levels with such a high density of electronics. In fact preliminary noise measurements on the first microplex chips show a level of around 2500 electrons RMS which (at 10% of the min-I signal from a 300 μm thick detector) is already very adequate.

5. CHARGE-COUPLED DEVICES

The charge-coupled device (CCD) was invented in 1970 at Bell Labs[31] and has been developed in two general directions. Linear CCD arrays are used for optical imaging, for analogue signal storage, as delay lines, and in a large variety of signal processing applications. Imaging 2-dimensional CCD's take advantage of the high performance of silicon for photon detection which we discussed in Section 3. These CCD's are used increasingly in TV cameras, for night vision systems (and general surveillance), in astronomy (visible and X-ray, where they have many advantages over photographic plates) and in hybrid form as infra-red detectors. In this last application, the primary photon detector is a narrow band-gap material, and the charge collected is transferred to a CCD via an array of bump bonds.

It is with the 2-dimensional area arrays that we have to deal in the context of particle detection. These consist in general of a fine matrix of potential wells just below the surface of the silicon, typically 20 μm x 20 μm x 10 μm deep. Each well constitutes an element of a picture in normal imaging applications and is referred to as a pixel. The CCD has in addition some surface structure which allows charges from each pixel to be transported and deposited in turn onto the output node of the detector. These signals are sensed by on-chip circuitry in order to minimise noise. Let us examine in some detail, with the aid of the general discussion of Section 3, how such a detector can be built. For more detailed information, there are some excellent books on CCD's[32,33] as well as CCD Conference proceedings and hundreds of published papers.

5.1 Structure and Operation of 2-d CCD's. 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 (see Figure 32(a) to (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 an np junction by the introduction of a shallow (~ 1 μm) implant of n-type dopant. The surface is oxidised to make an insulating layer and on top of this is deposited a thin conducting layer. The simplest would be aluminium, 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 Figures 32(d) to (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 np 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 Figure 32(g). For $V_c = 150$ V, such a device when

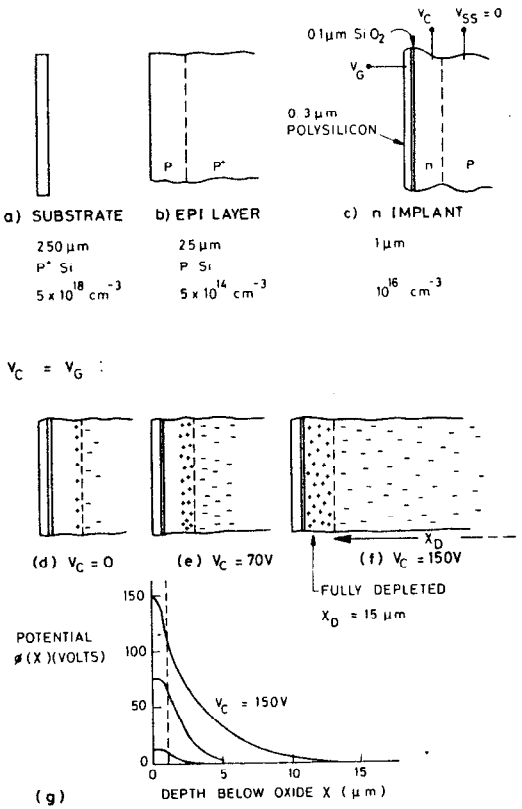

Figure 32 (a) to (c) Show (with increasing magnification for each stage) the successive stages in making a CCD-like structure.
(d) to (f) The depletion process which would apply if V_C and V_G were increased together.
(g) The corresponding potential distributions as a function of depth in the silicon.

traversed by particles would transport the generated electrons to the surface (Si/SiO_2 interface) and dump the holes into the undepleted substrate.

Now (Figure 33(a) and (b)) consider what happens if V_c is increased from 0 while V_G is held at 0 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 np junction proceeds as before, but the voltage across the oxide induces an increasing positive charge, starting from the Si/SiO_2 surface and growing into the body of the n-channel. At a very low value of V_c (about 8 volts) these depletion regions meet, causing the phenomenon known as pinch-off. The corresponding value of V_c 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 Figure 33(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 np 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

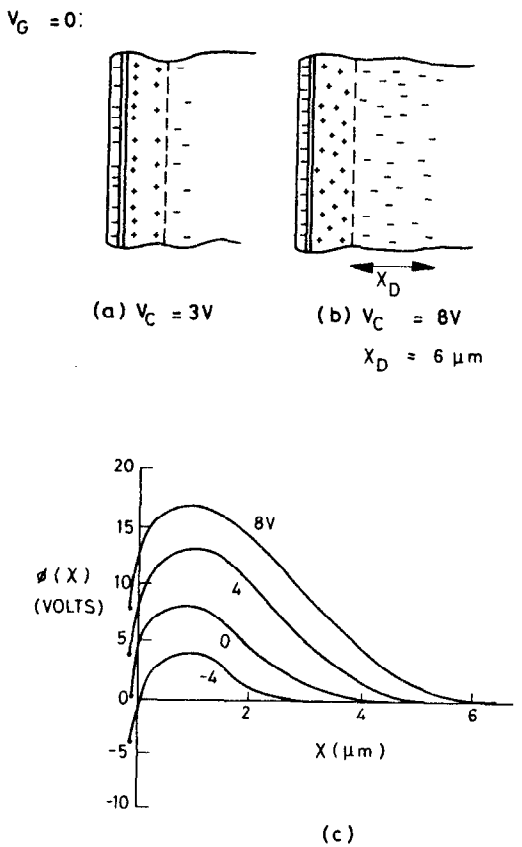

Figure 33 (a) and (b) The depletion process in normally biased CCD operation with V_G negative with respect to V_C.
(c) The corresponding potential distributions after channel pinch-off for various values of V_G.

vital ingredient in the design of CCD's for our application. Tiny charges (< 10 electrons) 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 Figure 33(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 Section 3; see for example Reference 33 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 Figure 34. Firstly, at the required pixel granularity (say 20 μm) p^+ implants are introduced of approximately 1 μm width

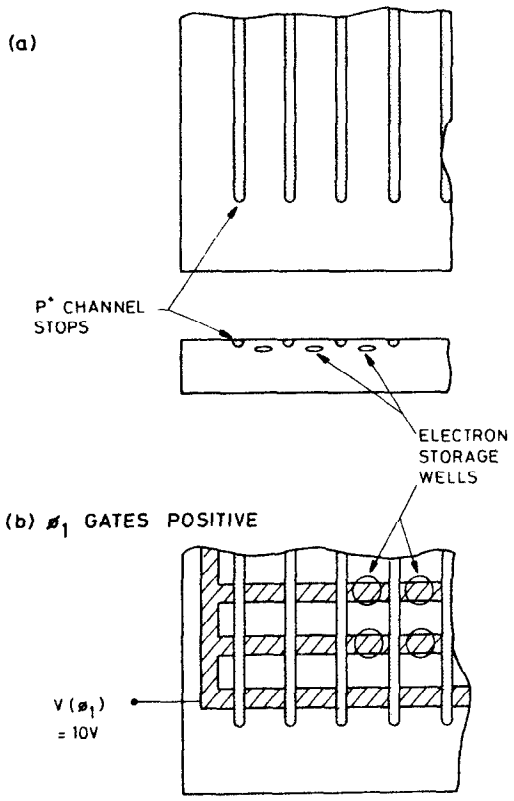

Figure 34 Establishing the potential well structure:
(a) Channel stops create potential barriers running vertically on the device.
(b) Gates create horizontal potential barriers. The combined result is a matrix of localised wells, each of which constitutes a pixel.

and 1 μm depth. These become fully depleted as part of the overall biasing of the CCD, and so provide strips of intense 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 Figure 34(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 Figure 33(c) and Figure 34(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 approach 10^6 in number on a typical CCD (400 channel stops x 600 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 (Figure 35). We arrange these gates in triplets (ϕ_1, ϕ_2, ϕ_3) in this so-called 3-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 electons 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

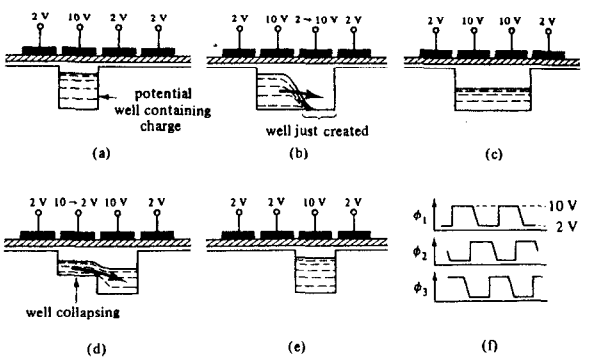

Figure 35 From Reference 32.
(a) to (e) Movement of potential well and associated charge packet by clocking of gate electrode voltages.
(f) Clocking waveforms for a 3-phase CCD.

$\phi_1 + \phi_2 + \phi_3$ electrodes together constitutes one pixel, eg. 3 x 7 μm = 21 μm.

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 3-phase CCDs, there exist other varieties (4-phase, 2-phase, virtual phase, etc) but we can ignore these in an initial discussion of these detectors.

One further element in the system is needed. 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, that row can be shifted sideways so that the charge contained in each pixel is sensed in turn by an on-chip circuit.

Figure 36 shows a diagram of one corner of a CCD detector, including the general charge collection operation for the passage of a min-I particle. One pixel is shown as a shaded area, covering the height of 3 I gates (Iφ1 to 3) and bounded by two channel stops. The charge in the bottom row of pixels can be shifted first to the bottom gate of those pixels (Iφ3) and then into the output register (Rφ1). Once in the R register, these charges are transported sideways to the output node of the CCD.

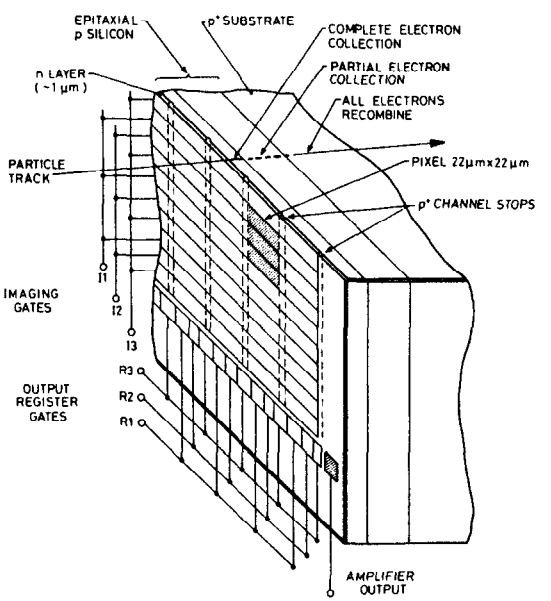

Figure 36 One corner of a CCD enlarged to show details of the pixel (storage element) structure.

Figure 37 is a photograph of the corner of a 3 phase CCD. Apart from the features noted in the caption, the elbow-shaped aluminium track to the left of the R register carries the stored charge from the R register to the gate of an on-chip FET. This gate (dark structure on the figure) can be seen sandwiched between the source and drain of the small on-chip FET, whose connections (further aluminium electrodes) can be seen disappearing off the bottom of the figure.

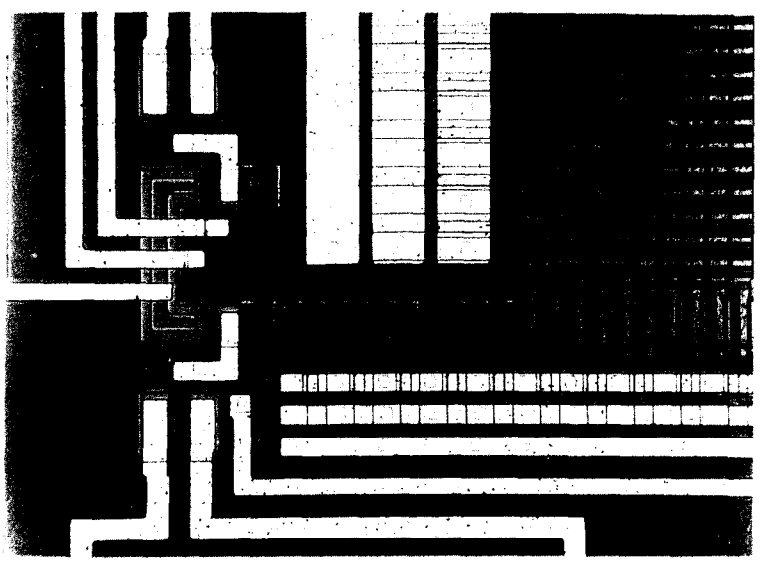

Figure 37 Courtesy of GEC, England. Photograph of one corner of a CCD showing the pixels of the imaging area (upper right quadrant), readout register (running along the bottom of the imaging area and extending 10 pixels to the left of it) and output FET (below and to the left of the readout register). The light coloured structures are aluminium tracks which carry the drive pulses to the gates, connections to the FETs, etc. The 3 broad bus-lines running vertically carry the Iϕ voltages, and the 3 narrow lines running horizontally carry the Rϕ voltages.

The CCD structure shown in Figures 36 and 37 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. As shown in Figure 36, the charge generated by a min-I particle along its track falls into 3 classes. There is a region of typically 10 μm below the surface for which the charge is within the depletion depth (see Figure 33) and is fully collected into the relevant pixel. Next, the charge from the 10 μm of undepleted epitaxial silicon (which generally has a long diffusion length) diffuses isotropically. About half of it diffuses into the depletion region and is caught in the relevant pixel or in neighbouring ones. Finally the electrons and holes generated in the p^+ substrate recombine very readily. The "effective thickness" of the CCD for particle detection is thus approximately 15 μm.

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

The rate capability of the driving electronics can be made quite high. It is (for example) no problem to shift the charges down the I array at 3 MHz or across the R register at 10 MHz with extremely high charge transfer efficiency. This quantity, often abbreviated to CTE, is the efficiency with which a bucket of charge is transferred from one pixel to the next. In buried channel CCD's, the charge transfer inefficiency may be as low as 10^{-5}. What has given CCD's their reputation of being very slow detectors has been the time required to sample the signal in order to achieve a sufficiently low noise level. This point will now be considered in some detail.

5.2 Readout Electronics. As shown in Figure 38, the charge from the pixel at the end of R register can be transferred onto the gate of an on-chip FET which is normally operated as a source follower. Its output is connected to a local preamplifier which senses the voltage change induced by the charge on the FET gate. This point can also be reset to some standard voltage V_R via an on-chip reset transistor, and this is conventionally done between each R transfer.

For high signal levels (eg. TV imaging) there is very little to add to this description. The outgoing voltage level provides a direct measurement of the stored charge (ie. brightness) of that part of the image, which is built up line by line as the contents of

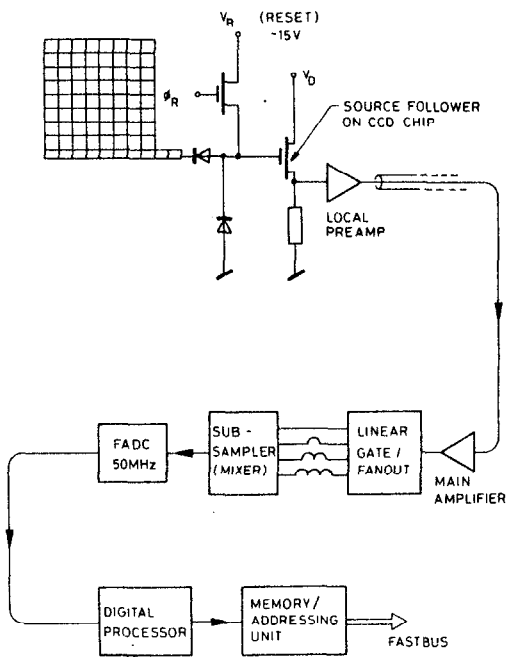

Figure 38 Block diagram of low noise signal processing, which uses a flash ADC for digital sampling of the pulse height information.

successive rows are sequentially transferred from I to R register and along the R register to the output.

Typical TV imaging involves signals near to the well capacity of the devices (several x 10^5 electrons/pixel). The sensing of very small signal levels in CCD's has been pioneered by night vision specialists and by astronomers, where total readout noise of less than 10 electrons RMS is regularly achieved. In this way images with

signals as small as 50 electrons/pixel can be reconstructed with good quality. The price paid by these special readout systems is speed. Given that in astronomical applications the images may be accumulated over periods of 20 minutes or more, a readout time of 10 seconds is completely acceptable. For particle detection in typical high energy physics experiments, a readout time of some milliseconds or at most tens of milliseconds is required. An important recent development has been the achievement of such fast readout without a serious degradation in noise.

Let us first consider the conventional method of low noise CCD readout, beyond the necessity already mentioned of adequately cooling the detector. The node is reset before each R transfer. Due to the inevitable existence of reset noise (the voltage actually set fluctuates uncontrollably by typically 200 electrons) it is next necessary to determine this precise voltage level. This is not entirely trivial for the following reason. In order to have good sensitivity (μV/pC of charge stored) the FET gate capacitance must be made small. It is typically held down to ~ 0.1 pF, but small FETs are inevitably noisy. Thus, in order to determine the output voltage to the required precision some signal averaging is necessary. Typically, astronomers use analogue integration of the output signal for a period of say 20μs. Next the R register is clocked, and the pixel charge is deposited onto the output node. A second integration is made and the voltage difference yields the charge contents of that pixel. This readout technique is generally known as correlated double sampling, and implies readout times of about 40 μs/pixel, ie. 10 s for a complete CCD.

Such a readout system was used in the first measurement of the efficiency and precision of CCDs as detectors of min-I particles.[22] Subsequently, my group addressed the problem of speeding up the readout to the point where it could be used in a real experiment, and the block diagram of Figure 38 and pulse trains of Figure 39 show how this was done.

The first (and in some ways the most difficult) step was to develop a sufficiently clean CCD drive system. Any of the external drive pulses (Iφ(1-3), Rφ(1-3) or ϕ_R, the Reset) can couple capacitively to the CCD output, inducing various types of feed-through. Given that these pulses are ~ 10 V and the min-I signal is ~ 1 mV, the feedthrough must be kept extremely small in order not to swamp the signal. This implies very clean drive pulses (sharp edges, no ringing) and careful layout of the external circuitry. The layout on the CCD itself has been very carefully designed to avoid problems in that area. Since the I gate capacitances are large (~ 10^4 pF) the drive system needs to provide these clean, relatively high voltage pulses also with high current. Fortunately, the overall readout time is dominated by the R drive system, since there are about 400 R transfers for every I transfer. The R gate capacitances are only about 1000 pF, so the drive currents are much lower than for the I gates. Eventually, a system has been developed which can drive the R register at 10 MHz with good charge transfer efficiency, as well as

being sufficiently well decoupled from the analogue output to allow sampling to commence within 20 ns of the end of the Rϕ triplet.

The local preamplifier has ~ 40 MHz bandwidth and sufficiently low noise (~ 3 nV/Hz$^{\frac{1}{2}}$) that it does not seriously increase the noise induced by the source follower (~ 10 nV/Hz$^{\frac{1}{2}}$). The signals are then fed via a remote main amplifier to a linear gate and fanout. The linear gate is opened only outside the duration of the Rϕ triplet, so

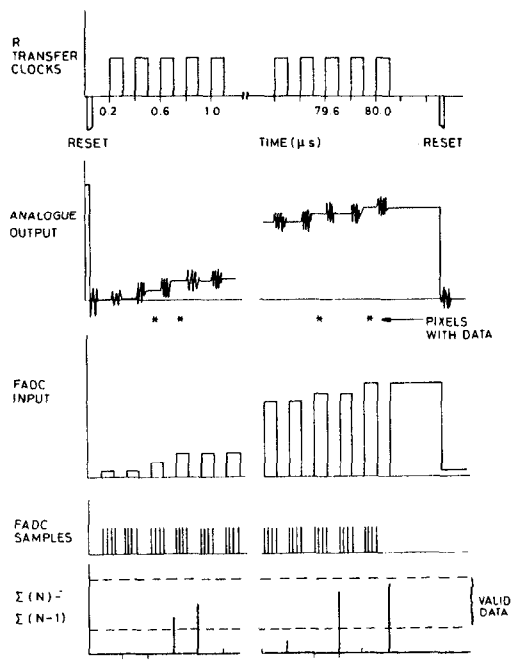

Figure 39 Sequences of signals through one row of CCD data processing. After processing a row, the output node is reset and the I gates are clocked to shift the next row into the now-empty R register. The process is repeated for each of the 600 rows of the detector.

protecting the later electronics from the large amplified spikes induced by the CCD clocking. Four outputs from the fanout are fed into an analogue mixer after relative delays of 0, 5, 10 and 15 ns. The mixer output is thus somewhat smoothed, as would be achieved by signal averaging at 200 MHz. This output is then sampled by an 8 bit flash ADC system operated at 50 MHz. By taking 4 samples over 80 ns (effectively 16 samples at 200 MHz) it is possible to achieve RMS noise levels on individual pixels of approximately 50 electrons equivalent pixel charge. While this does not match the measurement quality of the astronomers, as it obviously can not, in view of the much shorter sampling period, it is still less than 10% of the min-I signal, and so is entirely adequate for particle detection with good centroid finding. The transfer of charge and its sampling occupies in total 200 ns. The next feature of the fast readout is that instead of resetting after every pixel, as is customary in TV and astronomical imaging, we reset only once per row. Thus the analogue output builds up, with a series of steps, as shown in Figure 39. This arrangement is feasible because for particle detection we are always concerned with sparse data, which allows the contents of 400 pixels (mostly empty) to be summed on the output node without any problems of saturation or non-linearity in the on-chip or off-chip electronics. The flash ADC input (sub-sampler output) shown in Figure 39 thus consists of a series of levels, mostly equal (apart from noise fluctuations) but with occasional steps corresponding to pixels with non-zero stored charge. The digital output from the FADC sampler is fed to a digital processor which sums the samples in

groups of 4, then makes a subtraction between these sums for successive pixels. The result may be positive or negative due to noise, but in the case of a genuine min-I signal it will be positive and within a range for valid data (Figure 39). Signals from pixels having valid data (in particular satisfying a threshold requirement) are sent to a memory unit (Figure 38) which includes synchronised clocking information so that it can store the pixel address along with the digitised pulse height.

5.3 Use of CCD's for Min-I Particle Detection. In the first tests of a telescope of CCD's in a beam line,[22] it was established that they had high efficiency (98 ± 2%), good precision in x and y (~ 5 μm), and good 2-track resolution. The last feature is illustrated in Figures 40 and 41.

More recently, CCD's are being used in a charm production experiment (NA32 at the CERN SPS). By placing the detectors close to the target, the entire spectrometer aperture can be covered by $\frac{1}{3}$ of the CCD area, as shown in Figures 42 and 43. By using the fast readout system described in Section 5.2, this area can be read out within 12 ms, compatible with the readout time of the rest of the equipment in the spectrometer. However, for fixed target applications, this is not sufficient. The CCD's will be traversed by beam particles and by the products of out-of-time interactions which would build up a large background of spurious hits, making pattern recognition impossible. In the NA32 experiment, for example, the

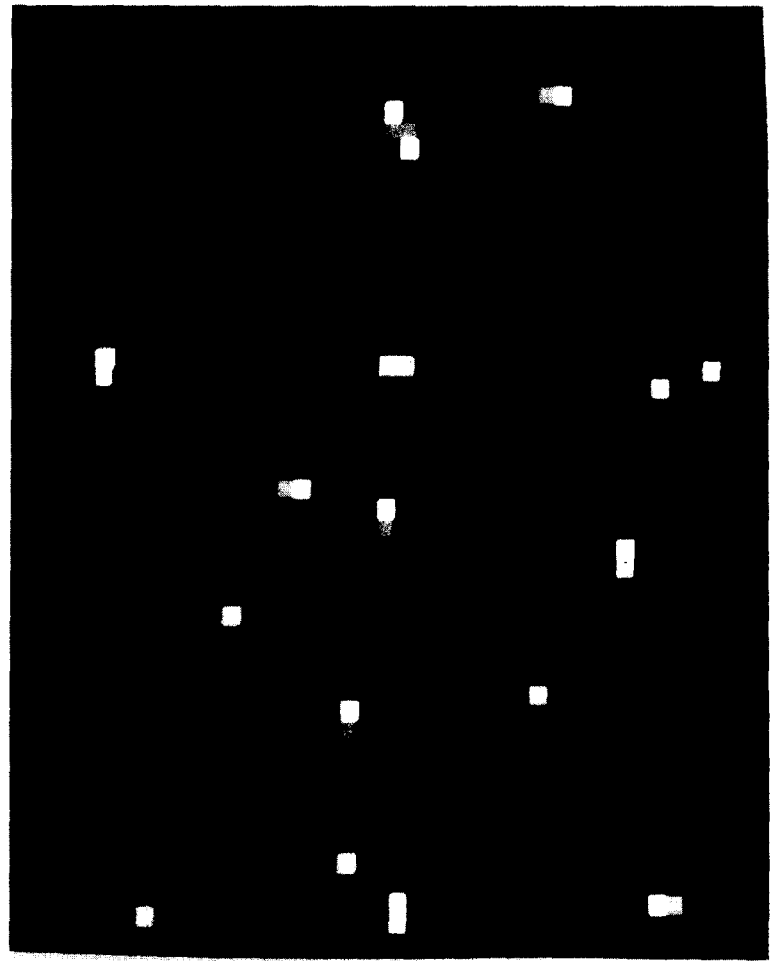

Figure 40 Zoomed online display of a high density region of tracks traversing a CCD detector. The display shows 17 beam tracks in an area of 1 square millimetre.

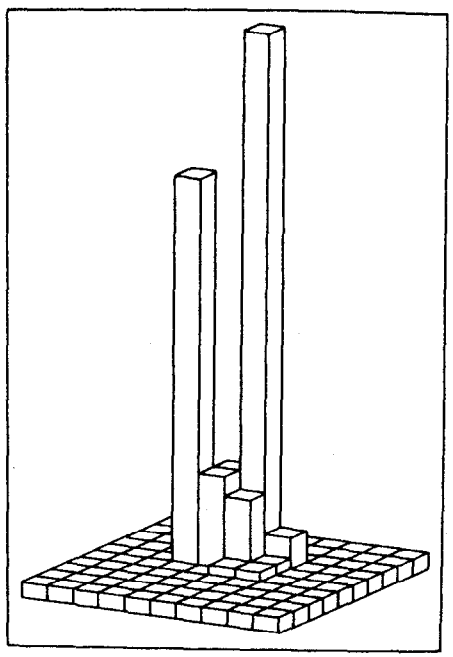

Figure 41 Isometric offline reconstruction of two particles separated by 40 μm in a CCD detector. The height of each element represents the pulse height measured in the corresponding pixel.

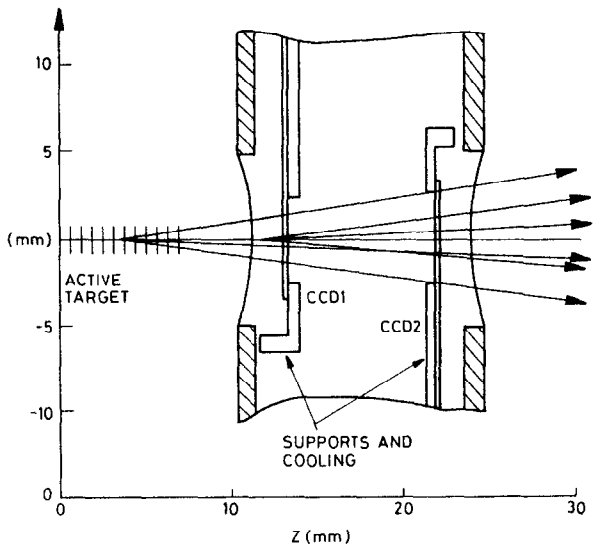

Figure 42 Layout of CCD's in experiment NA32. A pair of detectors, in a low temperature cryostat with thin vacuum windows, is located ~ 10 mm downstream of a silicon active target. The CCD's are in metal packs which provide thermal coupling to the refrigeration (cold nitrogen gas) but these packs are cut out in the regions used for particle detection.

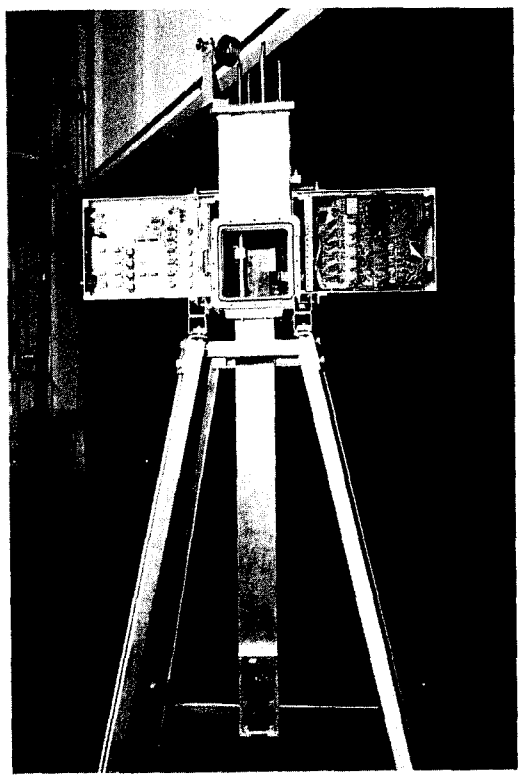

Figure 43 (a) General view of the NA32 CCD detector. The drive signals are generated on the large drive cards at the top and fed down to the CCD's on a system of low impedance strip-lines.

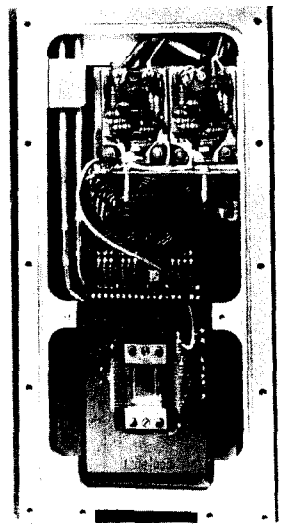

Figure 43 (b) Close-up of CCD1 at the bottom of the cryostat.

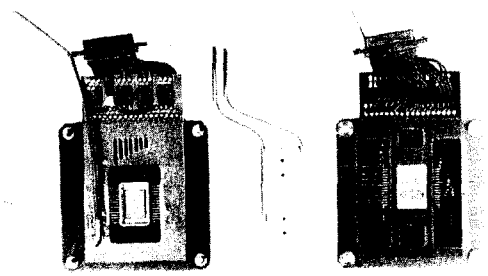

Figure 43 (c) Component elements displaced sideways; CCD1 seen from the front, cooling plate, CCD2 seen from the back.

CCD's are traversed by a beam of about 10^6 particles per second. How can this be tolerated?

The orignal idea is shown in Figure 44. By setting all the Iϕ gate voltages to be equal (say 0 V) it was thought that one would remove the potential well structure in one dimension, and the stored charges could diffuse vertically between the channel stops. By applying the gate voltages only on receipt of a trigger, the signals from early tracks would be washed out. This is in fact effective for large signals, but the charges from min-I particles are so small that they are trapped by tiny variations in the potentials below the Iϕ phases 1, 2 and 3 which inevitably arise in the manufacture of the CCD. What happens is that these gates are deposited as separate

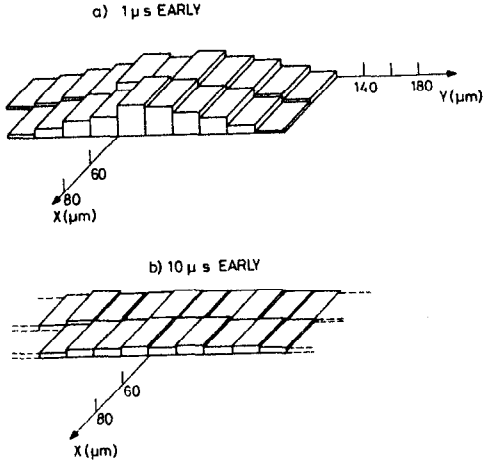

Figure 44 From Reference 34. Original idea for fast-clearing a CCD detector. This was unsuccessful, for reasons described in the text.

steps in the manufacture, inevitably under slightly different conditions. There are, therefore, small variations in (for example) the trapped charges in the oxide or at the Si/SiO_2 interface.

A second idea which has been suggested for fast-clearing CCD's is to pulse all gates negative, so destroying the potential wells, driving the stored electrons into the substrate. This special case of Figure 33 is shown in Figure 45. What happens is that as the gate goes more negative it tries to pull the surface negative, but this results in a flood of holes from the channel stops which effectively pin the Si/SiO_2 surface to 0 V, however negative the gate is driven. Thus the stored electrons remain in potential wells in depleted material and recombination will not occur.

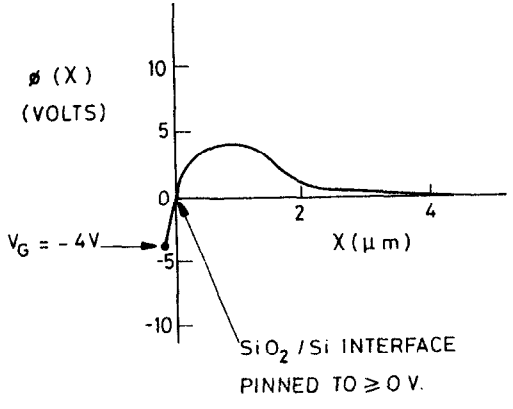

Figure 45 In order to try to clear the CCD, V_G is pulsed to -4V. This reduces but does not destroy the potential wells, and the Si/SiO_2 interface is now pinned to 0 V.

The method adopted in NA32 is illustrated in Figure 46. The beam profile is made thin vertically and wide horizontally, illuminating the regions of detector shown. The CCD's are run during the beam burst in a "drift chamber" mode. The I and R gates are clocked continuously together at about 2 MHz. The signals from beam particles traversing the CCD's are transported upwards in CCD1, downwards in CCD2, and eventually dumped at the output node. The detectors at any time have a low and perfectly acceptable density of background hits from recently arrived beam particles and interaction products. On receipt of a trigger the fast shifting continues for about 200 μs. The signals from the triggering event (shown as crosses in Figure 46) are shifted into "parking areas" at which stage the fast shifting stops. During this time, the beam is turned off with a simple kicker magnet, and is held off until the end of the

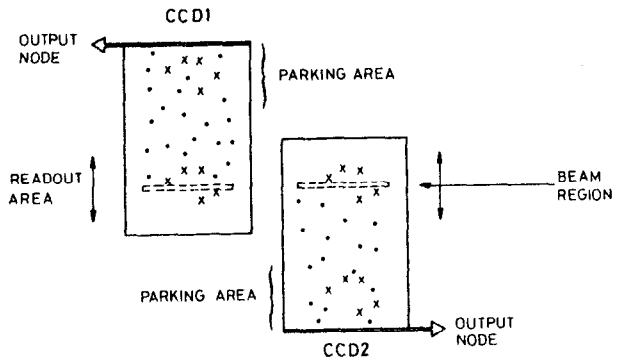

Figure 46 *Beam's eye view of the NA32 CCD's displaced sideways (as in Figure 43(c)) illustrating the "drift chamber" mode of readout.*

readout period. The data in the parking area ($\frac{1}{3}$ of the detector nearest the R register) can now be read out in background-free conditions.

This method of fast-clearing is applicable in general. Coupled with the high data storage capability of CCD's (10 spurious tracks per mm^2 would occupy only 1% of the pixels) these detectors, even without a genuine fast-clear capability, can easily be used in high rate environments.

5.4 CCD/MSD Comparison. So far, of the various ideas for silicon detectors, only MSD's and CCD's have been used for high precision particle tracking in experiments. Let us try to summarise their relative attributes.

Effective Detector Thickness:

CCD ~ 15 μm

MSD ~ 300 μm .

In both cases, these thicknesses are about at the lower limit for good signal/noise with current readout electronics. However, there is considerable scope for development. For example, CCD output circuits specially adapted for small signals are being designed with 1 electron RMS noise. CCD's are normally backed by inert silicon giving a total thickness similar to MSD's. If multiple scattering is important, they can conveniently be thinned to less than 100 μm.

Most Probable Energy Loss:

15 x 200 eV = 3.0 keV for CCD's

300 x 280 eV = 84 keV for MSD's.

Typical Readout Noise:

0.21 keV for CCD's

8.5 keV for MSD's.

Measurements:

x, y points for CCD's

x or y co-ordinates for MSD's.

Developments are under way to read x and y simultaneously from one MSD (orthogonal strips on front and back) but of course this does not resolve the pairing ambiguities.

Precision (0°):

5 x 5 μm for CCD's

3 μm for MSD's if every strip read.

Development of current ideas could lead to 1 μm MSD precision, and 0.2 x 0.2 μm CCD precision.

Precision (Angled Tracks):

Due to the thickness, this degrades much faster for MSD's than for CCD's. Energy loss fluctuations lead to non-Gaussian tails. At 45°:

CCD's have 10% probability of error > 4 μm

MSD's have 10% probability of error > 80 μm.

2-Track Resolution:

40 μm in space for CCD's (reading every pixel)

40 μm in projection for MSD's (reading every strip).

Flux Limit Per Crossing (Colliders):

Due to the much higher information storage capacity of CCD's (2500 per mm^2 compared with 50 per mm^2 for MSD's) they have a much higher capability of absorbing extraneous hits from background interactions, synchrotron radiation, etc. The actual limit depends on the quality of the overall tracking system (i.e. how precisely a track can be projected onto the CCD detectors in order to determine which hits are signal and which background).

Beam Rate (Fixed Target Experiment):

~ 1 MHz for CCD's

~ 10 MHz for MSD's with fast readout.

Use of multiplexing readout tends obviously to reduce the rate capability of MSD's.

Readout Time:

~ 40 ms for CCD's

~ 1 ms for MSD's via multiplexing.

MSD's can give fast outputs (eg. for triggering) in those cases where one can tolerate independent readout electronics on each strip (practicable over a small area only).

Area Coverage:

~ 4 cm^2 for currently available CCD's

~ 20 cm^2 for currently available MSD's.

For larger area coverage (eg. colliders) in both cases there are plans for mosaics of many detectors.

Overall, we may say that CCD's have a role when one wants to work as close as possible to the primary vertex, where track merging

and/or background flux would prohibit MSD's. They then have the advantages of minimal extrapolation length of tracks to the primary vertex (most important if multiple scattering is serious), unique space points and potentially higher precision.

MSD's provide less information but have advantages in the case of higher continuous rates and where larger area coverage is essential (eg. photon beams or LEP).

Note also that both forms of detector are advancing rapidly. The limitations of 2 years ago (slowness of CCD's, very large volume of off-chip electronics for microstrips) are now much reduced. Furthermore there are some very interesting ideas for hybrid detectors (MSD's with linear CCD readout, deep depletion silicon detectors connected by bump bonding to 2-d CCD's, etc) which will not be discussed here since they are rather speculative. But they may be very important in future. Hybrid detectors have been widely developed for other fields, especially in infra-red imaging. For a review of the inter-connection possibilities, there is an excellent paper by Chan[35] on this subject.

Finally, it is perhaps interesting to comment on a major technical challenge in this field. The granularity of CCD's in most vertex detector applications is more than adequate. But (remembering Figure 5) one would welcome improved measurement precision below 5 μm in all cases where multiple scattering is not the limiting factor. This applies especially in fixed target experiments where secondary momenta are high and the limit from multiple scattering is typically much less than 1 μm. The aim for improved precision can in principle be met by some technique for charge spreading. CCD's are used in star guidance systems for space-based applications where, by defocusing the star images on the CCD surface and using a centroid finding approach, precisions of $\sigma_x = \sigma_y = 0.1$ μm have been achieved,[36] despite the pixel sizes of 20 x 20 μm. The principle of this technique is illustrated in Figure 47. How could this be used for particle detection? In view of the very unfavourable diffusion vs. drift characteristics mentioned in Section 3.3.3, the solution probably does not lie in this direction. One idea, which has not been fully evaluated, is to build a very shallow depletion CCD

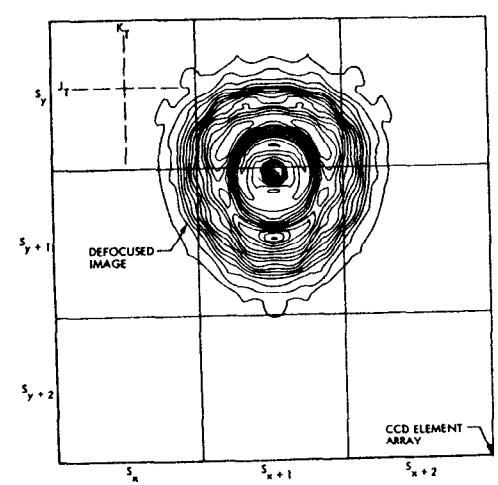

Figure 47 From Reference 36. Each square represents a CCD pixel. The contours represent a defocused star image of optimal size for position determination by centroid finding.

(~ 1 μm depletion depth) on an epitaxial layer of ~ 10 μm thickness (see Figure 48). In this case most of the signal would be generated by the isotropically diffusing electrons in field-free undepleted material, which might give adequate charge spreading for measurement precision of 0.2 μm to be achieved. But this is a very new area and there is certainly room for imaginative thinking on the problem.

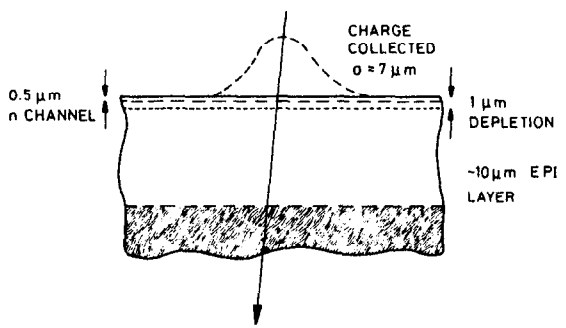

Figure 48 An idea for a particle tracking CCD with precision of about 0.2 μm. It consists of a very shallow depletion device, and relies for most of the signal on diffusion from the undepleted epitaxial substrate.

6. SILICON DRIFT CHAMBERS

There has been a good deal of speculation about novel silicon detectors which might satisfy the specific requirements for high precision tracking but operating according to different principles than MSD's or CCD's. So far, the only one which has been built is the silicon drift chamber, proposed by Gatti and Rehak[37] and built in Munich by Kemmer and co-workers. The basic idea is as follows. Starting from a wafer of high resistivity n-type silicon, make p^+ implants on both surfaces (Figure 49) and cover these with grounded metallic electrodes. Then, at a single electrode near the edge of the detector, apply a positive potential which increasingly reverse-biases both junctions. The depletion regions expand from both sides; eventually (just as in the CCD) they meet and produce a fully depleted (pinch-off) condition. In the present case, the

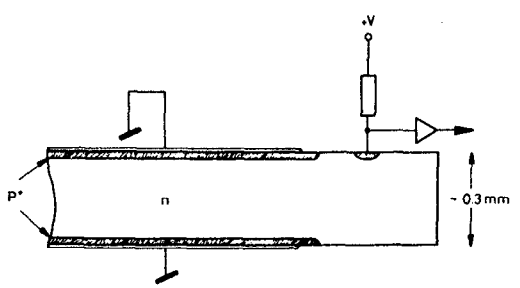

Figure 49 Cross-section of basic structure used for the silicon drift chamber.

pinch-off occurs in the mid-plane of the detector. Figure 50 shows the field and potential distributions through the silicon under several different conditions: (a) shows the case of both surfaces being grounded. The electric field grows rapidly once one penetrates the depleted p^+ silicon, has a maximum at the np junction, then falls gradually through the lightly doped bulk material. It changes sign at the mid-plane of the detector. The corresponding potential distribution has the familiar form of a combination of quadratics, and leads to a shallow potential energy well for electron storage at the detector mid-plane. Just as in the CCD, this is not an

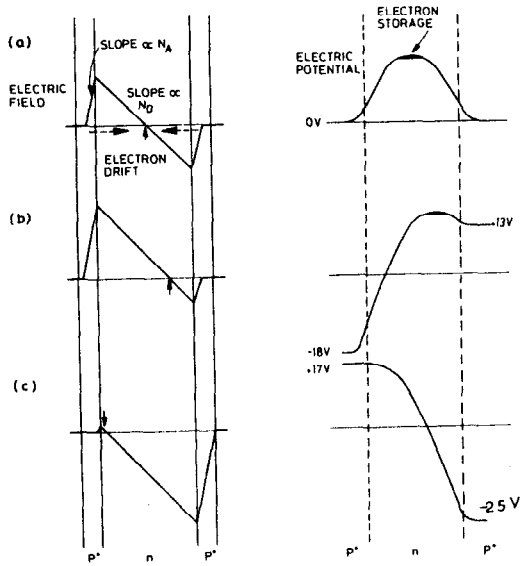

Figure 50 Field and potential distributions in the silicon drift chamber for different values of the surface potentials.

equilibrium condition, and thermal generation of electrons would cause an accumulation in this region leading to a loss of depletion. If the surface voltages are altered, the potential energy minimum can be shifted close to either face (Figure 50(b)) or even eliminated completely (Figure 50(c)). In the last case any generated electrons would be collected by the more positive surface electrode. The scheme used in the actual detector is indicated in Figure 51. The surfaces are subdivided into strips, whose potentials are graded, becoming steadily more positive from right to left, inducing a very gentle drift field within the detector. This field is small compared

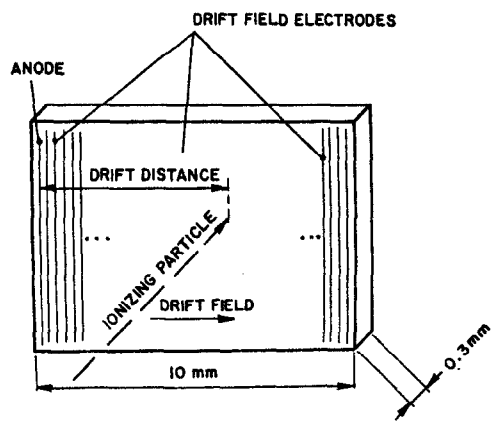

Figure 51 From Reference 37. Silicon drift chamber. The surface is covered by a strip array of p^+ junction electrodes, which provides the depletion and the drift field. Electrons produced by the passage of a fast charged particle drift towards the anode, which is the only readout channel on the wafer.

with the typical internal fields in silicon induced by the depletion process, as it needs to be if it is to lead to measurably long drift times. Typically, with an electrode pitch of 150 μm and ΔV = 2.3 volts/strip, we have a drift field of 150 V/cm.

From Section 3.3.3,

$$\text{drift velocity } V_d = -\mu_n E$$
$$= 1350 \times 150 = 2.0 \times 10^5 \text{ cm/s}$$
$$= 2.0 \text{ μm/ns}.$$

As the electrons approach the anode, the potentials of strips on opposite faces of the detector are made more and more unequal, so that the electrons are slewed towards one face of the detector as shown in Figure 50 and 3-dimensionally in the plots of Figure 52. Given a narrow anode strip and good timing based on zero crosser discrimination, approximately micron precision may be achievable in the co-ordinate of the particle normal to the drift direction. As with the microstrip detector, we are discussing a 1-dimensional detector, but with the advantage of a great reduction in the number of output channels. In principle some measure of 2-dimensionality is also possible by using a subdivided anode, though this has not yet been implemented.

For this idea to work, it is essential to have very uniformly doped material, in order to avoid inhomogeneities and consequent smearing of the electron times of arrival at the anode. The most uniformly doped material which can be grown is (for technical reasons) p-type, and this (with a resistivity of about 10 K Ω cm) is used to start with. It is then turned into n-type material by the

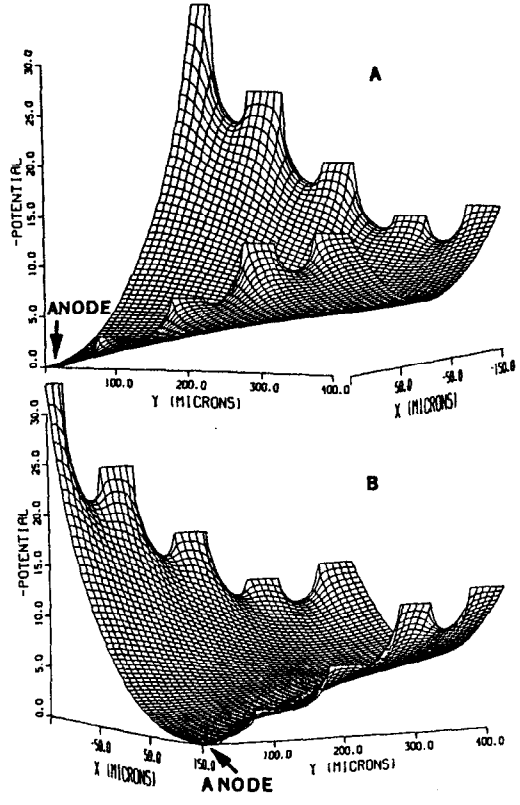

Figure 52 *From Reference 37.* The potential within the semiconductor drift chamber close to the readout anode. (Two views of the same potential from different points are shown.)

procedure known as neutron doping. The crystal is irradiated with slow neutrons and by means of the reaction

$$Si^{30} + n \rightarrow Si^{31} \rightarrow P^{31} + \beta^- + \bar{\nu}$$

is turned into n-type material. The resistivity is monitored and the irradiation ceases when this falls to 10 KΩ cm.

Tests with min-I particles have resulted in a measured precision of 20 μm over a 4 mm drift length, compared with 5 μm when a small light spot was used to generate a surface charge. The difference may be due to some smearing associated with the extended source of charge generated by the traversing particle.

One possible limitation of this type of detector may be the 2-track resolution, but this would be helped by a subdivided anode structure. Overall, results are very encouraging and certainly demonstrate the value of new ideas in silicon tracking detectors. There are also very interesting possibilities for high sensitivity photon detection using the drift technique without the position measurement, as discussed in Reference 37.

7. APPLICATIONS

We shall here look at some examples of the use of silicon detectors as high precision tracking devices for secondary vertex detection. One should not forget the contrasting approach of the FRAMM Collaboration and others (eg. NA1 experiment on the CERN SPS) who look for secondary vertices based on a change in pulse height using multiplicity counters downstream of the primary vertex. We consider one example from currently running experiments and one future application as an illustration of current trends.

7.1 Current Results. The first charm lifetime measurements using electronic tracking detectors came from the ACCMOR Collaboration in the NA11 experiment at the CERN SPS, using a telescope of 6 planes of microstrip detectors following a beryllium target. This setup is shown in Figure 53, and was followed by a large multiparticle spectrometer which included equipment for the

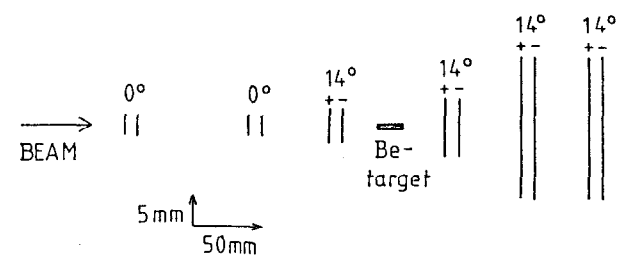

Figure 53 Side view of NA11 target region showing the beryllium target and the silicon strip detectors. The inclination of the strips to the horizontal direction is indicated.

detection of single electrons. This requirement in the trigger and offline analysis produced a charm enrichment factor of about 20 with respect to all inelastic interactions. The incident beam was 200 GeV/c π^-.

A typical reconstructed charm event is shown in Figure 54. Results on charged and neutral D lifetimes were reported at the Brighton Conference.[38] In addition, the lifetime of the charmed F meson has recently been measured[39] in the same experiment.

The microstrip detectors have given some beautifully clean charm events (like the one shown in Figure 54) but in the case of high multiplicity events there are often problems due to merging of the tracks in one or more MSD planes, despite the fact that the first plane is about 4 cm from the vertex. An example of such an event is shown in Figure 55, where merging of 2 tracks causes confusion in the

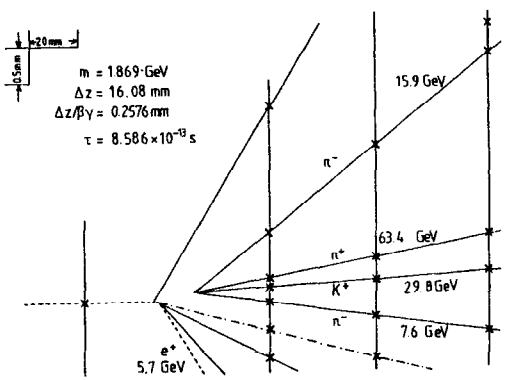

Figure 54 Reconstruction in one MSD view of the primary and decay vertices in a charm-production event. The decay $\bar{D}^0 \to K^+\pi^+\pi^-\pi^-$ is seen, with all final state particles identified by Čerenkov hodoscopes.

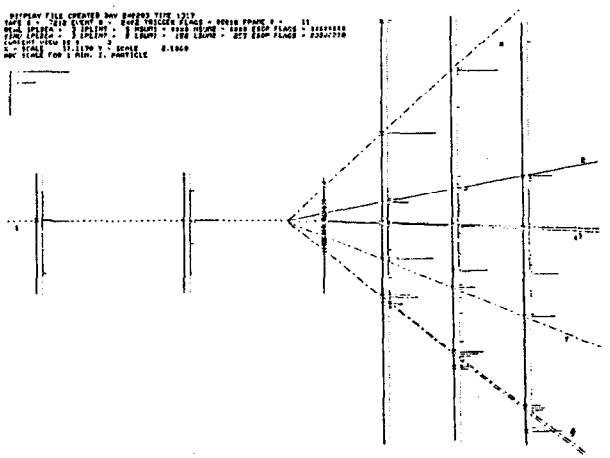

Figure 55 An event where the hits from 2 tracks (numbers 3 and 4) are merged in the first MSD plane giving rise to a bad fit to a single vertex. The data from one CCD plane are seen in projection.

fitting. In order to clean up such events, and in general to make precise measurements of x and y as close to the vertex as possible, it was decided to include CCD detectors in the setup, using the slim cryostat already described in Section 5.3. Initially, one CCD only was operating and the hits in it are visible in Figure 55. Of course, when viewed in projection, the data look confused, but since each cross represents a space point the track finding can in fact be done extremely easily. Figure 56 shows the reconstruction of the event in two orthogonal views with and without the help of the CCD's. The improvement in the vertex from the CCD's comes partly from the

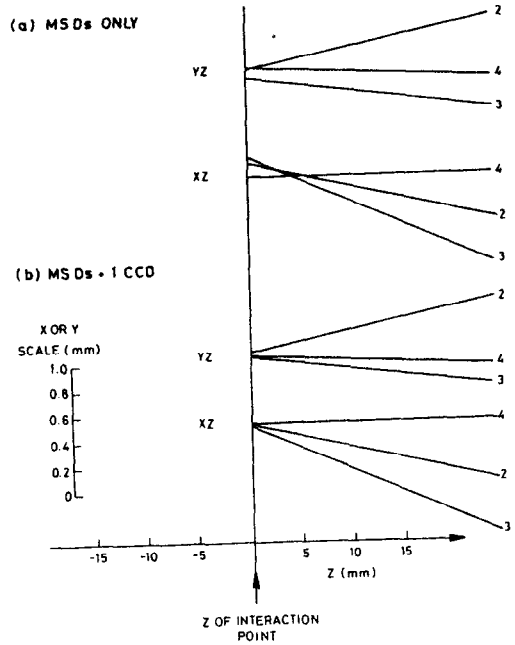

Figure 56 Reconstruction of the event in the vertex region
(a) using MSDs only and
(b) using the CCD data in the track reconstruction.

increase in precision close to the vertex but also (most importantly) from the fact that the CCD data are free of track merging effects.

These first tests were done in December 1983. By June of 1984 data were being taken with 2 CCD's included in the vertex region. Figure 57 shows a typical event, where there is again evident track merging in the first MSD planes. This is a representation of the event in one of the MSD viewing directions (14° to the horizontal).

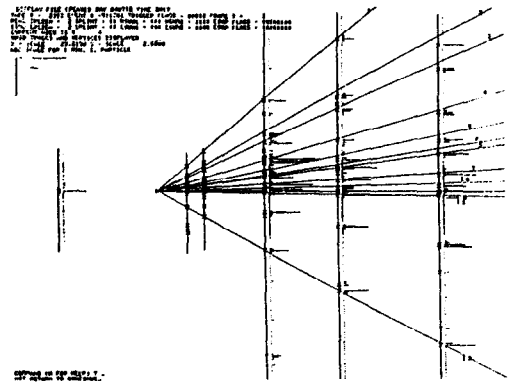

Figure 57 A complex event in which there is evident track merging in the first MSD plane. The data from two CCD planes are seen in projection.

The tracks shown are fits to the CCD data alone. In this view the CCD information looks very confused, but the correct way to look at the event in the CCD's is face on to the detectors, i.e. along the beam direction. This is shown in Figure 58. The upstream detector has 7 hits in 1 mm^2, and shows no problem at all of track merging in spite of the fact that the detector is only 15 mm from the vertex in a high multiplicity interaction at 200 GeV/c. The accidental background is completely negligible but it should be noted that this event was obtained with a beam flux of only 10^5 per second.

7.2 Proposed Future Applications. As we noted in Section 1, events of the type $e^+e^- \rightarrow Z^\circ \rightarrow q\bar{q}$ will be a rich source of physics involving short-lived particles. SLC is a particularly promising

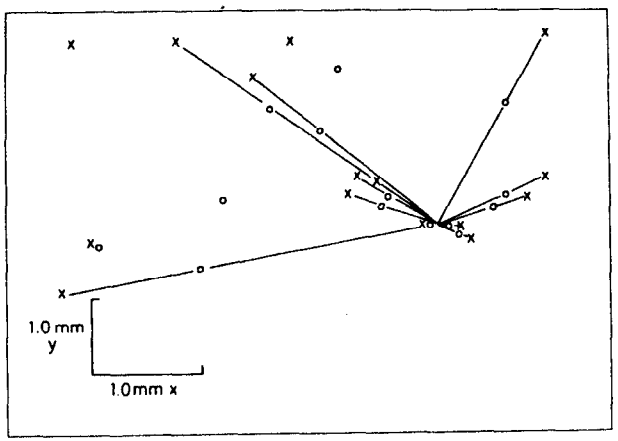

Figure 58 The event of Figure 57 as seen by the CCD detectors looking along the beam direction. The circles represent hits in CCD1 and the crosses represent hits in CCD2.

environment since the beam pipe can be made small (less synchrotron radiation at wide angles than in the circular collider LEP). Thus the detectors can be placed close to the primary vertex. Details have still to be finalised but Figure 59 shows the planned barrel vertex detector for SLD[40] Based on a beam pipe radius of 1 cm, this envisages an inner barrel of CCD's at this radius, and an outer barrel at double this radius. There will also be mosaics of end cap detectors to cover the production down to small angles.

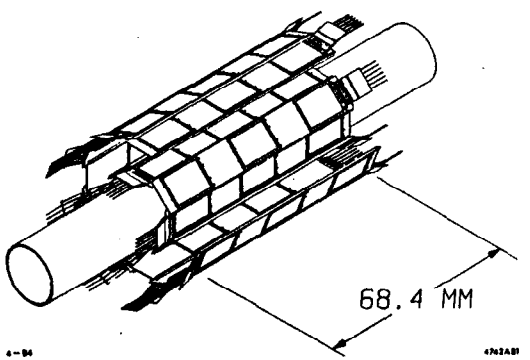

Figure 59 Arrangement of CCD's to make up the vertex detector for SLD. There is an inner barrel which fits closely around the beam pipe and an outer barrel (partly cut away in the drawing to reveal the inner one) at about twice the radius of the beam pipe.

The calculated detector performance is summarised in Figure 60. This shows the precision of the projected impact parameter with respect to the primary vertex, assuming the combined reconstruction power of the CCD vertex detector and the SLD central drift chamber. In the beam view of the event (XY plane) the full precision of the drift chamber gives the curve shown. The precision degrades badly below 1 GeV/c due to multiple scattering in the beam pipe and vertex detector. In the orthogonal plane (ZY') the precision in the vertex detector is unchanged, but the precision of the impact parameter is worse since here one relies on the Z precision of the central drift chamber in the overall fit, and this is considerably worse than the precision in the azimuth. Again at low momenta multiple scattering

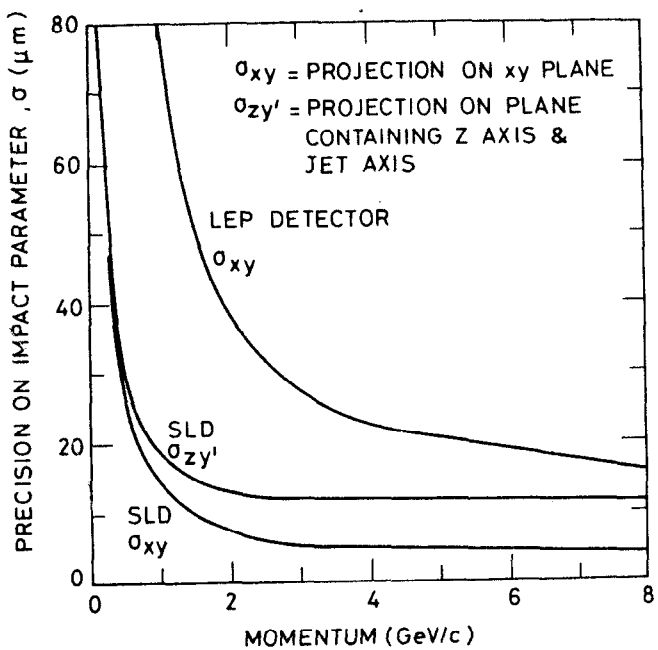

Figure 60 Precision on track impact parameters with respect to the primary vertex in two projections
(a) onto a plane (XY) normal to the beam direction, and
(b) onto a plane (ZY') containing the beam direction and the jet axis.
The lower curves refer to the SLD design including the CCD vertex detector and the upper curve refers to the LEP geometry and assumes 3 barrels of microstrip detectors.

dominates, but in both projections there is a large part of the useful momentum range over which the full precision of the vertex detector is used.

In contrast, the situation at LEP will be rather less favourable. Operating on an inner detector radius of 8 cm (due to synchrotron radiation background) even a sophisticated 3 barrel MSD system with 3 μm measurement precision will yield impact parameter precision about 5 times greater than at SLD, due to the larger lever arm. Indeed, at LEP the impact parameter precision will be multiple scattering dominated even far above 10 GeV/c, where very few secondary particles are expected. It is unlikely that MSDs at LEP will provide an orthogonal view of the event, and CCDs are excluded because of the very large detector area required around the thick beam pipe.

What does Figure 60 imply for the event reconstruction? An example is shown in Figure 61, which is the simulated reconstruction of the $b\bar{b}$ production event of Figure 4(b). Tracks are drawn with solid lines if they can be uniquely assigned to their true vertex (whether this be the primary vertex, the B (or $\bar{B}$) decay vertex or the D (or $\bar{D}$) decay vertex. Figure 61(a) assumes the SLD precision. We see that all the primary tracks are correctly associated, that 2 of the 6 B decay tracks are unique, and that both D decay tracks are unique. For the LEP reconstruction (Figure 61(b)) all but one of the primary tracks and all of the B decay tracks are ambiguous, while the D decay tracks are unique.

By looking at a large number of events of various types, we can arrive at the efficiency figures with which the planned SLD detector is able to uniquely associate tracks with vertices; these are

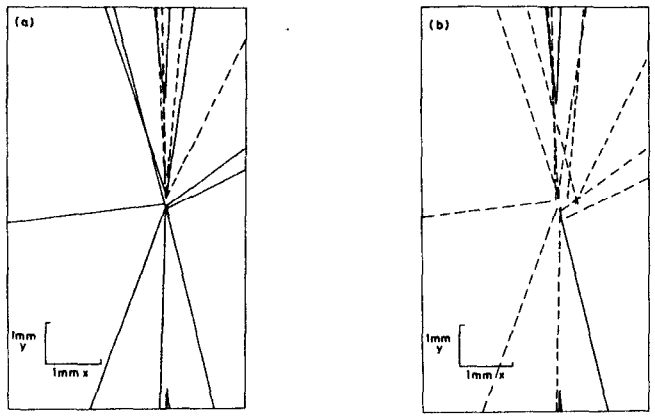

Figure 61

(a) SLD simulation. (b) LEP simulation.

Reconstruction of the event $Z^\circ \to b\bar{b}$ of Figure 4(b). Tracks which are ambiguous as to their vertex assignment are shown by broken lines.

summarised in the table. For comparison, the corresponding figures for the assumed LEP detector are included.

Event Type	Unique Tracks SLD		Unique Tracks LEP	
	Primary Vertex	Decay Tracks	Primary Vertex	Decay Tracks
$Z^\circ \to c\bar{c}$	91%	83%	51%	41%
$Z^\circ \to b\bar{b}$	86%	76%	48%	30%
$Z^\circ \to t\bar{t}$	85%	65%	43%	16%

In the case of the SLD detector, the efficiency falls slowly with increasing quark mass due partly to the increasing topological complexity of the events (more vertices give a higher probability of confusion) and partly due to the falling momenta and so worse multiple scattering of the final state particles (remember Figure 6). These efficiencies are sufficiently high that one can be assured of a great deal of interesting physics with such detectors looking at Z° decays. In the LEP case, the effects of multiple scattering are correspondingly worse.

Microstrip detectors at LEP can cover the large areas needed, which CCD's could not. Microstrip detectors may also have an important role at SLC. Their use is disfavoured for the reasons that

(a) they provide only one view, and

(b) they probably need to be placed at a larger radius in order to have a tolerably low level of track merging. This means that they will have poorer precision in measuring the impact parameters of low momentum tracks.

However, their use is favoured for the reason that they may be simpler to implement, especially if the SLC synchrotron radiation background necessitates operation of the inner barrel at a radius of 2 cm or greater.

Fixed target experiments have been a very important testing ground for these high precision detectors. Much physics may still be possible in such experiments, specially at the higher energies available at the Tevatron. From the point of view of event reconstruction, the fixed target environment is preferred to colliders, due to the higher momenta. But the low cross-section for charm and bottom production in hadronic collisions means that a lifetime trigger is highly desirable. To date, there has been less success with trigger schemes than with tracking detectors, and there is an urgent need for new ideas in this area.

8. RADIATION DAMAGE

Silicon detectors form a subset of all MOS devices. Radiation damage is relevant to many users of these devices and has been extensively studied. Much of the work is motivatd by industrial and military applications, but in the field of low level optical imaging the space-based astronomers have particular interest due to Van Allen radiation belts and on-board nuclear power plants.

Microstrip detectors, CCD's, silicon drift chambers, MOS multiplexing chips, all need individual consideration with regard to radiation damage. Rather than go through each in detail, we can get a general feeling for the different classes of effect, and consider CCD's as a typical example which embodies most of the effects relevant to the other devices.

8.1 General Discussion. There are essentially three classes of effect, namely ionization effects, atomic collisions with sufficient momentum transfer to disturb the atom in the crystal lattice, and nuclear interactions which may result in chemical changes (eg. Si → P) and large energy release by nuclear disintegration including α emission, etc.

Atomic collisions and nuclear interactions 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 photon or electron irradiation. Displacement damage may also give

rise to 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.

In undepleted silicon the effects of displacement damage are:

a reduction in the minority carrier lifetimes,

a decrease in carrier concentration (at higher irradiation), and

a reduction in carrier mobility (at still higher irradiation).

In depleted material, specially in silicon close to the internal breakdown fields, the damage centres can act as generators of large dark current spikes. (Virtual phase CCD's, for example, are particularly prone to this effect.)

Ionization generally induces transient effects in the bulk material, but long-term or permanent surface changes. These are:

trapped charge buildup in the silicon dioxide layer, and

increase in the density of trapping states at the Si/SiO_2 interface.

The first of these effects is due to the generation of electron-hole pairs within the oxide. In the absence of an electric field these largely recombine so that the radiation damage is almost non-existent. Unfortunately, devices frequently need to have applied bias voltages during irradiation, and these sweep out the electrons whose mobility in silicon dioxide is high. The result is an accumulation of immobile holes, i.e. positive fixed charge. Depending on the sign of the applied voltage, these accumulate mainly near the metal gate (gate negative) or near the silicon interface (gate positive). The latter case generally gives rise to more serious effects on the performance of the device.

The effects of the two abovementioned phenomena are:

threshold voltage shifts, and

reduced mobility of surface charge (eg. reduced MOSFET transconductance).

Threshold voltage shifts are best described in terms of the flat-band voltage of the gate. This is the gate voltage required in an otherwise unbiased device to induce depth-independent valence and conduction band edges in the silicon below the oxide. In the absence of fixed charges, the flat-band voltage is zero. In fabricating a device, there are inevitably some fixed charges at the silicon/silicon dioxide interface and within the oxide itself. Various special techniques (surface passivation, etc) are used by the manufacturers to minimise these effects, but at some low level every MOS structure is characterised by a non-zero flat band voltage. Radiation-induced charge-buildup in the oxide layer obviously disturbs the balance and leads to an ever increasing shift in the flat-band voltage. All thresholds and other operating levels associated with the gate voltages are correspondingly displaced.

Bulk damage mainly affects 'bulk-effect' devices such as bipolar transistors. Surface damage mainly affects MOS devices. CCD's are in fact sensitive to both types of damage. In general,

10^4 rads is acceptable (but there are exceptions!),

10^6 rads produces serious degradation or failure;

unless special precautions are taken in manufacture and operation;

10^6 rads corresponds to:

10^{15} 15 MeV neutrons/cm^2

2×10^{15} 1 MeV γs/cm^2

4×10^{13} Min-I particles/cm^2.

In collider experiments, the main concern comes from synchrotron radiation-induced soft X-rays. But the limit from synchrotron radiation will normally be set by the confusion from background hits in the detector. This occurs well below the level at which radiation effects become serious.

In fixed target experiments, the passage of an intense hadron beam through the detector can cause local radiation damage on a timescale shorter than the life of a typical experiment.

8.2 Radiation Damage in CCD's. There has been a great deal of excellent work on this problem, much of it by J. Killiany and co-workers at the Naval Research Laboratory in Washington. Reference 41 provides a very useful review of the subject.

8.2.1 Bulk Damage in CCD's. The bulk of the silicon is unaffected by low momentum transfer ionizing collisions but seriously affected by nuclear interactions. As expected, the effect of neutrons is (for example) a good deal more severe than that of photons or min-I particles.

A factor 100 increase in the room-temperature dark current is given by 10^4 rads of neutrons. The displacement damage can give rise to states within the band gap of the silicon which act as centres for the creation of electron-hole pairs. The resultant dark current may still be tolerable in cooled CCD's, except for virtual phase devices which can develop such a high density of intense dark current spikes as to be unusable.

The same neutron flux induces trapping centres in the silicon bulk which result in a loss of charge transfer efficiency. The probability of an electron failing to move from one pixel to the next during the charge shifting process becomes very high (about 2×10^{-3}). This gives rise to serious loss of signal and variation in response across the detector.

Fortunately, the scale and character of the damage resulting from the equivalent dose of γs or min-I particles is much reduced. More than 10^6 rads are needed to produce equivalent effects.

The relative seriousness of neutron-induced radiation damage should not be overlooked in experimental situations, where measurements of neutron background may explain the degradation of all types of silicon detectors, not only CCD's. Improved neutron shielding may be important for increasing the lifetime of the detectors.

8.2.2 Ionization-induced Damage. Consider first the problem of charge buildup in the oxide. This affects different sections of the device differently. The charge transfer section is mainly affected by the shift in the flat-band voltage, all gate voltages needing to be correspondingly displaced if the device performance (eg. well capacity) is to remain constant. The problems are obviously much

worse if the irradiation is non-uniform across the detector area, a phenomenon quite common in vertex detection systems, but not relevant to military or astronomical applications. In our case, there is not much to be gained by feedback-controlled gate adjustment, a technique which is useful in the case of uniform irradiation. The only clean solution is to select a technology which minimises the charge buildup.

For operation in the temperature range 200-300° K the situation can be made not too serious. Firstly, one should use an n-buried channel structure since in this case the gate voltages are negative with respect to the channel potential and the charge accumulates near to the metal/oxide interface, not the oxide/silicon interface. Secondly, one should use a planar insulator (avoiding the stepped oxide used in some CCD manufacturing processes). Thirdly, one should use a "hard oxide technology"; this covers a whole art which may mean different things to different practitioners. By combining these elements, it is possible to retain good CCD operation after 10^6 rads of ionizing radiation, whereas by neglecting them the devices may not be usable beyond 10^4 rads.

Figure 62(a) shows that while the situation can be made tolerable at higher temperatures, there is a rapid degradation below 200° K. The reason for this is that at lower temperatures relatively shallow trapping centres are able to retain fixed charges, whereas at higher temperatures the thermal energy makes them ineffective. Figure 62(b) shows that even at low temperatures, the problems are very dependent on the applied gate voltage. While it is not normally

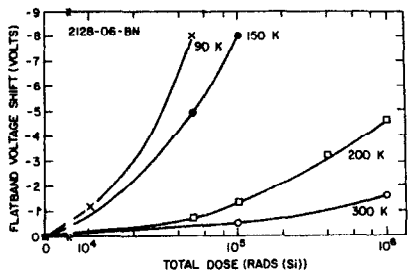

Figure 62 From Reference 41. Various plots indicating the sensitivity of radiation damage to time, temperature and other parameters.
(a) Flat-band voltage shift for the CCD radiation-hard oxide as a function of dose at several temperatures, illustrating the increased oxide charge trapping effects at low temperature.

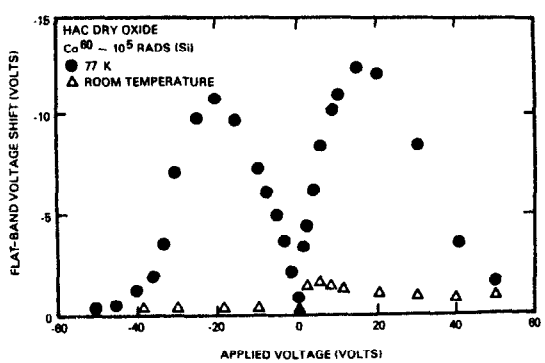

Figure 62 (b) Flat-band voltage shift at 10^5 rads (Si) as a function of applied voltage, illustrating the electric field dependence of the flat-band voltage shift at 77° K.

practicable to have zero bias during irradiation, it is important to be aware of this sharp voltage dependence in considering the optimisation of operating conditions.

These radiation effects are further complicated by the phenomenon of self-annealing. Figure 62(c) shows the time dependence of the flat-band voltage shift after a short burst of radiation. As expected, the self-annealing is most effective at higher temperatures, where the trapped holes have an increased probability of being released by fluctuations in the thermal energy. Even if a detector needs to be operated cold, it can be restored to health by periodically warming it to room temperature, as shown in Figure 62(d). This in fact refers to an input gate threshold voltage shift, but the principle applies generally. The cumulative radiation damage

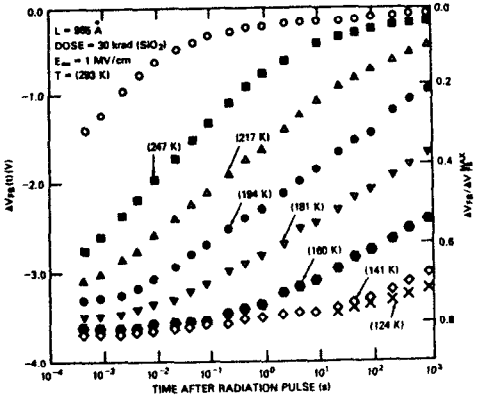

Figure 62 (c) Flat-band voltage shift as a function of time after 30 krad radiation pulse, illustrating annealing of the flat-band voltage shifts at various temperatures.

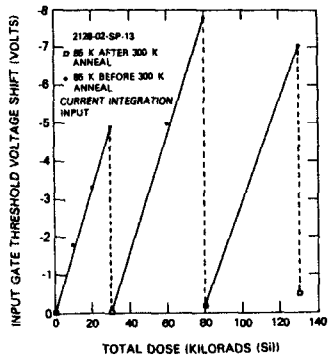

Figure 62 (d) Radiation-hard CCD input gate threshold voltage shift at 85° K as a function of dose and 300° K annealing.

after each anneal is very slow, and corresponds exactly to that which would be obtained in continuous room temperature operation.

If it is absolutely necessary to run a CCD at liquid nitrogen temperature, then a change of insulator is needed. Killiany has manufactured CCD's using a thin oxide/thick nitride insulator (1000 Å of Si_3N_4 and 100 Å of SiO_2) and obtained flat-band voltage shifts as low as -1 V/10^6 rads at 80° K. For particle detection applications, however, one should virtually never need to use such low operating temperatures.

So far we have been discussing the effect of charge buildup in the charge transfer section of the CCD. Let us now turn to its effect on the output section of the device.

The CCD output section is in fact the least radiation sensitive part of the detector due to the following factors:

(a) The reset MOSFET is purely a switch, and so is insensitive to threshold shifts.

(b) The threshold voltage shift in the output MOSFET is reduced due to the fact that the electric field strength within its oxide is typically 4 times less than in the charge transfer section. (Remember Figure 62(b).)

(c) The source follower is AC coupled to the local preamp, so small DC shifts are unimportant. DC coupled systems are much more difficult to control.

(d) The output FET is used as a source follower, and so we are insensitive to changes in g_m. In contrast, high gain stages will inevitably be more radiation sensitive.

This completes the discussion of charge buildup in the oxide. We have still to consider the ionization-induced increase in density of interface states.

This phenomenon (at very low radiation doses) completely ruins the charge transfer efficiency in surface channel CCD's (ie. devices in which the potential energy minimum sits at the Si/SiO_2 interface instead of about 1 μm below it). In any case, the charge transfer efficiency of non-irradiated surface channel CCD's is generally inadequate for very small signal levels; it is essential to use only buried channel CCD's.

For the same reason, it is important that the output MOSFET be run in buried channel mode (with the FET current confined to a channel displaced below the surface). This again is desirable in any case for noise optimisation in non-irradiated devices.

The interface states act also as generation centres for the surface component of the dark current. For example, at room temperature, after 10^6 rads the dark current density can increase by a factor of 500, from 2 nA/cm^2 to 1 $\mu A/cm^2$. The dark current may still be negligible at 200° K if reasonably fast device readout is employed.

While being applied specifically to CCD's, these comments on radiation damage illustrate most of the factors which apply to other silicon detectors and local preamplifier and multiplexing electronics. Each case is obviously different, and different parts of a device need to be considered separately. There are no simple numbers which one can quote as to the radiation hardness of silicon devices, simply because the variety of structures is so much greater than (for example) plastic scintillator. While this variety complicates the understanding of the problem, it also in many cases leads to a level of flexibility in which solutions can be found. The progress made in radiation hardening of CCD's is the best possible argument for a careful and systematic study of radiation damage, as a necessary prelude to minimising its effect in any type of detector.

9. PRACTICAL HINTS

Working with high precision silicon detectors is still very much a minority activity in the high energy physics community. Unlike (for example) drift chamber developments, there is not a large body of experts with a commonly known and widely shared pool of knowledge, expertise and exploratory techniques. Yet the physics interest of short-lived particles is obviously high, and the scope for technical developments is very great. It is to be hoped that the community of physicists and engineers interested in working in this field will continue to increase, to the point where we have an effective 'grapevine' for the exchange of ideas and experience.

Two contrasting examples come to mind. The phenomenon of whisker growth in drift chambers has been widely observed, and studied by several independent workers. By pooling their experience, one can begin to build up a detailed picture of what is taking place. The phenomenon of corrosion of wire bonds in microstrip detectors, in contrast, has been only superficially tackled. There is little doubt that the latter phenomenon would profit as much from a multi-pronged attack as did the former.

It may be useful, for the benefit of any potential newcomers to the field, to collect up a few hints which can ease the practical evaluation of detectors.

Firstly, particle detection enthusiasts should not forget the very useful responsiveness of silicon to optical input. A light spot, generated by an LED in the image plane of a microscope's camera attachment, can be used as a fine probe (few μm diameter) of the detector performance. Using the viewing channel of the microscope, one can place the spot precisely where one wants it on the detector surface, study the response versus position, etc. In this way, local sites of signal loss or degradation can be identified with high precision. By varying the LED on-time one can study linearity of response and possible threshold phenomena in the detector. By using a near-infra-red LED, it is possible to establish the response of the detector to charge generated deeply below the surface of the silicon.

Another powerful tool is provided by X-ray sources of several keV energy. They give signals which can be close to those from min-I particles, are (in contrast to min-I signals) absolutely mono-energetic, and so allow energy calibration and resolution to be established, and they provide essentially point distributions of charge generated uniformly within the bulk of the material (refer to Figure 10 for the electron ranges). As such they provide powerful tools for studying smearing of charge by whatever charge transport or collection procedure is used in that detector, whether microstrip, drift chamber or CCD.

A frequent mystery to newcomers is how one makes the link between the general discussion of Section 3 (for example) and the actual detector with its multitude of bias electrodes. A most useful tool, which establishes such features as depletion depth, channel pinch-off or the lack of it, etc., is the measurement of gate-to-substrate capacitance at a particular bias voltage or as a function of bias voltage. In this way one can determine whether the

conditions below the gate represent inversion, punch through, depletion or accumulation. This is illustrated very clearly in Reference 33, page 48.

The on-chip or off-chip circuitry needs careful tuning if operation conditions are to be optimised. Given the need for thin detectors, signal levels are minimal and external noise sources (pickup, etc) have to be carefully screened. We have found that an essential ingredient in this process is a good quality wideband spectrum analyser. With experience, this becomes as important a diagnostic tool as an oscilloscope and provides quite different information. It can immediately establish if there are sources of external pickup and if so measure their effect on the overall system noise. It can show whether a FET is operating in surface channel or buried channel mode, and it provides a quantitative basis for studying noise as a function of operating temperature and any other relevant parameters. When mysterious changes in signal occur, it can establish whether the cause is internal or external to the detector.

Finally, a few general comments. The production of the detectors we have been discussing is at present divided among different groups of specialists. There are those with a background in nuclear physics, who are building the microstrip detectors and silicon drift chambers. They are the experts in processing high resistivity silicon. Then there are those with a background in optical imaging, many having an industrial or military base. They are the experts in building complicated detectors (such as CCD's) using all the arts associated with the MOS technology, generally using low resistivity material. There are ideas for hybrid detectors, deep depletion CCD's, etc, which cross the boundaries between these groups of experts. This has led to the idea of a dedicated silicon detector group which would aim to solve the problems of high energy physics detectors by combining the expertise of the currently separate groups. This is an interesting idea, but personally I would not recommend it. The pressures for progress in other areas are very strong, and I believe we can do best (as in the old days of nuclear emulsion developments) not by trying to set up an independent laboratory for this work, but by co-operating very closely with industry. Scientists in industry enjoy working with physicists who have a systematic and thoughtful approach to the evaluation of their products, and who can provide useful technical input in some areas, for example in a close study of radiation damage phenomena. The development of detectors in this field is intrinsically more inter-disciplinary than in some other areas of high energy physics, and is more likely to profit from a collaborative approach rather than by trying to make the necessary developments in isolation.

ACKNOWLEDGEMENTS

Being a relative newcomer to the field of silicon detectors, I owe a great debt to the experts who have generously supplied me with information and generally helped in my education on this topic. Among these are David Burt of GEC (England), Joe Killiany of the Naval Research Laboratory, Washington, Velkjo Radeka of BNL, Emilio Gatti of Milano, Rangy Kandiah of AERE (Harwell), Pier Francesco Manfredi of Pavia, and Erik Heijne of CERN. In addition, some of my fellow high energy physicists who became interested in silicon detectors around the same time as I did have provided a very exciting and stimulating exchange of ideas. I am especially grateful to Berhard Hyams of CERN, Robert Klanner and Gerhard Lutz of MPI, Munich and Sherwood Parker of Hawaii. My colleagues at RAL have formed a splendid team, all of them making essential contributions to the success of the CCD detector work, viz. Bob English, Tony Gillman, Laurie Lintern, Richard Stephenson, Julia Walsh, Steve Watts and Fred Wickens.

Finally, I would like especially to thank David Leith for inviting me to talk at the SLAC Summer Institute, and for his hospitality and that of the other organisers during this excellent meeting.

REFERENCES

1. J. Ellis and M.K. Gaillard, CERN 76-18 (1976) 21.
2. Proceedings of the SLC Workshop, SLAC-247 (1982) p23 et seq.
3. B. Stech, Heidelberg Preprint HD-THEP-84-8 (1984).
4. A.J. Buras et al, MPI-PAE/PTh 7/84 (1984).
5. V. Barger and R.J.N. Phillips, MAD/PH/155 (1984).
6. H. Lynch, SLD New Detector Note No. 117 (1984).
7. L-L Chan, Physics Reports 95, No. 1 (1983).
8. V. Barger et al, MAD/PH/150 (1984).
9. V. Barger et al, MAD/PH/152 (1984).
10. V. Barger et al, MAD/PH/184 (1984).
11. M.I. Adamovich et al, Physics Lett 99B (1981) 271.
12. R. Bailey et al, Contribution to the XXII Conference on High Energy Physics, Leipzig (1984).
13. K.G. McKay, Phys Rev 84 (1951) 829.
14. B. Rossi, 'High Energy Particles', Prentice-Hall (1952).
15. H.A. Bethe, Handb Physik 24/1 (1933) 491.
16. L. Landau, J Phys USSR 8 (1944) 201.
17. O. Blunck and S. Leisegang, Z Physik 128 (1950) 500.
18. V.A. Chechin and V.C. Ermilova, Nucl Instr & Methods 136 (1976) 551.
19. V.C. Ermilova et al, Nucl Instr & Methods 145 (1977) 555.
20. W. Allison and J. Cobb, Ann Rev Nucl Sci 30 (1980) 253.
21. G. Hall, Nucl Instr & Methods 220 (1984) 356.
22. R. Bailey et al, Nucl Instr & Methods 213 (1983) 201.
23. R.S. Muller and T.I. Kamins, 'Device Electronics for Integrated Circuits,' Wiley (1977).
24. A.S. Grove, 'Physics and Technology of Semiconductor Devices,' Wiley (1967).

25. R.A.Greiner, 'Semiconductor Devices and Applications,' McGraw-Hill (1961).

26. J.B.A.England et al, Nucl Instr & Methods 185 (1981) 43.

27. B.Hyams et al, Nucl Instr & Methods 205 (1983) 99.

28. J.T.Walker et al, Nucl Instr & Methods 226 (1984) 200.

29. R.Hofmann et al, Nucl Instr & Methods 226 (1984) 196.

30. G.Zimmer, Nucl Instr & Methods 226 (1984) 175.

31. W.S.Boyle and G.E.Smith, Bell System Technical Journal 49 (1970) 587.

32. J.D.E.Beynon and D.R.Lamb, 'Charge-Coupled Devices and their Applications,' McGraw-Hill (1980).

33. M.J.Howes and D.V.Morgan (Ed), 'Charge-Coupled Devices and Systems,' Wiley (1979).

34. C.J.S.Damerell et al, Nucl Instr & Methods 185 (1981) 33.

35. W.S.Chan, Journal of the Society of Photo-Optical Instrumentation Engineers (SPIE) 244 (1980) 81.

36. P.M.Salomon, Journal of the Society of Photo-Optical Instrumentation Engineers (SPIE) 203 (1979) 130.

37. E.Gatti and P.Rehak, Nucl Instr & Methods 225 (1984) 608.

38. ACCMOR Collaboration, Proceedings of the International Europhysics Conference on High Energy Physics, Brighton, England (1983), p327.

39. R.Bailey et al. (ACCMOR Collaboration), Physics Letters 139B (1984) 320.

40. SLD Design Report (1984).

41. J.M.Killiany, Topics in Applied Physics 38 (1980) 147.