Preprint
RAL-P-95-008

Vertex Detectors: The State of the Art and Future Prospects

C J S Damerell

December 1995

COUNCIL FOR THE CENTRAL LABORATORY OF THE RESEARCH COUNCILS

RAL-P-95-008

VERTEX DETECTORS: THE STATE OF THE ART AND FUTURE PROSPECTS

C J S Damerell

Rutherford Appleton Laboratory

Chilton, Didcot, OX11 0QX, England

ABSTRACT

We review the current status of vertex detectors (tracking microscopes for the recognition of charm and bottom particle decays). The reasons why silicon has become the dominant detector medium are explained. Energy loss mechanisms are reviewed, as well as the physics and technology of semiconductor devices, emphasising the areas of most relevance for detectors. The main design options (microstrips and pixel devices, both CCDs and APSs) are discussed, as well as the issue of radiation damage, which probably implies the need to change to detector media beyond silicon for some vertexing applications. Finally, the evolution of key performance parameters over the past 15 years is reviewed, and an attempt is made to extrapolate to the likely performance of detectors working at the energy frontier 10 years from now.

Lectures Presented At the Twenty Third Annual Summer Institute
On Particle Physics

SLAC, Stanford, California 94309

July 10-21, 1995

TABLE OF CONTENTS

1 Introduction

There is for me a considerable sense of nostalgia in giving these lectures, since I previously gave such a series at the Summer Institute of 1984, which was especially noteworthy since it was coupled with the Pief-Fest to mark the retirement of Panofsky as Director of SLAC. Younger readers will be surprised to learn that the 1984 Institute, on the theme of the sixth quark, included evidence for the discovery of top with a mass of 40 ± 10 GeV.

In my 1984 lecture series, I suggested that these candidate top events really needed additional experimental evidence in order to be proved or disproved, and that this would best be provided by a precision vertex detector able to resolve the associated B decays. At the time, this suggestion was not taken particularly seriously. A lecture series relating to experimental methods of heavy quark detection at the same Institute made no mention of vertex detectors. Detectors with the required precision were only beginning to be used in the *fixed target* regime, and many of these were based on technologies such as bubble chambers that were manifestly not applicable to the collider environment. My own lectures made mention of techniques which have subsequently fallen into disuse for this reason. However, my main aim in those lectures was to establish a case for silicon vertex detectors in the collider environment. Our community was at that time in the early stages of planning the LEP and SLC detectors, and I focused particularly on Z^0 decays as the firm ground on which to base the case for these silicon vertex detectors. One was heavily dependent on Monte Carlo simulations of events with heavy flavour decays, where the possibilities for flavour tagging and some measure of topological vertexing could be demonstrated. Physicists at the time could be forgiven for not being wholly convinced by these simulations. Silicon detectors in those days were limited in size to a few square centimetres, were typically serviced by a huge amount of local electronics (easily accommodated in a fixed target experiment, but completely excluded in a collider), and detector reliability was a major problem. Here again, access for servicing which was easy in the fixed target environment would become much more difficult at the heart of a hermetic collider detector. In 1984, these Monte Carlo studies left on one side a host of technical problems which required many years of hard work to solve. Due to the loosely coupled R&D projects of many collaborations, the progress made since then has been immense. We now have a large variety of silicon vertex detectors in use in fixed target as well as collider experiments around the world. New designs are constantly being fabricated, and tried out in test beams. The associated local electronics has shrunk spectacularly and at the same time become much faster and more powerful.

My task is thus made easier than eleven years ago; silicon vertex detectors have become well established within the standard toolkit of high energy experiments. I no longer need rely on Monte Carlo studies to prove their usefulness; we can just look at the data. On the other hand, the array of detector types available has become somewhat bewildering, and I shall aim to provide some systematic guidance for non-experts. Furthermore, despite the fact that the proponents of silicon detectors have been able to expand their horizons, even planning in some case to displace gaseous tracking detectors with tens of square metres of strip detectors, they have begun to run into serious challenges in some vertex applications. In various hadron beam experiments, most

spectacularly at LHC at its design luminosity, silicon detectors as we now know how to build them will fail after an unacceptably short time, when placed close to the interaction region. This has stimulated a major effort with other materials of greater radiation resistance, as we shall see towards the end of these lectures.

We are seeing the beginning of a technology division between $e^+ e^-$ colliders and hadron colliders, in regard to vertex detection at the energy frontier. Both are well suited to the use of silicon at large radii, for general purpose tracking. But it is likely that at the luminosities needed for 'discovery physics' at the TeV energy scale, silicon detectors will continue to be useful for high resolution vertex studies in the $e^+ e^-$ collider environment but not at LHC.

There are clearly great advantages in remaining with the silicon technology as far as possible. A major reason for its rapid growth as a material for tracking detectors is that the *planar process* for manufacturing silicon integrated circuits has been developed to an extremely fine art. These developments are continuing at a pace which reflects the billions of dollars annually invested, for purposes which have nothing to do with scientific research let along particle physics.

Before plunging into our rather specialised topic in fine detail, it is useful to take a brief look at the overall scene of silicon devices, particularly regarding their utility as radiation detectors. For, unlike some detection materials which are not widely used outside of our field (e.g. liquid argon) silicon finds applications in a vast range of scientific sensors. We are in particle physics concerned with its use for tracking microscopes that allow us to probe the smallest and shortest lived particles in nature. Silicon devices also provide the means to see the largest and oldest structures in the universe. Between these extremes, these sensors find a vast number of diverse applications, some of great importance to mankind (e.g. in medical imaging). Technically, all these areas are closely linked, so progress in one field may be significant to many others. All these scientific applications are dwarfed by the use of silicon sensors in the mass consumer markets, notably in video cameras but with applications now extending into other areas. What makes this field particularly dynamic is the flow of ideas from people with very different aims and agendas. The next major advance for HEP detectors may come from an astronomer concerned about radiation damage to his space-based telescope, or from the designer of an output circuit able to function at HDTV readout rates. Similarly, those designing devices for HEP use, may dream up an advance that happens to be much more significant for some other field.

Why is silicon the preferred material for high precision tracking detectors, as well as for such a wide range of radiation detectors?

Firstly, a *condensed medium* is essential if point measurement precision below about 10 μm is required. Gaseous tracking detectors are limited by diffusive spreading of the liberated electron cloud to precision of typically some tens of microns. Such detectors are entirely adequate for a host of particle tracking applications, but not for precision vertex detectors. Having established the need for a condensed medium, one should in principle consider liquids. There was some work done on high precision liquid xenon tracking detectors in the '70s [1] but there were many problems, not least of which was maintaining purity in conditions where the high

mobilities of many contaminants rendered them particularly potent. In contrast, silicon wafers refined to phenomenal purity levels can then be sawn, exposed to the atmosphere, and assembled in complex geometries, with no degradation of their bulk electron lifetime characteristics. For these reasons, silicon and other solids are generally to be preferred, as opposed to liquids, for high precision tracking purposes. There are, however, many possible solid state detection media, so why pick silicon?

Silicon has a band gap of 1.1 eV, *low* enough to ensure prolific production of liberated charge from a minimum-ionizing particle, hereafter referred to as a MIP, (about 80 electron-hole pairs per micron of track length) but *high* enough to avoid very large dark current generation at room temperature (kT at room temperature = 0.026 eV). Being a low Z element of excellent mechanical properties (high modulus of elasticity) makes silicon an ideal material for use in tracking detectors where multiple scattering is of concern. This is nearly always the case in vertex detectors where tracks need to be extrapolated to the interaction region, and the dynamics of the fragmentation process ensures that even at the highest CM energies, many of the particles produced are of relatively low energy.

Besides these detector-related reasons, one has the vast IC technology developed specifically for this material. Silicon is currently unique in the combination of assets it brings with it; the growth of huge crystals of phenomenal purity, the possibility of n- and p-type doping, the possibility of selective growth of highly insulating layers (SiO_2 and Si_2N_3), and the possibility of doing all these using microlithographic techniques, allowing feature sizes of around 1 micron (and falling with time). A very readable account of the remarkable human stories associated with these amazing developments is to be found in George Gilder's book on the subject [2]. Very small feature sizes are of course precisely what one requires in order to construct detectors of precision below 10 microns. Overall, the art of producing integrated circuits is probably by far the most sophisticated, fastest developing area of technological growth in the history of mankind. Without these developments, silicon as a detector of nuclear radiation would have remained a minor player, subject to arcane production procedures, of limited use for the spectroscopy of low energy gamma rays, and wholly inappropriate for particle tracking purposes.

Though the scientific applications are of great importance, they are dwarfed by the use of silicon detectors for mass market consumer products and commercial interests. Accurate figures are not readily available, but it seems that approximately $100M per year is spent on R&D of CCDs for domestic video and still cameras. These are interline transfer devices of no direct use for most scientific imaging applications. About $10M is spent on CCD development for medical and other scientific imaging applications (mostly X-rays). Silicon devices specifically aimed at particle tracking (microstrip detectors, CCDs and active pixel sensors, hereafter referred to as APS devices) probably attract only $1M (order of magnitude) in R&D per year.

Even the consumer market for silicon sensors is dwarfed by the really hot commercial areas. For example, it was recently reported that NEC demonstrated a 1 Gbit DRAM. Production devices are expected to follow in three year's time, after the expenditure of *a further* $1.5B of R&D funding. Much of this will go in the development of sub-micron manufacturing capability, which ultimately will benefit the particle physics

instrumentation community. We can eventually look forward to *sub-micron* tracking precision with *sub-nanosecond* timing information. However, the pace of such developments will be determined by the major players outside our own field, and there will inevitably be a time lag of several years between a technology being available for mass produced ICs and it being affordable for our purposes.

While the silicon processing infra-structure and R&D for a specific device can be enormously expensive, once production begins the costs can be modest. The ingredients of integrated circuits (sand, air, aluminium) are ridiculously cheap, and this benefit can be seen dramatically in large production runs. For example, SONY produce approximately 5 million CCDs per year for the domestic video camera market, at a production cost of only around $10, including the micro-lens and colour filter system. This is a truly amazing achievement, as you can convince yourself by just looking through a microscope at one of these devices.

In summary, the match between silicon (and its attendant technologies) to the aspirations of the experimentalist wishing to construct tracking detectors of the highest possible precision, is evident. Were it not for the problems of radiation damage (which are most serious in the context of hadron colliders), there is little doubt that our field would by now have standardised completely on this material for vertex detection. Some time ago, test devices even surpassed photographic nuclear emulsions in precision, and with all the advantages of electronic readout. The challenge of hadron machines has stimulated some brave souls to undertake the monumental task of achieving similar technical performance using more radiation resistant materials than silicon. They have of course to solve the problems not only of the detectors but also of the local electronics. We shall take a brief look at what they are doing in Section 7 of this paper. Other than that section, we shall devote ourselves exclusively to a discussion of silicon detectors and electronics.

2 Energy Loss of High Energy Charged Particles in Silicon

High energy charged particles traversing crystalline silicon can lose energy in two ways. Firstly, by ionization of the atomic electrons. This simple picture becomes rather more complex in regard to the valence electrons, as we shall see. The second energy loss mechanism (the so-called non-ionizing energy loss or NIEL) consists of displacement of silicon atoms from the crystal lattice, mostly by the process of Coulomb nuclear scattering. Only if the energy transfer to the nucleus exceeds approximately 25 eV can the atom be displaced from its lattice site. Below that, the energy is dissipated by harmless lattice vibrations. This implies an effective threshold energy for displacement damage with incident electrons (for example) of around 250 keV. Displacement of silicon atoms to *interstitial* positions (creating a *vacancy* in the lattice where the atom had previously been located) is one of the main radiation damage mechanisms. For a high energy particle, the fraction of energy loss going into the NIEL mechanism is relatively small, but the cumulative effects on the detector performance can be severe.

A detector placed in a neutron flux experiences no signal from primary ionization, but the interactions can cause a high level of NIEL in view of the large neutron-silicon scattering cross-section. For both charged hadrons and neutrons, other mechanisms of energy loss and radiation damage exist, notably neutron capture followed by nuclear decay, and inelastic nuclear scattering. The effects of non-ionizing energy loss on silicon detectors are considered in Section 6. In this section, we focus on the ionization energy loss only.

2.1 Simplified Treatment

Let us first imagine all the atomic electrons to be free, as if the crystal consisted of the silicon nuclei neutralised electrically by a homogeneous electron plasma. As a charged particle traverses the material, it loses energy by collisions (Coulomb scattering) with the electrons. Close collisions, while rare, will result in large energy transfers, while the much more probable distant collisions give small energy transfers. The process can be thought of classically in terms of the impulse generated by the attractive or repulsive Coulomb interaction between the projectile and the electron. The nett impulse will be a kick transverse to the direction of travel of the projectile (see fig. 1). The greater probability of remote collisions arises simply from the greater volume of material available for collisions with a given impact parameter range, as the corresponding cylinder (of radius equal to the impact parameter) expands. In this simple case, the probability for a collision imparting energy E to an atomic electron is given by the Rutherford cross-section

$$\frac{d\sigma_R}{dE} = \frac{2\pi {q_e}^4}{m_e c^2 \beta^2} \times \frac{1}{E^2} \qquad (2.1)$$

where q_e and m_e are the charge and mass of the electron.

Note the mass of the struck particle in the denominator. This explains why scattering off the silicon nuclei, which are much more massive, causes very little energy loss, though these collisions do make the major contribution to the deviation in angle of the incident particle trajectory, via the process of multiple nuclear

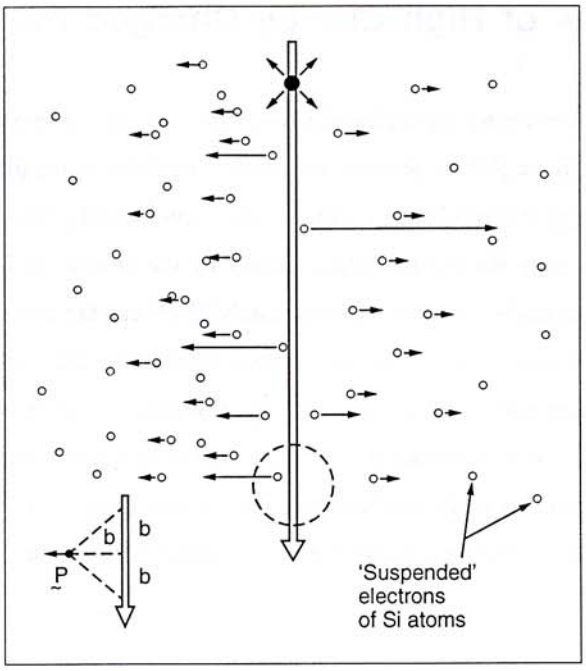

Fig. 1 Passage of charged particle through matter. Close collisions (electrons with small impact parameter b, shown by the inset) receive a powerful transverse impulse. Distant electrons receive a weak impulse.

Coulomb scattering. Also, for sufficiently large momentum transfers, these nuclear collisions contribute to the NIEL referred to above.

We are interested in evaluating the *mean* energy loss and also the *fluctuations*, for traversal of a given thickness detector. An apparently simple approach would be to perform the integration over all E to obtain the mean energy loss, and to run a Monte Carlo calculation with multiple traversals to determine the energy loss distribution (straggling formula). However, we see that the integral diverges like $1/E$. The stopping power of this free-electron plasma would indeed be infinite, due to the long-range Coulomb interaction. In practice, the electrons are *bound* and this prevents very low energy transfers to the vast number of electrons which are distant from the particle trajectory. This divergence is conventionally avoided by introducing a semi-empirical cutoff (binding energy) $E_{\min}$ which depends on the atomic number Z of the material. This is necessarily an approximate approach, since (for example) it ignores the fact that the outer electrons are bound differently in gaseous media than they are in solids. We shall need a more refined treatment to handle the cutoff in collisions with small energy transfer.

However, the Rutherford formula (with one small correction) is extremely useful as regards the close collisions, which are most important in defining the fluctuations in energy loss in 'thick' samples (greater than approximately 50 μm of silicon, for example). The required correction is the upper cutoff $E_{\max}$ in energy transfer imposed by the relativistic kinematics of the collision process. If the projectile mass is much greater than m_e, we have $E_{\max} = 2m_e c^2 \beta^2 \gamma^2$. Due to the $1/E^2$ term in the Rutherford formula, we find that there is for each sample thickness, an energy transfer range in which the integrated probability of such transfers

through the sample falls from almost unity to nearly zero. The Poisson statistics on energy transfers in this range gives rise to fluctuations on the overall energy loss for each traversal. Thus the overall energy loss distribution consists of an approximately Gaussian core plus a high tail, populated by traversals for which a few close collisions occurred, each generating several times the mean energy loss. While the energy transfer region in which the probability function falls almost to zero is dependent on the sample thickness, this merely introduces an overall scale factor, so the *form* of the overall energy loss distribution is constant (the famous Landau distribution) over a wide range of detector thicknesses.

The rare close collisions with energy transfer greater than approximately 10 keV generate δ-electrons of significant range, which may be important in tracking detectors due to their potential for degrading the precision. For these close collisions, all atomic electrons behave as if they are free and the Rutherford formula may be used with confidence.

For thin samples, the energy loss fluctuations are not adequately handled by the Rutherford formula with cutoffs $E_{\min}$ and $E_{\max}$. In this case, the bulk of the energy loss arises from low energy transfer collisions for which the binding of the atomic electrons must be handled in detail. We shall now consider the improved treatment of this case, specifically for crystalline silicon, though the same principles apply in general.

2.2 Improved Treatment

We note that energy loss is a discrete quantum mechanical process. We shall see that for very thin samples, a particle has even a finite probability of traversing the detector with no energy deposition at all.

For the low probability close collisions, as noted above, it is valid to consider all atomic electrons as free, and the Rutherford formula applies. Ejected electrons of energies greater than approximately 10 keV will release further atomic electrons along their path. See [3,4] for a detailed treatment. For our purposes, it is sufficient to note that the ultimate products that concern us are electrons, promoted into the conduction band of the material and holes (vacancies in the valence band), and that the generation of each electron-hole pair requires a mean creation energy W of approximately 3.6 eV. The precise value depends weakly on the temperature, see fig. 2, and reflects the temperature dependence of the silicon band gap. Since this is around 1.1 eV, we note that electron-hole generation is a somewhat inefficient process; approximately 2/3 of the energy transferred from the primary (hot) electrons gives rise to phonon generation, eventually appearing as heat in the detector. Beware, this has nothing to do with the non-ionizing energy loss (NIEL) referred to in the introduction to this section! Phonon generation (in contrast to NIEL) is a benign process which does not disrupt the crystal lattice and is usually ignored other than by enthusiasts for bolometric detectors. For our purposes, the δ-electrons ejected in close collisions can be considered to generate further electron-hole pairs at a mean rate of one per 3.6 eV of energy loss, *locally* on the track of the projectile, or *distributed* in the case that the δ-electron range is significant.

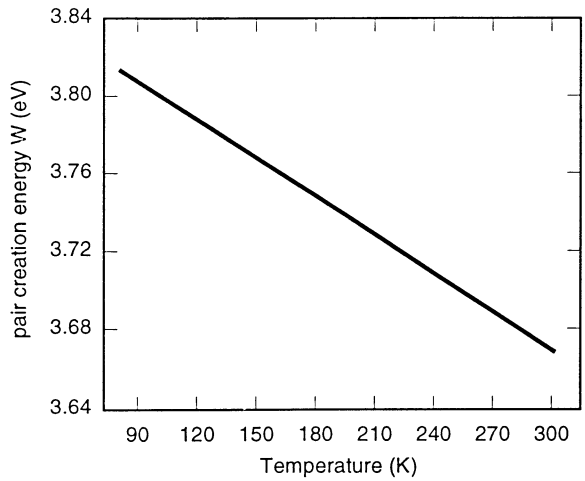

Fig. 2 Temperature dependence of the pair-creation energy W in silicon.

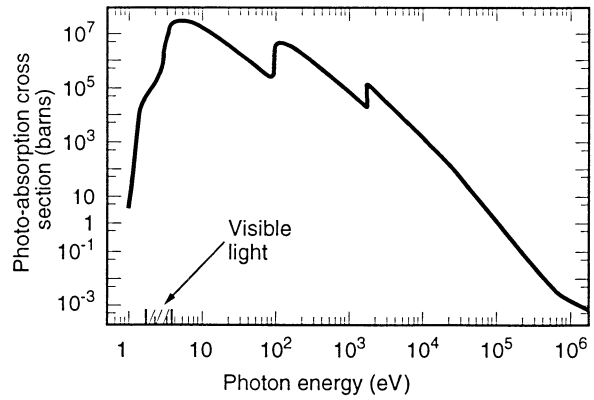

Fig. 3 Photo-absorption cross-section of silicon versus photon energy.

Qualitatively, the effect of the binding of the atomic electrons is to generate resonance-like enhancements in the energy loss cross-section, above the values expected from the Rutherford formula. The K-shell electrons produce an enhancement in the 2 to 10 keV range, the L-shell in the 100 eV to 1 keV range, and the M-shell a resonance at around 20 eV. Below this resonance, the cross-section rapidly falls to zero, in the region around 15 eV where the Rutherford formula would be cut off by the empirical ionization threshold energy.

The most satisfactory modern treatment proceeds from the energy-dependent photo-absorption cross-section (a clean *point-like* process in the terminology of solid state physics). This is of course closely linked to the energy loss process for charged particles, which fundamentally proceeds via the exchange of virtual photons. Combining photo-absorption and EELS (electron energy loss spectroscopy) data, Bichsel [5] has made a precise determination of the MIP energy loss cross-section for silicon. The most subtle effects are connected with the valence (M-shell) electrons. These behave as a nearly homogeneous dense gas (plasma) embedded in a fixed positive charge distribution. The real or virtual photons couple to this by generating longitudinal density oscillations, the quantum of which is called a *plasmon* and has a mean energy of 17 eV. The plasmons de-excite almost entirely by electron-hole pair creation. These somewhat energetic charge carriers are referred to as 'hot carriers'. Like the δ-electrons produced in the close collisions, they lose energy by thermal scattering, optical phonon scattering and ionization. The topic of hot carriers is a major area of research, but for our purposes (as with the δ-electrons) we can ignore the details, since the end product that concerns us is again electron-hole pair creation at a rate of one per 3.6 eV of primary energy deposition. Fig. 3 shows the photo-absorption cross-section for silicon. The plasmon excitation is responsible for the extremely large cross-section in the ultra-violet. It is by virtue of the low energy tail of this cross-section in the visible that silicon has its optical sensing applications. The material becomes almost perfectly transparent once the photon energy falls below the 1.1 eV band gap energy.

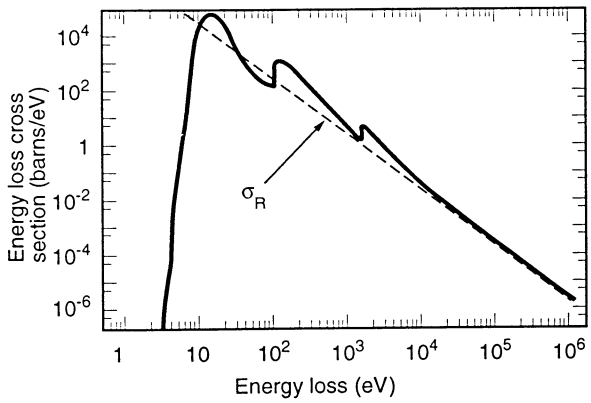

Fig. 4 Energy loss cross-section for minimum-ionizing particles in silicon, versus energy loss in primary collision. Rutherford cross-section σ_R is also plotted.

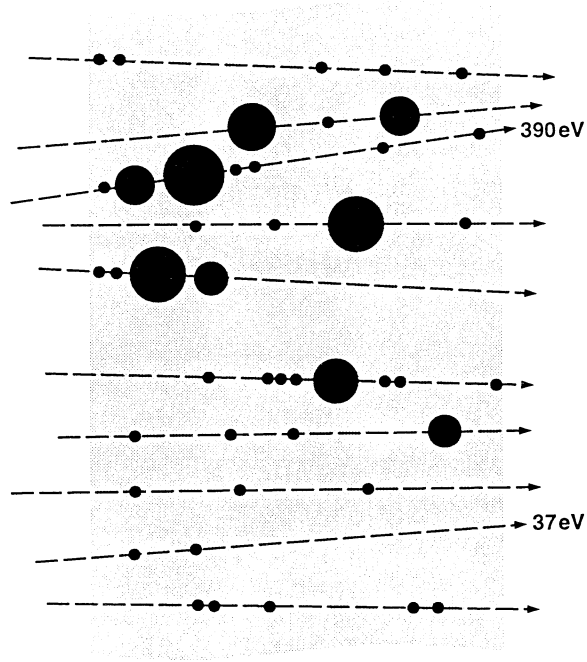

Fig. 5 Monte Carlo calculation of energy deposition in 1 μm thick silicon detector. Area of a blob represents the energy deposited in each primary collision process.

The closely related energy loss cross-section for a MIP is show in fig. 4. Note that on average it exceeds the Rutherford cross-section by a factor of several in the energy range 10 eV to 5 keV. Above 10 keV, it is very close to the Rutherford value. By integrating the different components of this cross-section, we can deduce the total mean collision rates associated with the different processes. These are as follows:

Electrons	**Collision probability per micron**
K (2)	8.8×10^{-3}
L (8)	0.63
M (4)	3.2

Thus, despite the fact that on average a slice of silicon 1 μm in thickness will yield 80 electron-hole pairs, the Poisson statistics on the *primary* process (on average 3.8 collisions per micron) clearly implies a very broad distribution, with even a non-negligible probability of zero collisions, i.e. absolutely no signal. For thin samples, a correct statistical treatment of the primary process is essential if realistic energy loss (straggling) distributions are to be calculated. Their shapes are a strong function of the sample thickness, quite unlike the thickness-independent Landau distribution. The situation is depicted graphically in fig. 5. The area of each circle represents energy loss in a primary collision process. Those of smallest size correspond to plasmon excitation, while the larger ones represent the ionization of L-shell electrons. For these 10 randomly selected tracks, the total energy deposition in the sample ranges from 37 eV to 390 eV.

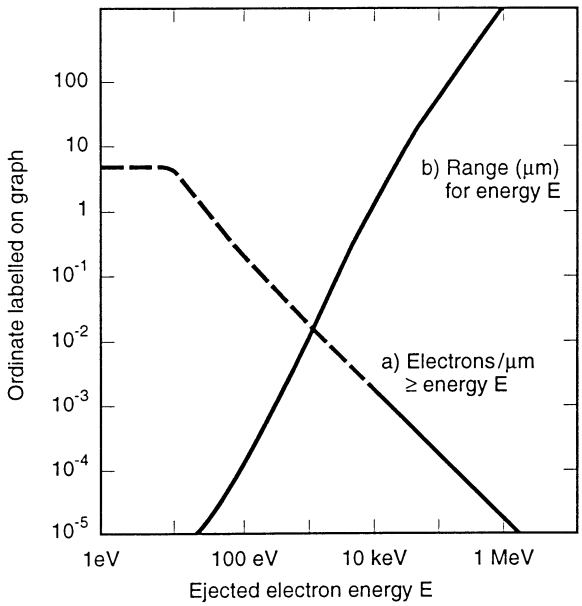

Fig. 6 (a) shows the number of electrons per micron of MIP track above a given energy, and (b) shows the range in silicon corresponding to that energy.

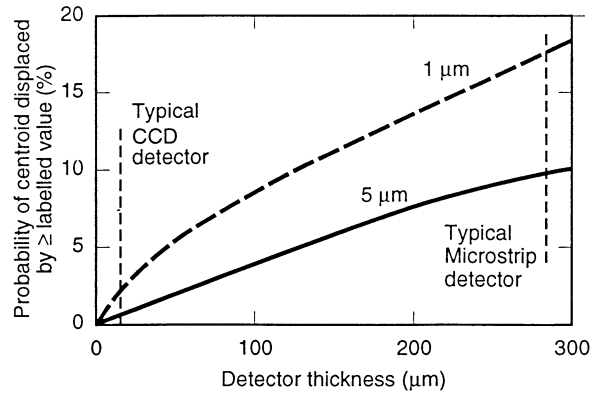

Fig. 7 Detector precision limitations from δ-electrons for tracks of normal incidence, as a function of detector thickness.

2.3 Implications for Tracking Detectors

For high precision tracking, there are clear advantages in keeping the silicon detector as thin as possible. A physically thin detector is optimal as regards multiple scattering. A detector with the thinnest possible active region (which may be less than the physical thickness, as we shall see) is optimal as regards point measurement precision, for *two distinct reasons.*

For normal incidence tracks, the concern arises from δ-electrons of sufficient range to pull the centroid of the charge deposition significantly off the track. Fig. 6 (a) is an integral distribution of the number of primary electrons per micron of energy greater than a given value, and fig. 6 (b) shows the range of electrons of that energy in silicon. The range becomes significant for high-precision trackers for E greater than approximately 10 keV, for which the generation probability is less than 0.1% per micron. Thus, a detector of thickness 10 μm is much less likely to yield a 'bad' co-ordinate than one of thickness 100 μm.

If the magnitude of the energy deposition in the detector is measured (by no means always possible) some of the bad co-ordinates will be apparent by the abnormally large associated energy. They could then be eliminated by a cut on the energy deposit, but this usually leads to unacceptable inefficiency and is rarely implemented. The situation is summarised in fig. 7, which indicates the probabilities of the centroid for a track being pulled by more than a certain value (1 μm and 5 μm) as a function of detector thickness. The advantage of a thin active medium is apparent.

The second reason for preferring detectors to be as thin as possible applies to the case of angled tracks. In principle (and occasionally in practice) it may be possible to infer the position of such a track by measuring the

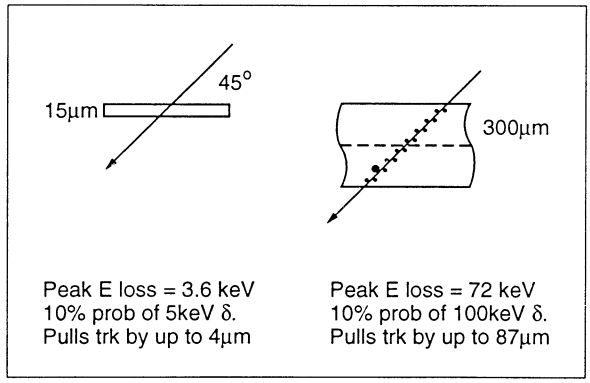

Fig. 8 Effect of energy loss fluctuations on detector precision for angled tracks.

entry and exit points in the detector, but more usually the best one can do is to measure the centroid of the elongated charge distribution, and take this to represent the track position as it traversed the detector mid-plane. In this case, large fluctuations in the energy loss (due to ejection of K- and L-shell electrons and δ-electrons) may be sufficient to cause serious track pulls for thick detectors. This is illustrated in fig. 8. In the thin detector there is a 10% probability of producing a δ-electron which, if it occurs near one end of the track, pulls the co-ordinate from its true position by 4 μm. In the thick detector, there is the same probability of producing a δ-electron which can pull the co-ordinate by 87 μm.

However, our enthusiasm for thin active detector layers must be moderated by the primary requirement of any tracking system, namely a high efficiency per layer. Fig. 9 (based on reference [5]) illustrates the problem we could already anticipate from fig. 5. For very thin detectors (e.g. 1 μm Si), we see a very broad energy loss distribution with peaks corresponding to 0, 1, 2, ... plasmons excited, followed by a long tail extending to very large energy losses. An efficient tracking detector could never be built with such an active layer. Even at 10 μm silicon thickness, the true distribution is much broader than Landau and has a dangerous low tail. By 300 μm, the Landau distribution gives an adequate representation. Thus, while very thin detectors are ideal from the viewpoint of tracking precision, great care must be taken to assure that *system noise* allows a sufficiently *low threshold* to achieve the desired detector efficiency.

2.4 Summary

The valence electrons of silicon are very easily excited into plasmon oscillations from which they dislodge a small number (typically 5) of electrons into the conduction band. A MIP thus creates a fine trail of electron-hole pairs along its track. The quantity W (energy needed to create an electron-hole pair) is approximately 3.6 eV, but depends on the band gap and hence (weakly) on the temperature. This energy loss process allows in principle unprecedented precision (much better than 1 μm) compared even to a nuclear emulsion (which needs typically a 400 eV δ-electron to blacken a grain). One does need to be prepared to exclude the measurements associated with large energy deposition, but these are rare in thin detectors.

How can this potential performance be achieved in practice? Standard IC processing (the planar technology) provides us with a host of suitable tools. This is after all one of the few areas of engineering in which sub-

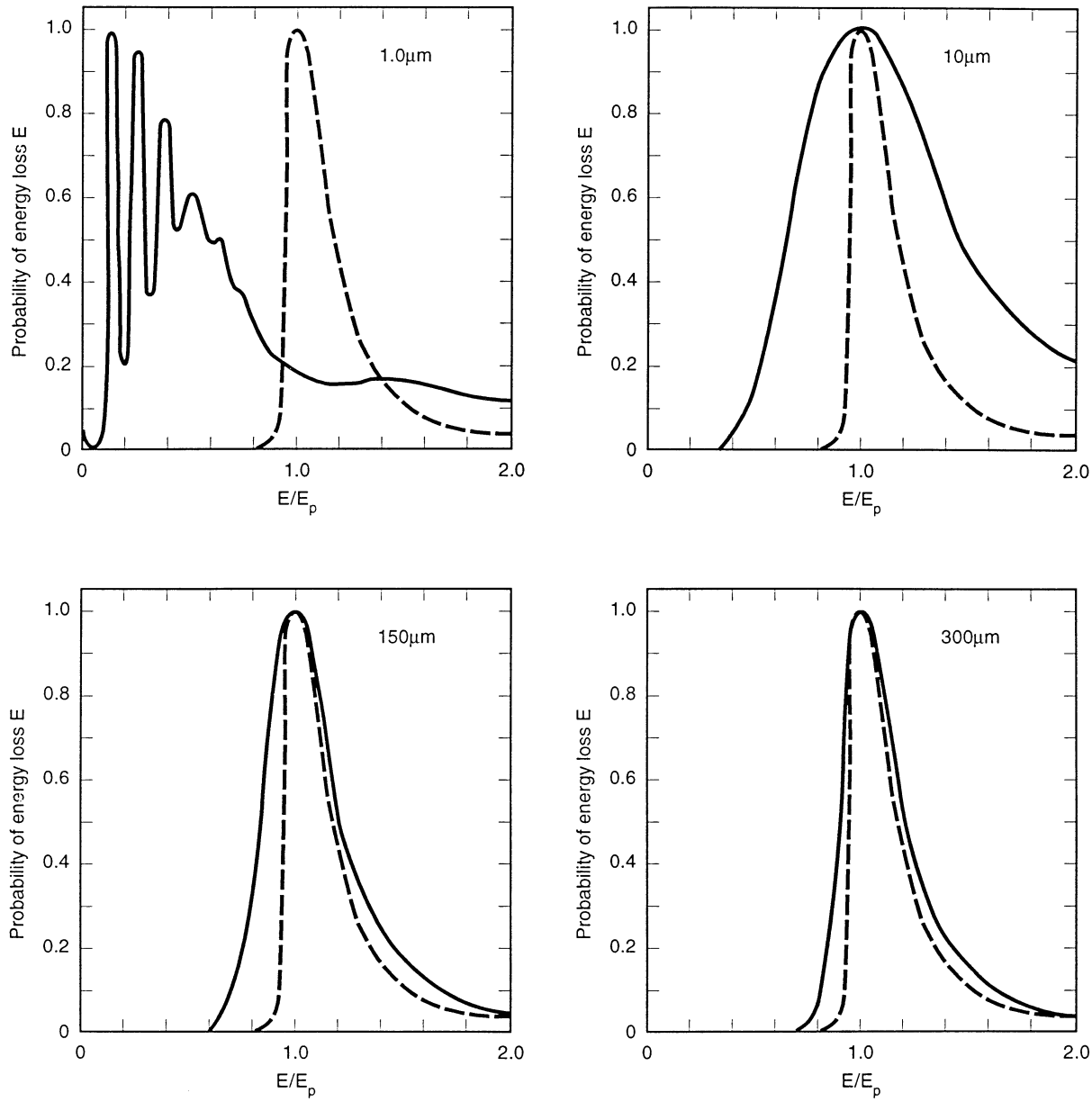

Fig. 9 Energy loss distributions for various silicon detector thicknesses, with (in each case) a Landau distribution for comparison. The separate peaks corresponding to 0, 1, 2 ... plasmon excitation are already merged by a thickness of 10 μm.

micron tolerances are now standard practice. In Sections 4 and 5, we shall explore some types of detectors currently available. But first we consider some of the basic properties of silicon which allow us in principle to collect and sense the signal charges we have been discussing in this section.

3 Physics and Properties of Semiconductors

Gaseous silicon has a typical structure of atomic energy levels (see fig. 10). It has an ionization potential of 8.1 eV, i.e. it requires this much energy to release a valence electron, compared with 15.7 eV for argon. As silicon condenses to the crystalline form, the discrete energy levels of the individual atoms merge into a series of energy bands in which the individual states are so closely spaced as to be essentially continuous. The levels previously occupied by the valence electrons develop into the *valence band*, and those previously unoccupied become the *conduction band*. Due to the original energy level structure in gaseous silicon, it turns out that there is a gap between these two bands. In conductors there is no such gap; in semiconductors there is a small gap (1.1 eV in silicon, 0.7 eV in germanium) and in insulators there is a large band gap. In particular, the band gap in silicon dioxide is 9 eV. This makes it an excellent insulator and, coupled with the ease with which the surface of silicon can be oxidised in a controlled manner, accounts partly for the pre-eminence of silicon in producing electronic devices.

We shall denote as E_V and E_C the energy levels of the top of the valence band and the bottom of the conduction band (relative to whatever zero we like to define). The energy needed to raise an electron from E_C to the vacuum E_0 is called the electron affinity. For crystalline silicon this is 4.15 eV.

3.1 Conduction in Pure and Doped Silicon

To understand the conduction properties of pure silicon, the *liquid analogy* is helpful. This is illustrated in fig. 11: (a) shows the energy levels in silicon under no applied voltage with the material at absolute zero temperature. All electrons are in the valence band, and under an applied voltage (b) there is no change in the population of occupied states, and so no flow of current; the material acts like an insulator. At a high temperature (c) a small fraction of the electrons are excited into the conduction band, leaving an equal number of vacant states in the valence band. Under an applied voltage (d) the electrons in the conduction band can flow to

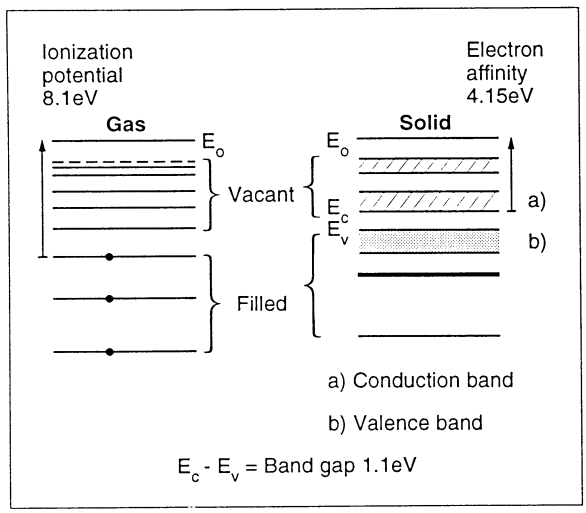

Fig. 10 Sketch of allowed energy levels in gaseous silicon which become energy bands in the solid material.

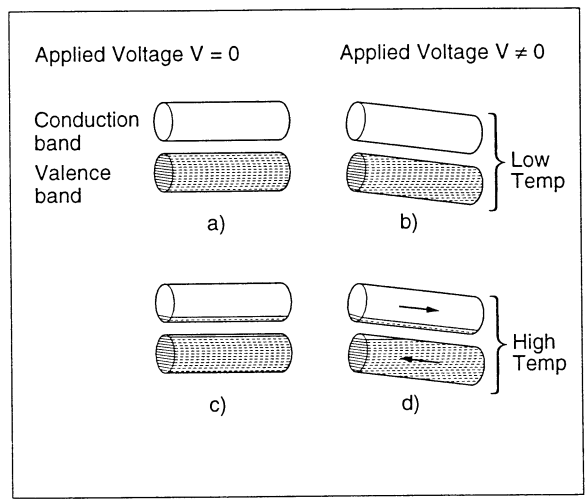

Fig. 11 Liquid analogy for a semiconductor.

the right and there is a re-population of states in the valance band which can be visualised as the left-ward movement of a bubble (holes) in response to the applied voltage.

Now kT at room temperature is approximately 0.026 eV. This is small compared with the band gap of 1.1 eV, so the conductivity of pure silicon at room temperature is very low. To make a quantitative evaluation, we need to introduce the Fermi-Dirac distribution function $f_D(E)$ which expresses the probability that a state of energy E is filled by an electron. Fig. 12 (a) shows the form of this function

$$f_D(E) = \frac{1}{1+\exp\left(\frac{E-E_f}{kT}\right)} \tag{3.1}$$

Note that E_f, the Fermi level, is the energy level for which the occupation probability is 50%. Fig. 12 (b) sketches the density of states $g(E)$ in silicon. The concentration of electrons in the conduction band is given by the product $f_D g$, and the density of holes in the valence band by $(1-f_D)g$, as shown in fig. 12 (c). In pure silicon, the Fermi level is approximately at the mid-band gap, and the concentrations of electrons and holes are of course equal. These concentrations, due to the form of f_D, are much higher in a narrow band gap semiconductor, fig. 12 (d) than in a wide gap material, fig. 12 (e).

So far we have been discussing pure (so-called *intrinsic*) semiconductors. Next we have to consider the *doped* or extrinsic semiconductors. These allow us to achieve high concentrations of free electrons (n-type, fig. 12 (f)), or of holes (p-type, fig. 12 (g)), by moving the Fermi level very close to the conduction or valence band edge. The procedure for doing this is to replace a tiny proportion of the silicon atoms in the crystal lattice by dopant atoms with a different number of valence electrons.

Fig. 13 shows the lattice structure characteristics of diamond, germanium and silicon crystals. Silicon, with four valence electrons, forms a very stable crystal with covalent bonds at equal angles in space. It is possible (e.g. by ion implantation) to introduce a low level of (for example) pentavalent impurities such as phosphorus. By heating (*thermal activation* as it is called) the phosphorus atoms can be induced to take up lattice sites in the crystal. For each dopant atom, four of its electrons share in the covalent bonding with neighbouring silicon atoms, but its fifth electron is extremely loosely bound. At room temperature this electron would be free, and hence available for conduction in a sea of fixed positive charge (the phosphorus ions, present at precisely the same average density as the liberated electrons). At absolute zero, all valence electrons would be bound and the phosphorus-doped (n-type) silicon effectively an insulator. The mathematical description of the effect of doping in silicon is to retain the Fermi-Dirac distribution function, but to raise the Fermi level (50% occupation probability) close to the binding energy of the fifth electron, i.e. close to the conduction band edge. The population of those electrons within the conduction band is again given by the overlap of the Fermi-Dirac distribution function (now shifted in energy) and the density of states in the conduction band. Except at very low temperatures (where the Fermi-Dirac function is extremely sharp) the result is a high density of electrons (*majority carriers*) and a negligible density of holes (*minority carriers*) in the n-type material in equilibrium, as shown in fig. 12 (f).

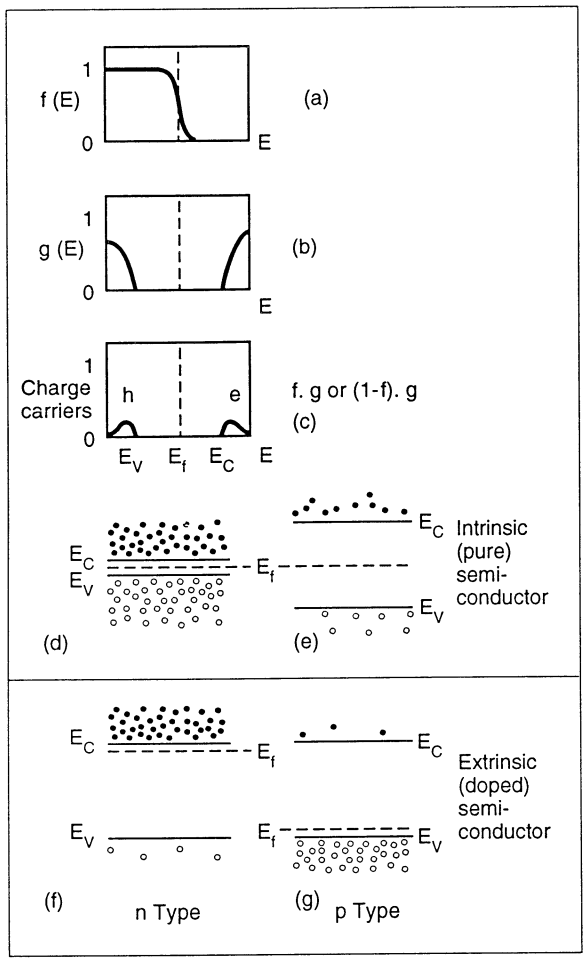

Fig. 12 (a) Fermi-Dirac distribution function. The slope increases as the temperature is reduced. (b) Density of states below and above forbidden band gap. (c) Concentration of charge carriers (electrons and holes) available for conduction. (d) and (e) Charge carrier distributions in narrow and wide band gap semiconductors. (f) and (g) Charge carrier distributions in *n*- and *p*-type semiconductors.

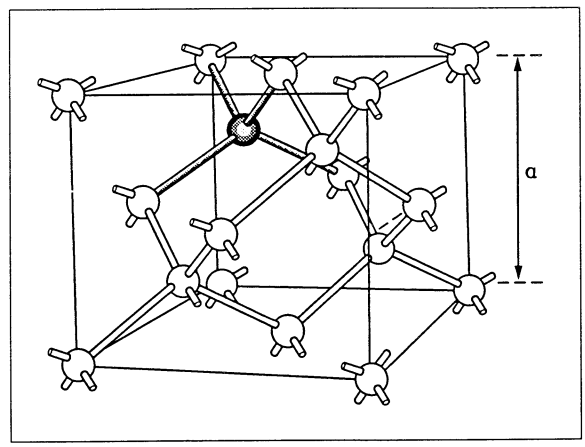

Fig. 13 Lattice structure of diamond, germanium, silicon, etc. where *a* is the lattice constant.

Alternatively, silicon may be doped with trivalent impurities such as boron. In this case, three strong covalent bonds are formed, but the fourth bond is incomplete. This vacancy (hole) is easily filled by an adjacent electron. Thus, as in the intrinsic material, holes behave as reasonably mobile positively charged carriers in a sea of fixed negative charge (the boron atoms with an additional electron embedded in the fourth covalent bond). The carrier concentrations (now with holes as majority carriers) are give by shifting the Fermi-Dirac distribution to within the hole binding energy i.e. close to the valence band edge as shown in fig. 12 (g).

The general situation regarding doped silicon is sketched in fig. 14, which indicates the energy levels corresponding to various commonly used dopant atoms. Pentavalent atoms are referred to as *donors* and trivalent atoms as *acceptors*. Note that the carriers are bound by only approximately 0.045 eV in the common *n*- and *p*-type dopants phosphorus and boron, compared to kT at room temperature of 0.026 eV.

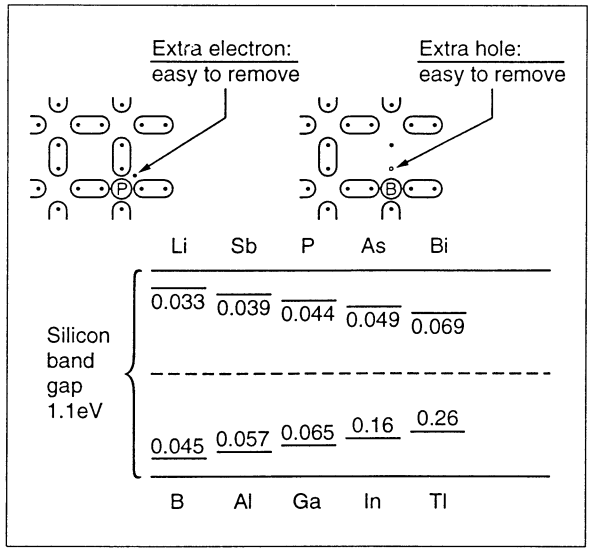

Fig. 14 Sketch of bond occupation in doped silicon (upper) and energy levels within the band gap corresponding to various *n*- and *p*-type dopants (lower). Levels of acceptor atoms are conventionally measured from the top of the valance band, and levels of donor atoms are measured from the bottom up the conduction band.

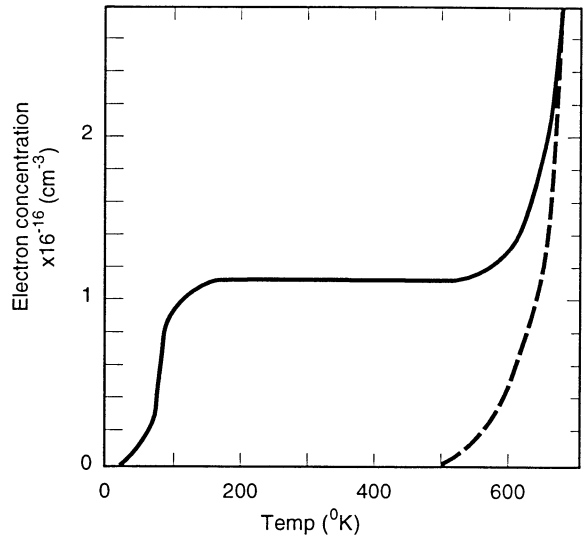

Fig. 15 Electron concentration versus temperature for *n*-type (arsenic doped) silicon. The dashed curve shows the concentration for intrinsic material.

Fig. 15 shows the concentration of electrons in *n*-type silicon (1.15×10^{16} arsenic dopant atoms per cm^3) as a function of temperature. Below about 100 K one sees the phenomenon of *carrier freeze-out*, loss of conductivitydue to the binding of the donor electrons. This is followed by a wide temperature range over which the electron concentration is constant, followed above 600 K by a further rise as the thermal energy becomes sufficient to add a substantial number of intrinsic electrons to those already provided by the dopant atoms. These will of course be accompanied by an equal concentration of mobile holes. The general behaviour show in fig. 15 is typical of all doped semiconductors.

The *resistivity* ρ of the material depends not only on the concentration of free holes and electrons but also on their *mobilities*. As one would intuitively expect, the hole mobility is lower than that for electrons. Both depend on temperature and on the impurity concentration. At room temperature, in lightly doped silicon, we have

electron mobility $\mu_n = 1350\ cm^2\ (V\ s)^{-1}$

hole mobility $\mu_p = 480\ cm^2\ (V\ s)^{-1}$

and the resistivity is given by

$$\rho = \frac{1}{q_e(\mu_n . n + \mu_p . p)} \qquad (3.2)$$

[n and p are the electron and hole concentrations]

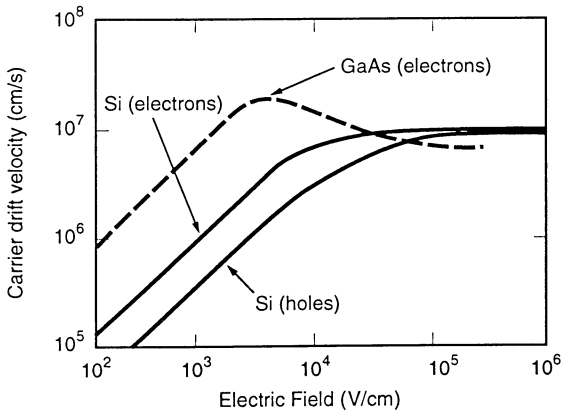

Fig. 16 Carrier drift velocity (electrons and holes) for silicon, and electron velocity for gallium arsenide as function of electric field in the material.

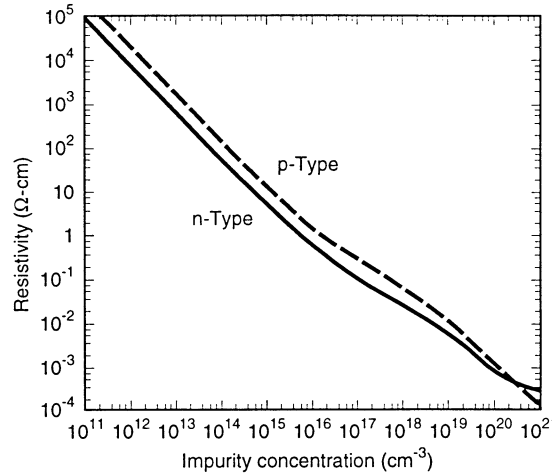

Fig. 17 Resistivity of silicon at room temperature as a function of acceptor or donor impurity concentration.

For pure silicon at room temperature $n_i = p_i = 1.45 \times 10^{10}$ cm^{-3} which gives $\rho_i = 235$ KΩ cm.

The carrier drift velocity (v_p for holes and v_n for electrons) is related to the mobility by $v_{p,n} = \mu_{p,n}\mathcal{E}$ where $\mathcal{E}$ is the electric field strength. This relationship applies only up to a maximum field, beyond which saturation effects come into play and one enters the realm of 'hot carriers' which lose energy by impact ionization (creation of additional electron-hole pairs). Fig. 16 shows the situation for silicon, as well as the fact that nearly ten times higher electron drift velocities are achievable in gallium arsenide, which therefore has the potential for much faster electronic circuits.

The ionization rate is defined as the number of electron-hole pairs created per unit of distance travelled, by an electron or hole. It depends primarily on the ratio $q_e\mathcal{E} / E_I$ where E_I is the effective ionization threshold energy, damped by terms reflecting the energy loss of carriers by thermal and optical phonon scattering, see reference [6]. For silicon, E_I is approximately equal to W (3.6 eV) for electrons and 5.0 eV for holes. The ionization rate becomes significant for electric fields in the range 10^5 to 10^6 V/cm in silicon, leading to the saturation of carrier drift velocity shown in fig. 16.

The resistivity as a function of dopant concentration is shown in fig. 17. For silicon detectors, as will be explained in the next section, we are often concerned with unusually high resistivity material, some tens of KΩ cm. From fig. 17, one sees for example that 20 KΩ cm p-type material requires a dopant concentration of 5×10^{11} per cm^3. Remembering that crystalline silicon has 5×10^{22} atoms per cm^3, this implies an impurity level for the *predominant* impurities of 1 in 10^{11}. Even in the highly developed art of silicon crystal growing, this presents a major challenge. The resistivity noted above in connection with pure silicon (over 200 KΩ cm) is entirely unattainable in practice. Very high resistivity n-type silicon can be produced in the form of *compensated* material. The most uniformly doped material which can be grown is (for technical reasons) p-type, and this (with a resistivity of about 10 KΩ cm) is used to start with. It is then turned into n-type material by the procedure known as neutron doping. The crystal is irradiated with slow neutrons and by means of the reaction

$$Si^{30} + n \rightarrow Si^{31}$$

followed by $$Si^{31} \rightarrow P^{31} + \beta^- + \bar{\nu}$$

is turned into n-type material. The resistivity is monitored and the irradiation ceases when this, having passed through a maximum, falls to the required value. In this way, material of resistivity as high as 100 KΩ cm can be made. Achieving reasonable uniformity through the wafer of such a high resistivity is obviously extremely difficult.

We now consider more quantitatively the relationship between the carrier concentration and the Fermi level. The number of conduction band states occupied by electrons is given by

$$n = \int_{E_c}^{E_t} N(E) f_D(E) dE$$

E_c and E_t are the energy at the bottom and top of the conduction band; $f_D(E)$ is the function (3.1); $N(E)$, the density of states, is given by the band theory of solids and is proportional to $(E - E_c)^{1/2}$. For the commonly encountered situation where Boltzmann statistics applies, for which the Fermi level is at least several times kT below E_c, the above integral can be approximately evaluated to yield

$$n = N_c \exp\left(-\frac{E_c - E_f}{kT}\right) \tag{3.3}$$

N_c is called the effective density of states. Its meaning is not as intuitively clear as the simple density of states $N(E)$; unlike $N(E)$ it is temperature dependent, being proportional to $T^{3/2}$.

The equivalent approximation for the hole concentration is

$$p = N_v \exp\left(-\frac{E_f - E_v}{kT}\right) \tag{3.4}$$

For *intrinsic* semiconductors, thermal agitation excites electrons from the valence band to the conduction band, leaving an equal number of holes in the valence band. In this case $n = p = n_i$, where n_i is the intrinsic carrier density. There is a dynamic equilibrium between thermal generation on the one hand, and recombination of electrons in the conduction band with holes in the valence band on the other. The neutrality condition obtained by equating (3.3) and 3.4) leads to

$$E_f = E_i = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln\left(\frac{N_v}{N_c}\right) \tag{3.5}$$

Thus the Fermi level of an intrinsic semiconductor lies very close to the middle of the band gap. The intrinsic carrier density is given from (3.3) and (3.4) also:

$$pn = n_i^2 = N_c N_v \exp(-E_g / kT) \tag{3.6}$$

where $E_g = E_c - E_v$

Note that

$$n_i = \sqrt{N_c N_v} \exp(-E_g / 2kT) \tag{3.7}$$

$$\propto T^{3/2} \exp(-E_g / 2kT)$$

Thus n_i has a rapid temperature dependence, doubling for every 12° C rise for silicon around room temperature.

For doped silicon, e.g. n-type, the neutrality condition is between the ionized donors and the conduction band electrons created by the ionization process. For a dopant energy level E_d, the number of ionized donors is related to the Fermi level by the relation

$$N_d^+ = \frac{N_d}{1 + 2\exp\left(\frac{E_f - E_d}{kT}\right)} \tag{3.8}$$

See reference [7]. From (3.3) and (3.8) we have the neutrality condition

$$N_c \exp\left(-\frac{E_c - E_f}{kT}\right) = \frac{N_d}{1 + 2\exp\left(\frac{E_f - E_d}{kT}\right)} \tag{3.9}$$

Fig. 18 shows graphically the solution of (3.9) for two temperature values. At room temperature, the donor atoms are completely ionized and the carrier concentration is essentially equal to N_d, with $E_f = E_{f1}$ a little below E_d. At the reduced temperature, $E_f = E_{f2}$ falls in the small energy range between E_d and E_c and the carrier concentration plummets. Conversely, at very high temperatures, thermal excitation of valence band electrons would become dominant, causing the carrier concentration to rise rapidly, and the Fermi level to stabilise near the middle of the band gap, off-scale to the left in the figure.

For p-type material the number of ionized acceptors is given by

$$N_a^- = \frac{N_a}{1 + 4\exp\left(\frac{E_a - E_f}{kT}\right)} \tag{3.10}$$

The difference in the factors in the denominator arises from the difference between the ground-state degeneracy for donor and acceptor levels.

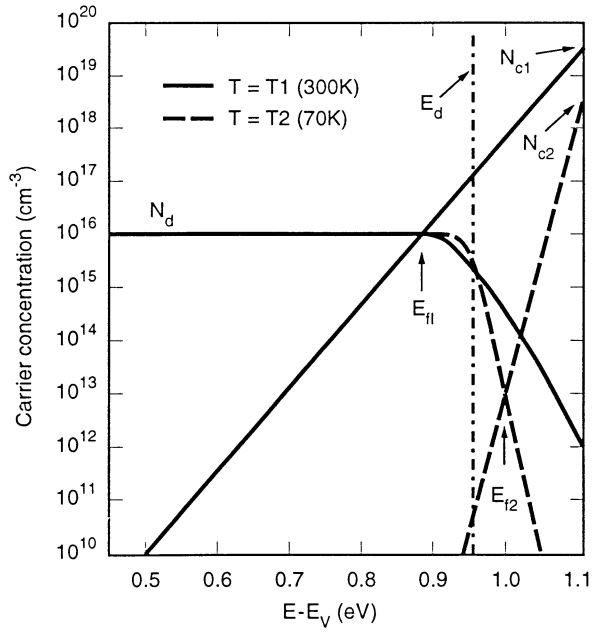

Fig. 18 Number of ionized donors and number of conduction band electrons versus the Fermi energy level E_f.

In general, for doped material we have

$$\left.\begin{aligned} n &= n_i \exp\left(\frac{E_f - E_i}{kT}\right) \\ p &= n_i \exp\left(\frac{E_i - E_f}{kT}\right) \end{aligned}\right\} \qquad (3.11)$$

and $pn = n_i^2 = N_c N_v \exp(-E_g / kT)$ *just as for intrinsic material.* Thus the deviation of a doped semiconductor from the intrinsic condition can be simply represented by a shift in the Fermi energy level with respect to the intrinsic level. The constancy of the *pn* product for different doping conditions is a particular example of the very important *law of mass action* which applies as much in semiconductor theory as it does in chemistry. In thermal equilibrium, the increase in electron concentration by donor doping causes a decrease in the concentration of mobile holes (by recombination) such that the *pn* product is constant. The ionized donors in this sense are passive bystanders, serving to preserve charge neutrality. It is generally valid to think of *n*-type material in equilibrium as containing only mobile electrons and *p*-type material as containing only mobile holes, the majority carriers in each case.

3.2 The *pn* Junction

We now need to introduce a most important fact related to conducting materials which are electrically in contact with one another and in thermal equilibrium; *they all must establish the same Fermi energy*. This applies to

metal/semiconductor systems

n-type/*p*-type systems, etc.

Charge flows from the high to low energy region for that carrier type until this condition is established. For example, at a *pn* junction there develops a fixed space charge of ionized donors and acceptors, creating a field which opposes further drift of electrons and holes across the junction. The *depletion approximation* says that the semiconductor in this condition changes abruptly from being neutral to being fully depleted. This is far from obvious and in fact there is a finite length (the *Debye length*, typically 0.1 μm) over which the transition takes place. But the depletion approximation will be adequate for all the examples we need to consider. Let us look in some detail at the important case of the *pn* junction. Before contact (fig. 19 (a)) the surface energy E_0 is equal in both samples; the *p*-type Fermi level is close to E_v and the sample is densely populated by holes; the *n*-type Fermi level is close to E_c and the sample is densely populated by electrons.

On contact, the electrons diffuse into the electron-free material to the left, and the holes diffuse to the right. In so doing the electrons leave exposed donor ions (positively charged) over a thickness x_n in the *n*-type material, and the holes leave exposed acceptor ions (negatively charged) over a thickness x_p in the *p*-type material. This builds up an electric field which eventually just balances the tendency for current to flow by diffusion. Once this condition is reached (fig. 19 (b)) the Fermi levels in the materials have become equal. The electrical potentials in the two samples (for example the potential energy at the surface E_0 or at the conduction band edge E_c) are

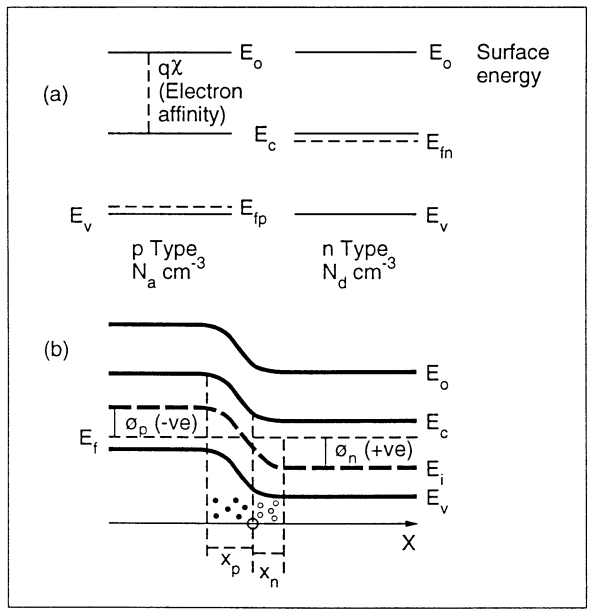

Fig. 19 (a) Energy levels in two silicon samples (of *p*- and *n*-type) when electrically isolated from one another. (b) When brought into contact, the Fermi level is constant throughout the material. The band edges bend in accordance with the space charge generated.

now unequal.

Intuitively, this can be understood as follows. Initially, the electrons at a particular level in the conduction band of the *n*-type material see energy levels in the *p*-type material at equal or lower energy which are unpopulated, so they diffuse into them. The developing space charge bends the energy bands up, so that these levels become inaccessible. Eventually, only very high energy electrons in the *n*-type material see anything other than the absence of states of the band gap in the *p*-type material, and conversely for the holes in the *p*-type material.

Let us develop this quantitatively, adopting a co-ordinate system in which the *pn* junction of fig. 19 (b) is at position $x = 0$. E_0, E_c, E_i and E_v all follow the same x dependence. The zero of the electric potential ϕ is arbitrary, so we define

$$\phi = -\frac{(E_i - E_f)}{q_e} \tag{3.12}$$

Thus ϕ	is	0	for intrinsic material
		positive	for *n*-type
		negative	for *p*-type

From (3.11), in the case of fully ionized donors and acceptors,

$$\phi_n = \frac{kT}{q_e} \ln\left(\frac{N_d}{n_i}\right)$$

$$\phi_p = -\frac{kT}{q_e} \ln\left(\frac{N_a}{n_i}\right)$$

The potential barrier

$$\phi_i = \phi_n - \phi_p = \frac{kT}{q_e} \ln\left(\frac{N_d N_a}{n_i^2}\right) \tag{3.13}$$

Notice that the potential barrier falls linearly with temperature since it is sustained by the thermal energy in the system. We may deduce the electric field strengths $\mathcal{E}(x)$ near the junction by using Poisson's equation

$$\frac{d^2\phi}{dx^2} = -\frac{d\mathcal{E}}{dx} = -\frac{q_e}{\varepsilon_s}\rho(x)$$

ε_s is the permittivity of silicon $= \varepsilon_r \varepsilon_0$

ε_0 is the permittivity of space $= 8.85 \times 10^{-14}$ F cm^{-1}

$= 55.4$ e$^-$/V μm

ε_r is the dielectric constant or relative permittivity of silicon $= 11.7$

For $x_n \geq x \geq 0$

$$\left.\begin{array}{ll} \dfrac{d\mathcal{E}}{dx} = +\dfrac{q_e N_d}{\varepsilon_s} & \therefore \mathcal{E}(x) = -\dfrac{q_e N_d}{\varepsilon_s}(x_n - x) \\ \text{For } -x_p \leq x \leq 0 & \\ \dfrac{d\mathcal{E}}{dx} = -\dfrac{q_e N_a}{\varepsilon_s} & \therefore \mathcal{E}(x) = -\dfrac{q_e N_a}{\varepsilon_s}(x + x_p) \end{array}\right\} \quad (3.14)$$

The *undepleted* silicon on either side of the junction is *field-free*. The depleted silicon close to the junction experiences an electric field whose strength is maximal at the junction and is directed always to the left, i.e. opposing the flow of holes to the right and opposing the flow of electrons to the left.

Requiring continuity of the field strength at $x = 0$ implies

$$N_a x_p = N_d x_n \quad (3.15)$$

Thus, if one wants to make a deep depletion region on one side of the junction (important, as we shall see, for many detectors) we need to have a very low dopant concentration, i.e. very high resistivity material.

The electric field strength varies linearly with x; the electric potential, by integration of (3.14), varies quadratically.

For $x_n \geq x \geq 0$

$$\left.\begin{array}{ll} & \phi(x) = \phi_n - \dfrac{q_e N_d}{2\varepsilon_s}(x_n - x)^2 \\ \text{For } -x_p \leq x \leq 0 & \\ & \phi(x) = \phi_p + \dfrac{q_e N_a}{2\varepsilon_s}(x + x_p)^2 \end{array}\right\} \quad (3.16)$$

Requiring continuity of the potential at $x = 0$ implies

$$x_n + x_p = \left[\frac{2\varepsilon_s}{q_e}\phi_i\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2} \quad (3.17)$$

From (3.13) ϕ_i depends only weakly on N_a and N_d.

If, for example, $N_a >> N_d$ we have $x_p \approx 0$ and (3.17) gives $x_n \propto N_d^{-1/2}$.

So a factor two increase in resistivity leads to a factor of only $\sqrt{2}$ increase in depletion depth.

Fig. 20 summaries these results on the characteristics of an unbiased *pn* junction, with the inclusion of some typical numerical values based on $N_a = 10^{14}\,\text{cm}^{-3}$ and $N_d = 2\times10^{14}\,\text{cm}^{-3}$. The peak field in this case is about 3 kV/cm.

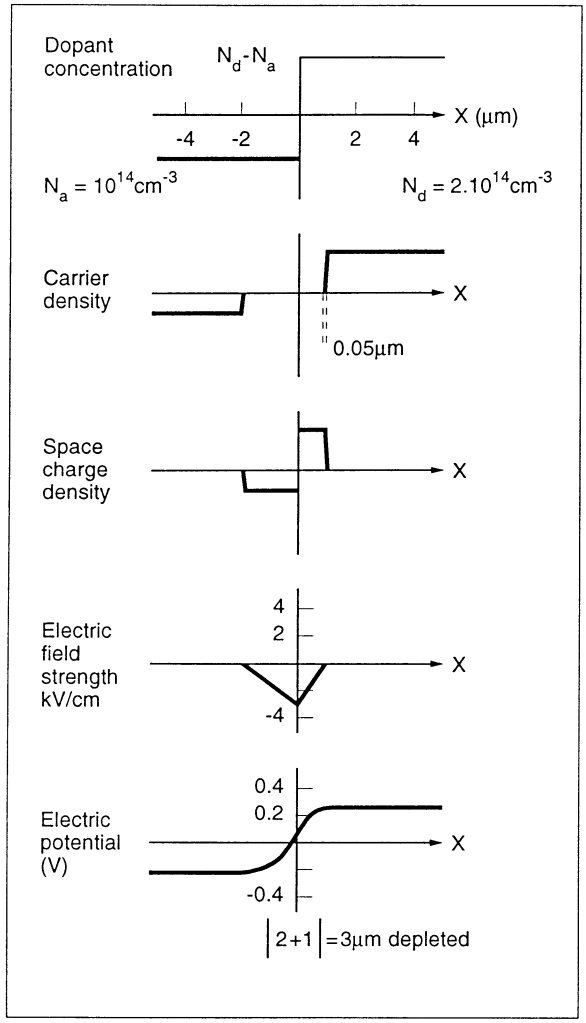

Fig. 20 Summary of various quantities across an unbiased *pn* junction.

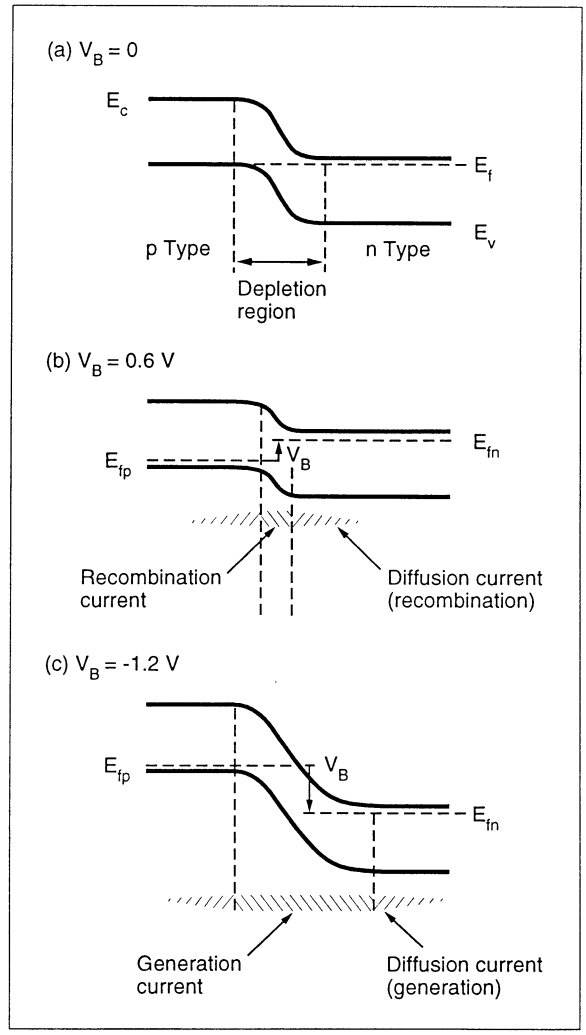

Fig. 21 Effect of an applied voltage across the semiconductor junction.

We now consider the effect of applying a voltage across the junction. Under equilibrium conditions, electron-hole pairs are continually generated by thermal excitation throughout the semiconductor. In the case of zero bias (fig. 21 (a)) the electrons and holes generated within the bulk of the semiconductor recombine. Those generated in the depletion region are swept into the undepleted silicon, holes to the left, electrons to the right. This effect would act to reduce the potential barrier and so is compensated by a small flow of *majority carriers* which find themselves with just sufficient energy to diffuse across the barrier in the opposite directions at just the rate needed to cancel the charge generation in the depleted material. The overall effect is of zero current flow, i.e. equilibrium.

By applying a forward bias (fig. 21 (b)) we separate the previously equal Fermi levels by an amount equal to the bias voltage; the system is no longer in thermal equilibrium or this condition could not be maintained. Although there is still an electric field in the depletion region which is directed against the current flow, the depletion region is narrowed and the potential barrier is now inadequate to prevent majority carriers from flooding across it, holes from the left and electrons from the right. Many of these will recombine within the depletion region giving rise to the *recombination current.* Those which survive are absorbed within one or two diffusion

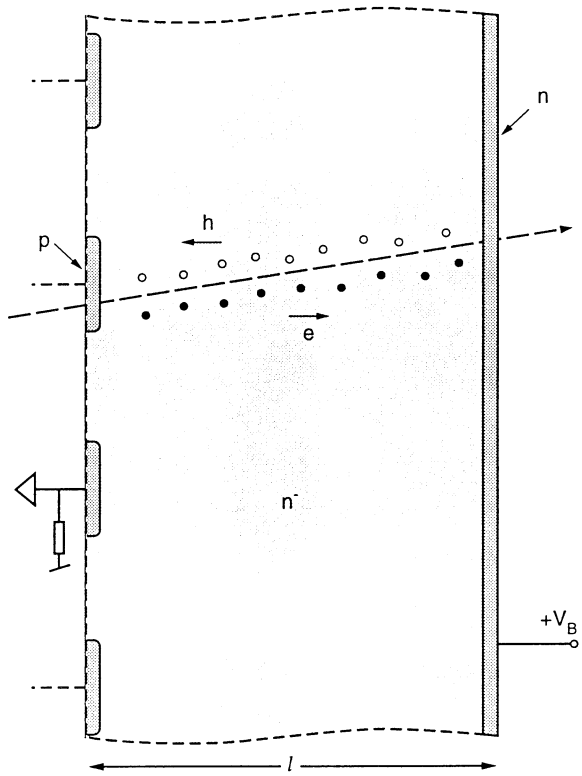

Fig. 22 Operating principles of a simple microstrip detector. It consists essentially of a reverse biased diode, operated at a voltage sufficient to fully deplete the thick, high resistivity, *n*-layer.

lengths by recombination with the majority carriers on that side of the junction, giving rise to the *diffusion current.* Beyond these regions there is just a steady flow of majority carriers supplied from the voltage source to keep the current flowing. Notice that in a forward biased junction the current flow results entirely in electron-hole *recombination.*

With a reverse bias, we have the situation shown in fig. 21 (c). The depletion region is now much wider and electron-hole pairs generated within it are efficiently swept into the undepleted silicon, electrons to the right and holes to the left, giving rise to the *generation current.*

Unlike the case of the unbiased junction, there is now no supply of majority carriers able to overcome the increased potential barrier across the junction. On the contrary, the thermal generation of *minority carriers* within one or two diffusion lengths of the depletion region leads to some holes generated in the *n*-region reaching this depletion region and then being briskly transported across it, and conversely for electrons generated in the *p*-region. This leads to the so-called *diffusion current.* In the case of the reverse-biased junction, the current flow is thus caused entirely by electron-hole *generation.* The current flow across reverse-biased junctions is of great importance in determining the noise limits in silicon detectors. An immediate observation is that, since this current arises from *thermal* generation of electron-hole pairs, the operating temperature will be an important parameter.

Before continuing to discuss this point, it is worth noting that we have finally collected up enough information to calculate the characteristics of a typical particle detector, and it is instructive to do so. Referring to fig. 22, we have a silicon detector of thickness ℓ made of good quality, high resistivity n-type silicon (ρ = 10 KΩ cm). On the front surface we make shallow implants of acceptor atoms (the p-strips) and on the back surface we make a highly doped n-type implant to provide a good low-resistance ohmic contact. The terms n^+ and p^+ are conventionally used to represent high doping levels, n and p represent moderate levels, n^-, p^- or π and ν represent low levels, and i is used for intrinsic or compensated material of the highest possible resistivity.

Now we apply a positive voltage V_B to the n-type surface with the aim of completely depleting the detector. In this way we shall ensure complete collection of the electrons and holes generated by the passage of a charged particle; with incomplete depletion we would lose signal by recombination. Equation (3.17) applies, with the difference that we replace ϕ_i by $V_B + \phi_i$ since the junction is biased in the direction which assists the previously existing depletion voltage.

We have

$$x_n + x_p \approx x_n = \left[\frac{2\varepsilon_s}{q_e}(V_B + \phi_i)\left(\frac{1}{N_a} + \frac{1}{N_d}\right)\right]^{1/2}$$

$$\approx \left[\frac{2\varepsilon_s}{q_e} \times \frac{V_B}{N_d}\right]^{1/2}$$

From fig. 17, we see that $N_d \approx \dfrac{6\text{x}10^{15}}{\rho}$ and we require $x_n = \ell$

$$\therefore V_B = \frac{q_e}{2\varepsilon_s} \times \frac{6\times10^{15}}{\rho} \times \ell^2$$

$$= \frac{10^{-4}}{2\times55.4\times11.7} \times \frac{6\times10^{15}}{\rho} \ell^2 \times 10^{-8}$$

where ℓ is in μm and ρ in Ω cm

$$\therefore V_B = \frac{4.63\ \ell^2}{\rho}$$

For the above example, $V_B = 42$ V is the potential needed to fully deplete the detector. We also find $\mathscr{E}_{max}$ = 2.8 kV/mm. This looks comfortable in terms of the breakdown field in silicon. However, in a real detector, it is important to pay attention to the regions near the edges of the p-strips, where the fields can be very much higher.

Returning to the general properties of the reverse biased junction, the most important parameter influencing the leakage current is the operating temperature. At high temperatures, above 100 °C typically, the leakage current

is dominated by thermal electron-hole generation within approximately one diffusion length of the depletion edge. The diffusion length for minority carriers is

$$L_D = \sqrt{D\tau_m} \tag{3.18}$$

where D is the diffusion constant and is related to the mobility μ by the Einstein relation

$$D = \frac{kT}{q_e}\mu \tag{3.19}$$

For electrons $D_n = 34.6 \text{ cm}^2 \text{ s}^{-1}$
For holes $D_p = 12.3 \text{ cm}^2 \text{ s}^{-1}$
} at room temperature

τ_m is the minority carrier lifetime, and it can vary from about 100 ns to more than 1 ms depending on the quality of the silicon. This point will be discussed further. This leakage current (termed the diffusion current, as previously noted) depends only weakly on the reverse bias voltage, but is highly temperature dependent due to its origin in the thermal generation of minority carriers.

At lower temperatures (less than about 100°C) the diffusion current becomes negligible and the generation current dominates. This continues to show a similarly fast temperature dependence, but is now also quite voltage dependent, since the depletion width is proportional to $V_B^{1/2}$.

The diffusion and generation currents depend on the rate of generation of electron-hole pairs, and the diffusion current depends also on the minority carrier lifetime. These quantities are in fact closely related. Direct thermal generation of an electron-hole pair is quite rare in silicon for reasons which depend on the details of the crystal structure. Most generation occurs by means of intermediate generation-recombination centres (impurities and lattice defects) near the band gap centre. Thus an electron-hole pair may be thermally created in a process where the hole is released into the valence band and the electron is captured by the trapping centre in one step, to be subsequently emitted into the conduction band. These *bulk trapping states* vary enormously in their density and can be held down to a low level by suitable processing. It is precisely these states which determine the minority carrier lifetime already mentioned. Reducing the density of bulk trapping states does two things. It cuts down the thermal generation of charge carrier pairs in the material, so reducing the concentration of minority carriers available for the generation of current across a reverse-biased junction. It also increases the minority carrier lifetime and so the diffusion length (but only at $\tau^{1/2}$). The first effect vastly outweighs the second, so that a low density of bulk trapping states is highly advantageous in ensuring low leakage current. As we shall see later, even originally high grade silicon can deteriorate due to the production of bulk trapping states by radiation damage. Mid-band gap impurities such as gold are a particularly serious source of bulk trapping centres. Even in low concentrations, gold atoms strongly reduce the carrier lifetimes, and lead to greatly increased leakage current.

These effects are less serious in cases where one is collecting large signals promptly. But in cases of small signals and/or long storage times (such as in a silicon drift chamber, or CCD), particular care is needed. One important design criterion is to keep the stored charges well away from the surface of the silicon, since the silicon/silicon dioxide interface always has a high level of lattice defects. This criterion has led to the development of various forms of *buried channel* radiation detectors, to be discussed later.

3.3 Charge Carrier Transport in Silicon Detectors

While the charge generated by an ionizing particle is being transported by the internal field in the detector, the process of diffusion spreads out the original very fine column of charge. In the case of very highly ionizing particles (such as alphas) the original density of electrons and holes can be so high that space-charge effects are important. In the case of MIPs, however, such effects are negligible and the time development of the electron and hole charge distributions may be treated by simple diffusion theory.

Consider a local region of electron charge, for example a short section of the particle track length within the silicon. Under the influence of the internal field, this will be drifted through the material and at the same time will diffuse radially as indicated in fig. 23.

The RMS radius of the charge distribution increases as the square root of drift time t_d, as in (3.18), with standard deviation $\sigma = \sqrt{2Dt_d}$. Thus 50% of the charge is contained within a radius of 0.95 $\sqrt{Dt_d}$. Assuming a 'typical' drift field in depleted silicon of 1 kV/cm, and using the fact that the drift velocity $v_n = \mu_n \mathcal{E}$, we obtain the following indication of the growth of a charge packet with time:

Drift Time	Charge Radius	Drift Distance
10 ns	6 μm	135 μm
1 μs	60 μm	14 mm

Diffusive charge spreading is an attractive option for improving spatial precision beyond the limits of the detector granularity. For example, one might hope to achieve precision of one or two microns from a strip detector with 25 μm pitch, by centroid finding on the basis of measured charge collection in adjacent strips. This depends on achieving a charge radius of ≥ 30 μm which (from the above table) implies large drift distances and/or gentle drift fields. Ideas for improved precision by centroid finding may be limited by the available resistivity of silicon.

There is however, an alternative approach that has so far been applied only to CCD detectors, but which could be of more general interest. A wafer cut from a silicon crystal will normally have a rather uniform dopant concentration. It is possible subsequently to grow relatively thick (up to around 100 μm) *epitaxial layers* on the substrate wafer, of excellent crystalline quality and quite different (but also uniform) dopant concentration. For detector applications, a low resistivity substrate with a high resistivity epi layer is of particular interest. In the CCD case, as we shall see, the epi layer would be implanted with an *n*-layer, and biased so as to deplete only approximately 3 μm depth. The charge carrier transport associated with (for example) a charged particle track

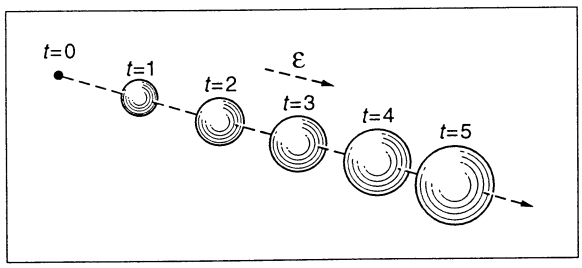

Fig. 23 Combined drift and diffusion of an initially compact charge cluster (electrons or holes) as a function of time over equal time intervals.

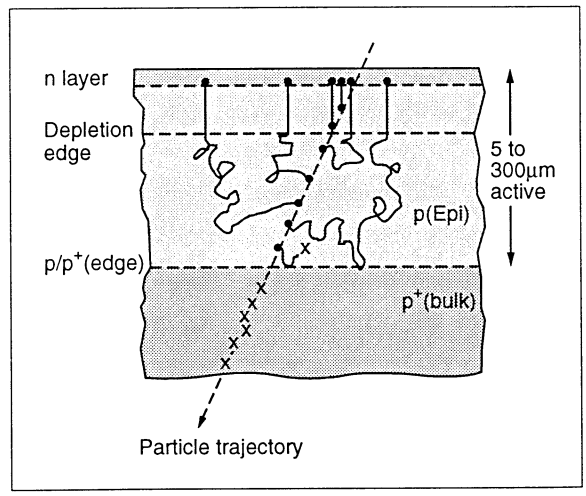

Fig. 24 Charge collection from a silicon structure as used in some pixel devices.

traversing such a structure is depicted in fig. 24. Electrons within the thin depletion region are promptly collected into the buried channel, with no time for lateral diffusion. Electrons from the highly doped p^+ bulk are completely disposed of by recombination (very short minority carrier diffusion length in this material). However, electrons generated in the undepleted epitaxial layer find themselves able to diffuse homogeneously in all directions. Those which approach the p / p^+ junction experience a potential barrier as we have already discussed in the case of the unbiased *pn* junction, of magnitude

$$\phi_B = \frac{kT}{q_e} \ln\left(\frac{N_{p^+}}{N_p}\right)$$

For a 20 Ω cm epi layer on a highly doped 0.1 Ω cm substrate, we find

$$\phi_B = 180 \text{ mV compared with } \frac{kT}{q_e} = 26 \text{ mV}$$

at 300 K. The p / p^+ interface therefore acts as a *perfect mirror*, and the electrons continue diffusing until they happen to approach the *pn* depletion edge, at which point they are stored. Thus a MIP leaves an electron charge cluster which is transversely spread by an amount related to the epi layer thickness. Such a detector made with partially undepleted thick-epi material is in principle better for precision tracking by centroid finding than a fully depleted detector. To fully exploit this concept, one has to pay attention to the detector granularity, epi layer thickness, readout noise, etc. The most spectacular results in precision centroid finding in CCDs have been obtained not as yet with MIPs but with defocused star images in satellite guidance system, where precision below 0.1 μm has been achieved using 20 μm pixels. This constitutes a very important demonstration of the inherent pixel-to-pixel homogeneity possible with high quality silicon processing.

4 Microstrip Detectors

4.1 Introduction

Charged particles deposit a significant fraction of their energy by ionization in all types of materials, but only some are suitable as detector media. The conceptually most elementary detector types are insulators in which the signal is collected simply by applying a voltage to a pair of metal plates attached to the opposite faces of the detector layer, so creating an electric field within the material. The detection medium may be a gas (ionization chamber), a liquid (e.g. liquid argon calorimeter), or a solid (e.g. diamond detector). However, this principle cannot be applied to semiconductor detectors since even the highest purity material would generally have unacceptably low resistivity (i.e. excessive leakage current) except at extremely low temperature. As we have seen, it is possible to generate a region of internal electric field devoid of free charge carriers, and hence having greatly reduced leakage current, by creating a reverse biased junction. Electron-hole pairs generated within the depletion region, for example by thermal or optical excitation, or by the passage of a charged particle, are promptly swept to the surface for collection. This principle has been used for the detection of ionizing particles in silicon for over 40 years [8]. We have already noted some variations on this theme in connection with pixel devices (collection of minority carriers from undepleted material adjacent to depleted silicon) but the microstrip detector follows exactly this simple tradition.

The pioneering microstrip detectors of the early '80s [9] were based on the processes used for many years to manufacture non-segmented semiconductor detectors for nuclear physics applications. The diodes were simply formed by the surface barrier between metal (aluminium) strips and the high resistivity substrate. The strips were wire bonded to huge fanout boards which housed local pre-amplifiers connected to every Nth strip ($N \approx 5$). The principle of capacitive charge division was used to interpolate the track co-ordinates for signals collected on floating strips. The ratio of board area to detector area was almost 1000 to 1; this was tolerable in fixed target experiments having unlimited space for local equipment outside the aperture of the forward spectrometer.

Closely following on these early developments, two revolutions took place which totally changed the technology of these detectors, opening up for them a much more powerful role in particle physics.

The first of these revolutions was to switch from surface barrier detectors to ion implantation, thus adopting the highly developed techniques used for processing integrated circuits. The microstrip detector becomes essentially a p-i-n diode structure, as we discussed in Section 3.2. The p-strips (fig. 22) were overlaid with metal (aluminium) to provide a low resistance path, and connected to external electronics. This development [10] had been considered impossible by many semiconductor detector experts at the time. The high resistivity material used almost uniquely by detector people was supposedly incompatible with the high temperature processing required for the activation stage of ion implanted material. Kemmer showed that these experts were incorrect; it was problems of cleanliness in processing, rather than the high temperatures themselves, which led to the dreaded resistivity drops. The first result of this revolution was *more* robust detectors and hence the possibility of much larger areas. As important, the door was opened for the inclusion of a host of features already developed

for ICs, such as techniques for isolating edge-related leakage currents (guard rings), for biasing with high dynamic resistance, and so on. Some of these will be discussed in the Section 4.3.

The second revolution was the development of readout chips with high density front-end amplifiers [11, 12]. Using integrated circuit technology, the front-end could be shrunk to a pitch of 50 μm, permitting the microstrip channels to be wire bonded directly to these compact ICs located along the edge of the detector. Furthermore, the readout chips embodied resettable storage of the analogue signals, and multiplexed readout. Thus, the number of cables needed for the detector readout was reduced by about a factor of 100. We shall in Section 4.3.3 record great ongoing progress in developing special readout ICs to suit a wide range of experimental conditions.

The combination of robust, sophisticated microstrip detectors and extremely compact electronics has led to their application in a host of experiments. With the SLC Mark II and LEP detectors, they crossed the barrier from fixed target experiments into the collider environment, with excellent results in heavy flavour physics.

4.2 The Generic Microstrip Detector

Microstrip detectors come in a large variety of designs, each with its own strengths and weaknesses, each with a certain range of applications.

Due to the fact that high resistivity n-type material is more readily available, most detectors have used n-type wafers as starting material, though this may be changing in some application areas. The 111 crystal-orientation is conventionally used, but reasons why this too may be changing are discussed in the next section. As already mentioned, the pioneering detectors all used p^+ strips, collecting holes from the track of the ionizing particle. More recently, the back surface (n^+ implant) has also been subdivided into strips (which can as well be angled, perhaps at 90° to the p-strips) giving us double-sided microstrip detectors.

Such a detector, and the associated internal electric field, is sketched in fig. 25. The reverse bias is achieved by applying a positive voltage to the n-strips, the p-strips being grounded. In each case, series resistors (usually on-chip polysilicon) are used to create a high impedance path. The electric field (directed in the negative Z direction) would be uniform across the depleted n^- substrate, were it not for the finite resistivity and hence the presence of a low density of fixed positive charges. Due to this space charge the magnitude of the field falls steadily from its peak value at the pn junction, towards the n-side. The sketch shows an over-depleted detector. For the just-depleted case, the field would sink to zero at the surface of the n-strips. Once we enter the heavily-doped p- or n-strip region, the field develops a large gradient, falling abruptly to zero.

The sketch indicates an AC coupled detector. The metal readout strips are isolated from the implanted strips by a thin layer of dielectric (silicon dioxide). Thus the amplifier inputs sense the fast signal without also being obliged to sink the DC leakage current. Both AC and DC coupled microstrip detectors are common. In applications where radiation levels are low, and hence degradation in leakage current is not a problem, the extra simplicity of DC coupled detectors may be advantageous. Early microstrip detectors were all DC coupled.

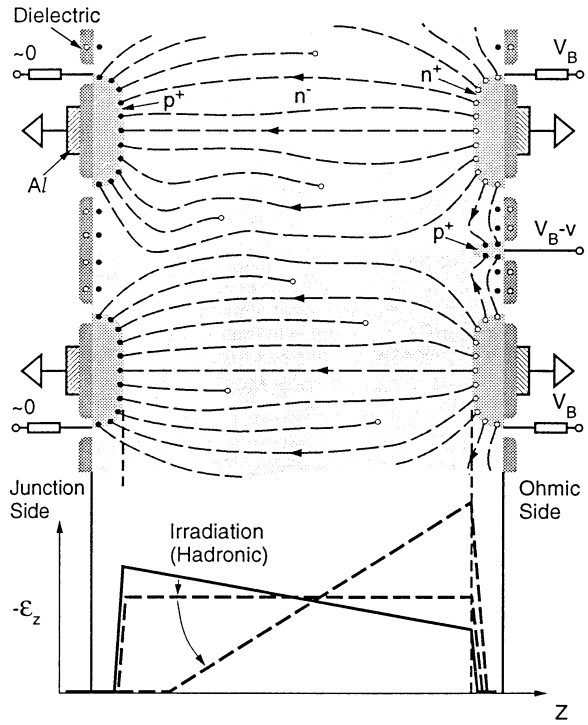

Fig. 25 Sketch of cross-section of a generic double-sided microstrip detector. Exposed fixed charges are shown by open circles (positive) and filled circles (negative). Also shown is the electric field distribution in such a detector before and after radiation-induced displacement damage in the silicon.

Between neighbouring charge collection strips on both sides is a passivation layer of silicon dioxide. Such oxide layers inevitably collect some positive charge (holes trapped as interface states) which is compensated by a very thin accumulation layer of mobile electrons in the bulk material. On the *p*-side, these are repelled by the exposed negatively charged dopant atoms in the *p*-strips. However, on the *n*-side, they create a low resistance inter-strip leakage path. Signal electrons collected on one *n*-strip will readily flow to neighbouring strips; the strips are effectively shorted together. This problem can be overcome in a number of ways; fig. 25 shows one of the cleanest solutions which is drawn straight from the textbooks of IC design. p^+ 'channel stops' are implanted between the *n*-strips. They are biased somewhat negatively relative to the strips, and hence acquire a negatively charged depletion layer which repels the mobile electrons in the surface accumulation layer, so blocking the leakage path that would otherwise be present.

Before leaving this figure, there is one further point worthy of note, relating to the collection of signal charge. After the passage of an ionizing particle, holes begin to drift to the left, electrons to the right. Once the charges separate, the space-charge self-repulsion in principle leads to expansion of the charge cloud during the drift time. A localised charge distribution of N carriers (holes or electrons) will expand with time to a sphere of radius r_s, where

$$r_s = \left[4.5 \times 10^{-7} \frac{\mu N}{\varepsilon_s} t_d \right]^{1/3} \text{cm}$$

ε_S is the permittivity of silicon and t_d is the drift time in seconds. For collection of holes or electrons in a microstrip detector, r_s amounts to less than 1 μm and can be neglected (while the signal from an α particle can expand to $r_s \approx 100\,\mu$m ; see reference [9]). As we saw in Section 3.3, diffusive charge spreading can on the other hand be considerable. This is sensitively dependent on the type of charge carrier collected, on the detector resistivity, and on the biasing conditions.

For the conditions shown in fig. 25, a strongly over-depleted detector, the electric field is reasonably uniform. For a just-depleted detector, the *holes* would all pass through the high-field region close to the *pn* junction, and those generated in that half of the detector would be entirely drifted through a fairly high field. For the *electrons*, on the contrary, all would pass through the low field region before reaching the *n*-strips. Hence (even without the effect of the relative mobilities) the electron cloud will experience greater diffusive charge spreading than the hole cloud. In principle, this would give us higher precision (by centroid fitting) on the *n*-side than on the *p*-side. This question is discussed in more detail in the next section.

There are, however, several reasons why such fine tuning of detector parameters may not yield the desired improvement in precision.

Firstly, in a radiation environment, the effective dopant concentration varies with time. As depicted in fig. 25, and discussed in detail in Section 6, hadronic irradiation causes the depleted material to become steadily more *p*-type. Having passed through the compensated condition (when it could be depleted with a few volts) the resistivity falls steadily. After a certain dose (for fixed operating voltage) the detector would fail to deplete fully and the hole signal would be lost (no longer collected on an individual *p*-strip). The electron signal would still be collected, but from a steadily decreasing thickness of detector. Thus any precision advantage gained by fine tuning the depletion conditions could not be preserved through the life of the detector.

Secondly, due to their thickness, microstrip detectors have a significant probability of loss of precision due to δ-electrons, as discussed in Section 2.3. Results published from test beams often limit the signal charge to less than approximately 1.7 times the MIP mean value, in order to restrict the tails on the co-ordinate residuals. In tracking detectors with a limited number of points per track, one would not normally have the luxury of such a filter. For binary readout detectors, one would not even know which were the large signal clusters.

Thirdly, detector precision is seriously degraded for angled tracks, as we shall see in detail in the next section.

Finally, most tracking detectors in experiments operate in a magnetic field which (because of the Lorentz angle) degrades the measurement precision. In a conventional collider geometry with a solenoid magnet, the Z measurements are unaffected but the precision of the $R\,\phi$ measurement is degraded. For details, see the next section.

4.3 Microstrip Detectors: Detailed Issues

4.3.1 Design Optimisation

All silicon microstrip detectors are of approximately 300 μm thickness. For much thinner detectors, the loss of signal charge, exacerbated by the reduction in signal voltage due to the increased capacitance from strip to substrate, results in a poor signal-to-noise performance. Even thicker detectors might be required for example in cases of modules having several long strips linked together and to a single readout chip. The capacitance to substrate is a particularly important issue in cases where capacitive charge division is used for the readout of floating strips. To avoid serious signal loss, it is essential that the geometry be chosen so that the inter-strip capacitance greatly exceeds the strip-to-substrate capacitance, or one would suffer from serious loss of signal from floating strips. In some large systems currently under design (e.g. the ATLAS Silicon Central Tracker or SCT) the individual modules are 12 cm in length, with strip capacitances of around 18 pF (1-2 pF/cm is typical). Such large capacitances represent a considerable challenge for readout electronics, as we shall see in Section 4.3.3.

As already mentioned, a high resistivity *n*-type substrate is conventionally used. High resistivity *p*-type material is now available (both bulk and epitaxial), providing an interesting option for detector fabrication. Such detectors would have the advantage that under irradiation they simply become steadily more *p*-type. Thus one would avoid the complications (e.g. in guard ring structures) associated with the junction shifting over from the *p*-side to the *n*-side during the life of the detector.

The 111 crystal orientation is conventionally used in microstrip detectors, since it provides the densest surface and hence the lowest probability of 'spiking' (growth of aluminium deeply into the crystal in local regions, possibly shorting out the diode structure). For IC manufacture (and also for MOS detector types such as CCDs), the 100 crystal orientation is preferred due to the lower density of dangling bonds at the silicon/silicon dioxide surface, and hence lower trapped charge at the interface. This may be particularly important in some microstrip detector applications, and for this reason some groups are doing exploratory work with 100 material. For AC coupled detectors, the area of metal in contact with silicon is reduced by many orders of magnitude compared to the early DC coupled devices. Also, metallization equipment is now extremely refined compared to 10 years ago, so the problem of spiking should be largely in the past.

For biasing microstrip detectors, the most commonly used method (also the simplest) is via on-chip polysilicon resistors. A problem with this approach is that as one has to allow for higher leakage current (due to radiation damage and/or longer strips) the resistance value needs to be reduced in order not to disturb the bias voltage excessively. This in turn can lead to loss of signal and worsening signal-to-noise ratio. The ideal solution would be a low DC resistance and a high dynamic resistance. Two approaches have been adopted, the reach-through structure [13] and the FOXFET biasing scheme [14]. This Field OXide FET structure, which employs a thick gate oxide, is vulnerable to radiation damage effects [15, 16]. The present situation appears to be that polysilicon biasing is the only safe solution for detectors to be used in a high radiation environment.

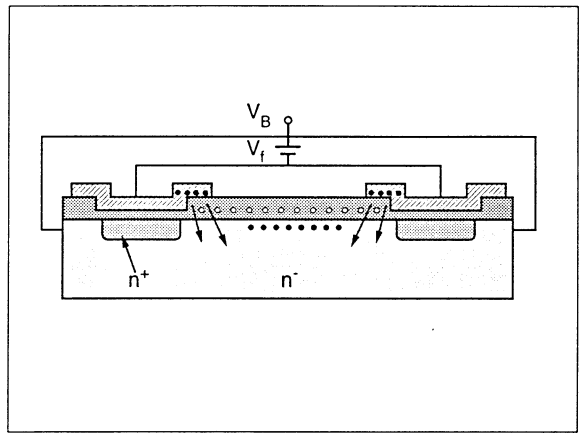

Fig. 26 The technique of *n*-stop isolation by field plate separation with extended AC coupled electrodes (one of several field plate approaches).

For the *n*-strip isolation in detectors (single- or double-sided) where the electron signal is collected, two methods have been adopted. The channel stop approach [13] has been illustrated in fig. 25. An alternative 'field plate' method uses an MOS gate structure, in the form of 'wings' attached to the aluminium readout strips in AC coupled detectors [17]. This is illustrated in fig. 26.

For all these various microstrip detector structures, careful attention should be paid (by 2-D simulation) to the peak electric fields induced near the strip edges. Poorly understood leakage current has characterised many of the designs which at first glance looked quite reasonable. In a pioneering paper [18], Ohsugi and co-authors demonstrated the sensitivity to geometrical details in the specific case of AC coupled *p*-strip sensors. Breakdown was demonstrated in structures where the relative edges of the p^+ implant and the aluminium electrode led to peak fields at the edge of the implanted strips exceeding the breakdown limit in silicon of 30 V/μm. While such problems can in principle be avoided by careful design, it is very easy to encounter some local variations, edge effects at the ends of the strips, etc. which can still cause problems. To this end, the diagnostic tool demonstrated in their paper is of enormous value. Using an infra-red microscope equipped with a CCD camera, very small regions of avalanche breakdown can be seen clearly. This marvellous tool [19] is of value wherever anomalous leakage currents are encountered either due to design deficiencies or to process faults. One of the problems that has plagued manufacturers of large area microstrip detectors, particularly in the case of double-metal structures (see below), is that of pinholes in the dielectric, permitting unwanted leakage paths. An infra-red microscope can be used to explore the positions of these defects, and possibly to suggest solutions (e.g. improved step coverage across gate edges). Similar problems have been encountered and solved in this way in the world of CCD detectors. For *n*-strip microstrip detectors, there is evidence (not surprisingly) that field plate devices are more susceptible to micro-discharges than *p*-stop devices. However, much depends on the specific design details.

It is hardly surprising that another issue which still causes problems in microstrip detector design is that of uncontrolled oxide layers (e.g. inter-strip, as depicted in sketch form in fig. 25). In other detector types such as CCDs, care is taken to avoid even fine cracks between gate electrodes (by overlapping neighbouring electrodes)

since gate oxide inevitably contains trapped interface charge, the magnitude of which increases with irradiation. The electrical effects of such trapped charge can be minimised by the presence of a metal or polysilicon cover layer held at a well defined potential. Microstrip detectors do not easily lend themselves to such design rules, but one may escape from trouble due to the accumulation layer of electrons already referred to. However, particularly if one is aiming for high efficiency for detection of (say) soft X-rays which deposit their signal near the surface, there are numerous examples of anomalous dead layers and other effects probably related to the uncontrolled oxide. This is an area for ongoing concern regarding the design of microstrip detectors.

The use of high resistivity silicon is driven by the desire to have a manageable operating voltage for full depletion; 150 V is commonly considered an upper limit. Under intense hadronic irradiation, this may set an uncomfortably short lifetime for the detector. It has been pointed out [20] that careful design of microstrip detectors (particularly as regards implant profiles, strip edges, guard rings, etc.) may enable operating voltages to be set even above 1 kV before micro-discharges or breakdowns occur. Such a design would considerably extend the usefulness of microstrip detectors in high radiation environments. Note that it is usually the breakdown voltage rather than the leakage current which shortens lifetime of a detector in a radiation environment. The leakage current can always be reduced by cooling. There is long experience of this in the area of CCD detectors, and large systems of cooled microstrip detectors are now in the planning stages [21].

We have discussed briefly the availability of double-sided detectors, which are of interest in that they provide apparently two advantages over (for example) a pair of single-sided detectors. Firstly, less material (of particular significance for vertex detectors) and secondly some degree of resolution of the ambiguity problem for multi-hit events. Regarding the latter, the idea is that one can measure the signal charges in the p- and n-side clusters and use the correlation between them to rule out some of the associations (e.g. between a below-average cluster in one view and a multi-MIP cluster in the other view). In fact, this is not a very practicable idea, since the level of ambiguity is not greatly reduced.

Regarding extraneous material in the active volume, much depends on the angle between the strips on the two sides. If this is small (e.g. a few degrees) both sides of the module can be read out from the end without complications. If however, one requires a large angle between the two strip planes (e.g. 90°), there are two options. Consider the case of a Z view as well as the conventional $R\phi$ view in a collider environment. The most obvious option, implemented in the pioneering double-sided ALEPH vertex detector [22], would be to place the Z readout chips along the long edge of the module. This results in a large amount of electronics in the active volume of the barrel detector system, which is not a good idea if precision vertexing is the goal. Later detectors have followed one of two different approaches. Both move the Z readout chips to the ends of the barrel, outside the active volume, in the same general area as the $R\phi$ readout chips. The most ambitious approach is to integrate the linking traces onto the detector modules themselves, using a double-metal technology [23, 24]. A dielectric layer separates the Z-strips from the orthogonal readout strips, and metallized vias provide the connections between the two levels. Due to the larger number of Z-strips than readout strips in a typical module (a long rectangle) the Z-strips may be connected in a repeating pattern, resulting in some degree of ambiguity as

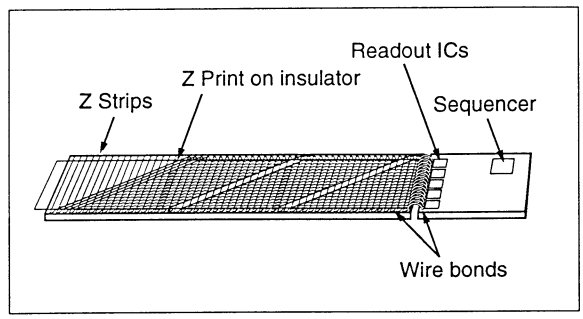

Fig. 27 A scheme for Z-strip readout using a separate metallized substrate (glass or kapton).

to the spatial position (normally not a problem given the overall track-finding software). Alternatively, the Z-strip pitch may be made correspondingly coarser than the pitch of the $R\phi$ readout strips. There is one inevitable disadvantage to the double-metal approach, which is the increased capacitance of every strip; the detector strips and readout strips form a web of closely linked electrodes, separated only by the thickness of the dielectric layer. This, coupled with the fact that tracks at the ends of the polar angle range may deposit their charge over a number of Z-strips, can lead to a serious degradation in the signal-to-noise performance of the detector. The capacitance problem can be greatly reduced, with only a modest degradation in terms of material in the active volume, by routing the readout traces on separate thin substrates (e.g. copper traces on kapton) [25]. The Z-strips are wire bonded to diagonal readout strips at the edge of the detector, the signals being carried to the electronics in a zig-zag geometry, using additional Z-strips to link the diagonal readout strips. This idea is illustrated in fig. 27. In this way a low and acceptable ambiguity level as to which of a few widely separated strips was hit, is the price paid for accessing the data in an economical form with little additional material, and a generally acceptable overhead in capacitance.

There remains the choice between double-sided detectors and two back-to-back single-sided detectors, one for $R\phi$ and one for Z. As has been noted, the correlated cluster signal information is not often very useful, so the key issue is that of the additional material in the back-to-back approach. In vertexing applications this is always important, though seldom decisive. There is necessarily additional material in the form of support structures, etc., so we are certainly not talking about a factor of two, and the multiple scattering is proportional to the square root of the thickness. If the double-sided option came free of additional costs, it would clearly be preferred. However, this is far from the case. Double polished silicon wafers are available and are not in themselves particularly expensive. However, for bulk production, it is desirable to use as far as possible the standard IC manufacturing equipment, which is all explicitly geared to single-sided processing. It has been claimed that the cost of double-sided relative to single-sided detectors is 3:1. This may be true for some small production runs, where it merely reflects the reduced yield of the double-sided devices. However, for large volume production such as we are now seeing planned (e.g. for the LHC SCTs) it should be possible to greatly reduce the cost per unit area of detectors made with standard processing equipment. In this case the cost ratio mentioned above is likely to become much more unfavourable. Time will tell.

4.3.2 Spatial Precision in Microstrip Detectors

Early microstrip detectors with very fine readout pitch (and huge fanout factors) had wonderful spatial precision, but are now only of historical interest. We are at present effectively constrained by the readout pitch of all existing front-end electronics, namely 50 μm. This can be reduced by a factor of two by attaching readout ICs at each end of a module, and this has been done in environments of high track density. Also, one can include floating strips as has already been discussed. Spatial precision of approximately $\frac{25}{\sqrt{12}}$ μm$=7.2$ μm is thus in some ways natural for a silicon microstrip detector when read out with currently standard electronics. In large tracking systems, one has frequently to work very hard to achieve such levels of stability and systematic precision, for many reasons. Having said this, considerably better spatial precision has been achieved, mostly in test beam situations.

Let us consider first the case of normal incidence tracks. As we saw in Section 4.2, the extra diffusive spreading would suggest that (for a given strip pitch) one might be able to achieve a higher precision in the charge collection on the *n*-side (electrons) as opposed to the *p*-side (holes). However, most experimental results to date have been obtained with detectors made with *p*-strips on *n*-bulk silicon.

Using a single-sided detector with *p*-strips on a 20 μm pitch, and analogue readout on every strip, Belau et al [26] were able to measure the *spatial distribution* of the hole charge collected. This varied from an RMS width of 2.5 µm to 1.9 μm as the operating voltage was raised from 120 V (just-depleted) to 200 V (over-depleted). From this they *calculated* the precision achievable with a readout pitch of 20, 60 and 120 μm to be $\sigma=$ 2.8, 3.6 and 5.9 μm, in the optimal case of the just-depleted detector. Measurements with 60 and 120 μm readout pitch [27] yielded precisions of 4.5 and 7.9 μm, a little worse than calculated, but better than $\frac{20}{\sqrt{12}}=5.8$ μm which would be the limit for a digital system with 20 μm readout pitch. Evidently, some degree of useful charge spreading is achieved with detectors having narrow strip pitch. For electron collection, the lower average electric field yields even better *calculated* precision, 0.8 μm to 3.6 μm for the three cases mentioned above. In this case, they did not have data for comparison.

In all this, please remember the caveat about δ-electrons mentioned in Section 2.3. In these test beam studies, clusters with more than 1.7 times the mean MIP signal were discarded, with the consequential efficiency loss that could probably not be tolerated in a detector used for a physics experiment.

Results with a more typical arrangement of readout of every strip on a pitch of 50 μm have been reported for double-sided detectors [28]. For normal incidence the precision achieved was 8.8 μm on the *p*-side. This slightly worse figure is attributed to the higher electronic noise in that system. The signal-to-noise was 16 for the *p*-side and degraded (for not completely clear reasons) to 10 for the *n*-side. The precision for the *n*-side signal was 11.6 μm, confirming the suggestion that the system noise played a large part in the measured spatial precision.

For normal incidence tracks, we may conclude that spatial precision in the region 5 - 10 μm is typical for strip pitch ≤ 50 μm, and with readout pitch ≤ 150 μm. The degradation in precision with increasing readout

pitch is fairly modest. The usual reason for requiring a fine readout pitch (typically, equal to the strip pitch) is the need to preserve an optimal two-track resolution.

Once we permit angled tracks (which really only are of concern for the *RZ* view as opposed to the $R\phi$ view in colliders) the situation deteriorates fairly rapidly. The particle leaves a trail of charge carriers which are collected on a number of *Z*-strips. Taking the overall centroid is a bad approximation to the track position at the centre plane of the detector, due to the energy loss fluctuations along the track. The problem has been studied theoretically [29] and experimentally [30], as a result of which Hanai et al have developed an algorithm ('convoluted Gaussian centroid') which leads to an experimental precision as function of α, the track angle to the detector normal, varying from 8 μm ($\alpha = 0°$) to 40 μm ($\alpha = 75°$). These results were obtained using a single-sided *p*-strip detector with 25 μm strip pitch and 50 μm readout pitch.

A dangerous factor affecting spatial precision in microstrip detectors is the effect of magnetic fields. Empirical measurements have been reported in [26]; these agree well with calculations. For the *p*-strip signal in a just-depleted detector, a magnetic field of 1.7 T shifts the measured co-ordinate by about 10 μm and increases the width of the collected charge distribution from 5 to 12 μm. The relevant parameter determining these effects is the Hall mobility μ_n^H for electrons and μ_p^H for holes; see Shockley [7]. With the usual arrangement in collider barrel detectors (magnetic field $\mathscr{H}$ perpendicular to electric field) the charges drift at the Lorentz angle θ^L with respect to the electric field, where θ^L is almost independent of the magnitude of the electric field and is given by

$$\tan\theta^L_{n,p} = \mu^H_{n,p} \times \mathscr{H}$$

Now

$$\mu_p^H \approx 310 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$$

and

$$\mu_n^H = 1650 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$$

For a typical case of a magnetic field of 1.5 T and a 300 μm thick detector, the charge distribution of the holes shifts by approximately 7 μm while that for the electrons shifts by approxiomately 37 μm [31]. Thus collection of the *electron* signal in future collider experiments is liable to serious systematic effects, unless the *n*-strips are oriented at least approximately along the direction of drift induced by the magnetic field (the azimuthal direction in a barrel detector).

Finally, a reminder that in any silicon detector of thickness approximately 300 μm, the production of δ-electrons of significant range is quite a common occurrence, so the residual distributions will inevitably have a significant non-Gaussian tail, unless steps are taken to exclude large-signal clusters, with the attendant loss of efficiency.

4.3.3 Electronics for Microstrip Detectors

We have seen that silicon microstrip detectors have developed and diversified to an extraordinary degree, due partly to the ingenuity of those involved, and partly to the tools and devices provided for them by the integrated circuit industry. As regards the readout electronics, the progress has been at least as spectacular, for the same two reasons. The current picture is in fact one of somewhat bewildering complexity, since the diversity of options is so great. Part of this diversity reflects the variable detector applications, but even for one single application (e.g. the ATLAS SCT) there is not yet unanimity among the experts as to the optimal approach. The issues are quite subtle and the boundary conditions keep shifting. In this section, we shall aim to take a general look at the principles leading to these various options, and make some remarks about the relevant areas of application. What is clear, however, is that the ASIC designer now has enormous power and flexibility at his disposal, so that a new application area is likely to lead to the very rapid evolution of one or more new readout schemes full of wonderful ideas to handle the peculiarities of that particular application.

Even from the very beginning of the ASIC initiative which opened the door for silicon microstrip detectors to find a home in collider detectors, there was not a unanimous approach. At that time, there *was* unanimity at the level of the functional requirements (amplifier, sample-and-hold, multiplexed analogue output) but two technological solutions; nMOS [12] and CMOS [11] were pushed by different groups. In the event, the 'low and slow' CMOS solution proved superior, largely due to the much lower power dissipation (around 2 mW per channel compared with ten times that for nMOS). This pioneering CMOS chip, the first of a family of CAMEX chips, was joined by others, of which the most commonly used are the MX (3-7) [32], SVX (1-3) [33] and AMPLEX [34] families. More recently, a bipolar chip for the front end electronics has made its appearance [35].

Why is the user of silicon microstrip detectors faced with such a large array of readout options? Some part of the reason is socio-cultural. There never was a 'standard' drift chamber preamplifier; different laboratories like to do their own thing, and this competition is extremely healthy in encouraging new ideas. But mostly, these various approaches have been driven by the need to equip detectors to work in increasingly varied and hostile conditions. Beam crossing intervals at SLC (8 ms) and LEP Phase 1 (22 μs) allowed very relaxed shaping times of 1 or 2 μs The detector modules were small (strip lengths $\leq$ 6 cm) and the radiation environment almost non-existent. Under these benign conditions, the ASIC designers were able to achieve spectacularly good signal-to-noise from a variety of single- and double-sided detectors. Moving to the Tevatron (originally 3.5 μs upgrading to 396 ns and eventually 132 ns), HERA (96 ns) and in future the SLAC and KEK *B* factories (4 ns) and LHC (25 ns) represents a phenomenal challenge. Compounded with the escalating beam crossing rate is the need to increase the module sizes (strip lengths of 12 cm will be used in the large ATLAS SCT, for example), plus the fact that the detectors at all hadron machines will encounter significant, if not fatal, radiation damage. Some relief is provided by cooling the detectors to reduce leakage current, but for the most part it has been up to the chip designers to get the physicists out of a very uncomfortable situation. This is a rapidly evolving story, and it is far from clear where we shall end up. In the case of the LHC detectors several critical decisions have to be taken over the next year, and these will be based on the results of much hard work going on

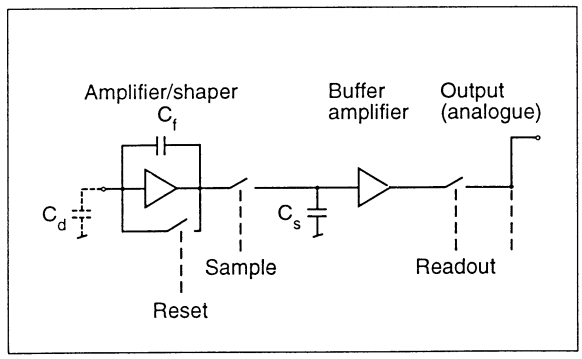

Fig. 28 Block diagram of one channel of a typical analogue readout chip.

in design labs and in test beams. Let us review in very general terms the main approaches, all of which are certainly appropriate for some applications.

Firstly, the generic analogue chip comprises typically 128 channels, one of which is shown in its essentials in fig. 28. The amplifier/shaper may include a CR-RC filter. It has been shown [36] that more sophisticated filtering schemes do not lead to a major improvement in noise performance. On receipt of a trigger, the signals are sampled and stored on capacitors C_S, which are read out (sequentially for each channel on the chip) via the analogue output, for remote digitisation. Such a readout chip minimises the logic local to the detector (and hence is optimal from the viewpoint of power dissipation, which is usually an important issue), but it cannot be used in high rate environments where even the first level trigger appears after several beam crossings. The most obvious response to this situation is firstly to reduce the shaping time so as to retain an analogue signal which is unambiguously associated with its beam crossing. However, this causes inevitably a penalty in noise performance, and may not be necessary. Given the sparsity of the tracks in the detector, each strip has a low probability of being hit on successive beam crossings. Then one may retain a longer shaping time, and use a filtering approach [37] to recover the fast timing information by deconvoluting the sampled voltages of a shaped pulse, to retrieve the original impulse signal with high precision. This ingenious approach may extend the range of applicability of CMOS front-end electronics into the realm of LHC operating conditions, and has been adopted by the CMS collaboration. Their analogue signal (50 ns shaping time) is sampled at the beam crossing rate of 40 MHz. The samples are stored in an analogue pipeline of 128 cells and, if a positive level-1 trigger signal is received, are deconvoluted by the analogue signal processor. All this happens in parallel for each channel.

The stored signals are read out at leisure via a multiplexer, connected off-chip to an electro-optical modulator. This consists of a multi-quantum-well device which amounts essentially to a mirror of voltage-controlled reflectivity. Consuming almost no power, this device permits a change of reflectivity from 30% to 60% by changing the voltage across an InP/InGaAs sandwich [38]. The device is connected to an optical fibre, at the remote end of which is the drive laser, receiver module, flash ADC and event builder memory. The beauty of such links is that they permit very high speed data transmission with almost no power dissipation at the detector

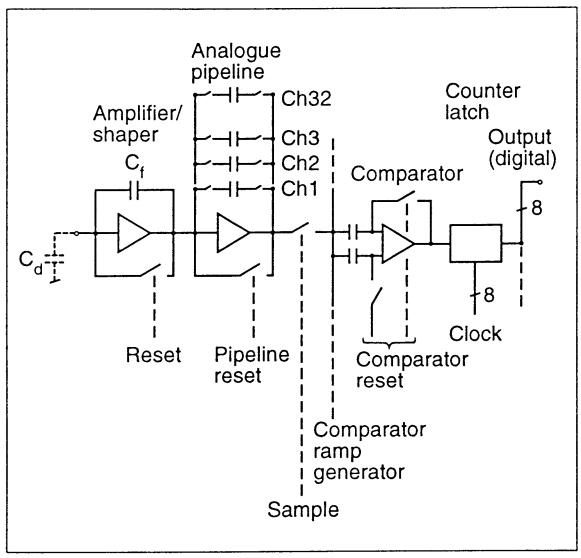

Fig. 29 Block diagram of one channel of a typical digital readout chip, of the SVX type.

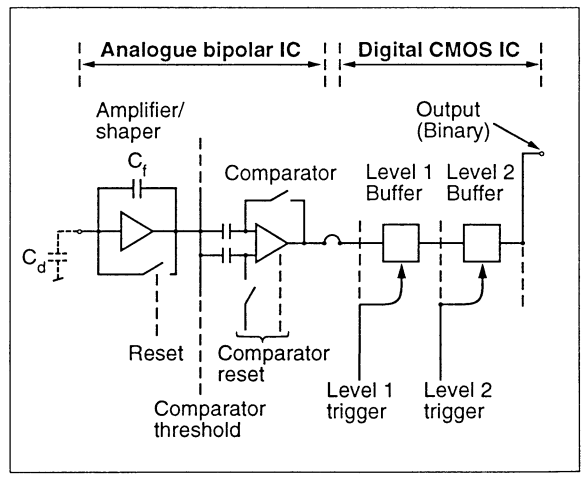

Fig. 30 Block diagram of an FEE system based on a bipolar analogue chip followed by a digital readout chip (binary readout).

end. Used (as here) in analogue mode, they permit 7 bit resolution which is entirely adequate for microstrip detector applications.

The SVX family of readout chips has pioneered the digital approach. An example is shown in fig. 29. Analogue signals are again put into a pipeline (one per channel). On receipt of a level-1 trigger, the relevant signal is transferred to a storage capacitor which serves as one input to a comparator used as a Wilkinson ADC circuit. The other comparator input is ramped at a fixed rate, and the time to reach equality of input is stored digitally as a measure of the signal amplitude. The digital data are then read out via a multiplexer.

Finally, we consider the bipolar option. Bipolar IC technology has been making great strides in recent years, and it has become possible to shrink amplifiers down to a pitch of 50 μm, as has been true for some time with CMOS systems. As a result, stray capacitances have been greatly reduced, and furthermore very small transistors can be made with high bandwidth and low current. In short, the power dissipation has dropped to an extremely competitive level. At hadron colliders, even with cooled detectors, the problem of leakage current in long-strip modules after a few years of radiation damage will be considerable. The shot noise associated with the leakage current tends to favour short shaping times as opposed to the longer shaping time with deconvolution mentioned previously. The lower limit on the useful amplifier shaping time is given by the charge collection time of typically 20 ns. Below that, one encounters increasingly severe signal loss (the *ballistic deficit* effect). The superior transconductance of the bipolar transistor compared with CMOS (even if run in the weak inversion mode) suggests that to achieve adequate signal-to-noise performance for long microstrip modules in fast readout conditions, the bipolar option may be superior.

A disadvantage (possibly minor) of the bipolar approach is that (due to the near non-availability of rad-hard bi-CMOS), one necessarily has an analogue chip followed by a digital CMOS readout chip. Doubling the number of wire bonds in the system is not a major overhead, and there are advantages. For LHC applications,

the size of the digital processing chip is such that yield is a significant consideration. Having the analogue front-end as a separate chip may be more economical overall.

This bipolar/CMOS combination has been used with excellent performance in the demanding environment of the ZEUS Leading Proton Spectrometer (LPS) at HERA [39, 40]. The basic system (fig. 30) consists of a bipolar amplifier/comparator chip with 20 ns risetime, followed by a low power digital pipeline. Not only does the front-end break with tradition in microstrip readout systems, but so does the digital system. The designers have adopted the simple 'binary' approach of recording only the addresses of above-threshold strips, not the pulse heights. In fact, their system (which has been carefully designed to minimise common-mode noise) operates extremely stably with a constant threshold of 0.78 fC set for all channels.

Lack of pulse height information of course limits the spatial precision to $\frac{p}{\sqrt{12}}$, where p is the strip pitch, but as we have seen, this precision is in any case close to the limit achieved in nearly all systems. Furthermore, it is only in small radius vertex detectors that there is a major physics advantages in pushing the point measurement precision to the highest achievable value.

The readout system takes advantage of the hierarchical trigger structure of ZEUS, which will also be followed in LHC. In the ZEUS application, they use a synchronous level-1 buffer of about 6 μs followed by an asynchronous level-2 buffer. Data are thus stored on-chip until a valid level-2 trigger arrives after about 1 ms.

All these considerations of readout options are complicated by another question, that of radiation damage. The move to hadron colliders has stimulated a major effort to develop rad-hard versions of the local detector electronics.

In the case of CMOS, a number of companies (Harris, UTMC, Honeywell and DMILL) are involved with the chip designers already mentioned. For example, a 100 Mrad-hard version of the MX7 chip in 1.2 μm CMOS exists. These chips tend to somewhat exceed the 50 μm channel width, but for applications such as the LHC SCTs this is acceptable. One cloud on the horizon is that, with the downturn in military spending, there is less funding for development of rad-hard electronics. However, as the industry moves into sub-micron processing, the devices have improved radiation resistance as a by-product (thinner oxide, etc.) so the trend may be to add a few steps to achieve adequate performance from a process not specifically developed for optimal radiation hardness.

For the bipolar ICs, the radiation damage situation is more favourable, due to the lack of sensitivity to oxide charge. The main cause of deterioration is bulk damage, which results in a reduction of the current gain β at high doses. Typically an *npn* transistor suffers a β degradation of approximately a factor two after 5 Mrads. The circuit designer can allow for such degradation, which makes these ICs usable at all but the smallest radii needed for vertex detectors at LHC. This region (as we shall see) is territory almost certainly out of bounds for silicon strip detectors due to the radiation damage in the detectors themselves.

Very recently, one company, DMILL (LETI), have produced some bi-CMOS chips using a rad-hard process. Whether they will find a sufficient market to sustain this initiative, and if so whether these will offer a way to the future for HEP detectors, remains to be seen. At least for the time being, the combination of bipolar chips with rad-hard CMOS digital chips appears to be the safest means to satisfy our requirements.

Thus, in conclusion, both the CMOS and bipolar ICs we have discussed can, it appears, be designed to tolerate the worst radiation conditions likely to be encountered by silicon microstrip detectors. The inevitable noise degradation due to growth of leakage current in the detectors, plus other detector-related issues to be discussed in Section 6, are what finally limit the scope for these detectors. There is no possible cure in the electronics for these deficiencies, once they reach an unacceptable level in the detectors.

4.4 Physics Performance and Future Trends

Silicon microstrip detectors were originally developed as vertex detectors for charm physics at fixed target experiments. Here, with the benefit of the high track momenta, they were able to achieve excellent impact parameter precision and hence clean reconstruction of a wide range of charm particle decays.

The move to e^+e^- colliders (initially SLC and LEP) called for much larger detector areas (and here the detector manufacturers were well able to oblige) and much more compact electronics (and, as we have seen, the ASIC designers solved these problems for us). Nevertheless, the physics capabilities of the detectors took a step backwards. Due to the lower particle momenta and the large radius beam-pipe (initially 10 cm at LEP, eventually reduced to 5.5 cm) the impact parameter precision for tracks in hadronic jets was relatively poor. Non-specialists were at first understandably ignorant of the situation, because all groups were proudly demonstrating beautiful miss-distance plots based on muon pairs. The situation for tracks in jets was, of course, much worse. The Holy Grail for vertex detectors is to present to the experimentalist a clear topology of the event, with high efficiency for associating all tracks uniquely with their true parent vertices. Fortunately for the detector builders, there is a host of lesser objectives which are still extremely useful for physics. The long lifetime of beauty particles means that b-tagging is relatively straightforward. The cleanliness of the $\tau^+\tau^-$ signal at the Z^0 means that lifetime measurements (though imprecise at the individual event level) can be made accurately, given high event samples. Areas such as charm tagging and the separation between charged and neutral B states are much more challenging.

The 1- and 2-dimensional microstrip vertex detector systems at LEP have covered the range of radii typically 60 to 110 mm, and (in their finally upgraded forms), delivered precision in impact parameter as a function of momentum p GeV/c of:

$$\sigma^b_{XY} \approx 20 \oplus \frac{80}{p \sin^{3/2}\theta} \ \mu\text{m}$$

and

$$\sigma^b_{RZ} \approx 30 \oplus \frac{80}{p \sin^{3/2}\theta} \ \mu\text{m}$$

With an average track momentum of about 5 GeV/c, this implies a mean impact parameter precision for normal incidence ($\theta = 90^{\circ}$) of around 30 μm. For reasonable topological vertexing (including charm) one would like to do 5-10 times better than this. Another problem for the extraction of physics with microstrip detectors is that of poorly understood tails on residual distributions. These are presumably due to a combination of factors such as energy loss fluctuations, δ-electrons, cluster merging (by no means negligible in the core of jets) and so on. The general approach has been to artificially broaden the Monte Carlo residual distributions to match the data. This is only a partial solution since it ignores the correlations that are surely present (e.g. a pair of tracks closely spaced in one view, giving poor co-ordinates on *all* planes due to partial merging of clusters).

The overall picture is one of impressively high efficiency and purity for b-tagging, with flagging performance in the more challenging areas. For LEP2, the b-tagging requirement is considered to be of paramount importance (Higgs and SUSY search). To do better as regards topological vertexing at the Z^0 would have required a smaller beam-pipe, giving a shorter extrapolation length to the interaction point (IP), and hence better impact parameter precision. But then, the track merging problem on the inner barrel would have been more severe. In any case, the time for such discussions is past.

The pioneering silicon microstrip vertex detector at hadron colliders has been the SVX family (same name as their readout chips) at the Tevatron. SVX1, the original detector, played a crucial role in the discovery of the top quark, again by performing the relatively simple task of b-tagging. This is the first major discovery in particle physics in which a silicon vertex detector has been essential in achieving a convincing signal, and I am sure it will not be the last. After all the technical complications we have been discussing, it is somewhat comforting to note that the detector used for the top discovery was a single-sided, DC coupled, low signal-to-noise, radiation-soft detector! Such a detector would never survive the conditions after the Tevatron upgrades, and this vertex detector has already been upgraded at least once.

New microstrip vertex detectors are on the way. CLEO II has one (on a small radius beam-pipe, necessarily tackling the more challenging requirements of charm decay) and the SLAC and KEK B factories are building them, primarily to measure the longitudinal position of the B and $\overline{B}$ decay points with respect to the IP.

The ZEUS LPS set the trend for microstrip detectors to be used as momentum spectrometers in high intensity conditions in which wire chambers could not survive. This trend continues with the D0 silicon tracker (approximately 5 m^2) and the CMS and ATLAS SCTs (40 m^2 for ATLAS). What has happened is that the high energy, high luminosity hadron collider environment has become too unfriendly for wire chambers on almost any radius. Therefore, silicon microstrips are taking over as detectors with tracking precision < 100 μm, and able to handle the hit rates and the integrated radiation doses. For such large detectors, spatial precision is less of an issue (it will in fact be a challenge to build them with few micron stability, so the intrinsic detector precision may not be the driving factor). This is one reason for the interest in (for example) binary readout.

However, these detectors clearly have their limitations. There is a nasty hole of radius about 30 cm in ATLAS and CMS within which microstrip detectors dare not venture, due (as we shall see in Section 6) to problems of radiation damage. With the huge event multiplicities, track merging would also be very serious. In this region, silicon pixel detectors may find a home, and (at the smallest radii) other detector options, as we shall discuss in Section 7. The overall result is that the main emphasis in the world of silicon microstrip detectors at the energy frontier has shifted from aiming to achieve the ultimate in spatial precision with the minimal detector thickness (including pushing for double-sided detectors) to aiming to cover very large areas as economically as possible, with electronics having an extremely high rate capability. The pressure for the most economical solution may argue against double-sided detectors, particularly since the material associated with the additional silicon layers is not seriously detrimental to the momentum resolution of the tracks that are important for physics. Fortunately, the size of the collaborations has grown at least a fast as the areas to be covered, so there in every reason to believe that they will succeed in these challenging tasks.

To describe any advanced technology as mature is usually misleading. Silicon microstrip detectors and particularly the associated electronics will continue to evolve for many years. However, as the OPAL collaboration demonstrated when they decided they needed a silicon microstrip vertex detector to retain LEP competitiveness, it is possible starting from scratch to build a sophisticated detector with this technology within a year, provided one does not try to invent a lot of new features.

For e^+e^- linear 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 1-D microstrip detectors and the 2-D 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 2-D images of excellent quality.

More usually, the term pixel detector is taken to mean a device equipped with a 1- or 2-D array of pixels (picture elements). The 2-D 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 1-dimensional precludes any application in which the desired output is some form of picture. Hence, pixel devices are of much greater inter-disciplinary 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? Fig. 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 recognisable 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 3-particle case shown (6-fold ambiguous) but for a high energy jet of 10 tracks there are 3.6×10^6 possibilities! 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

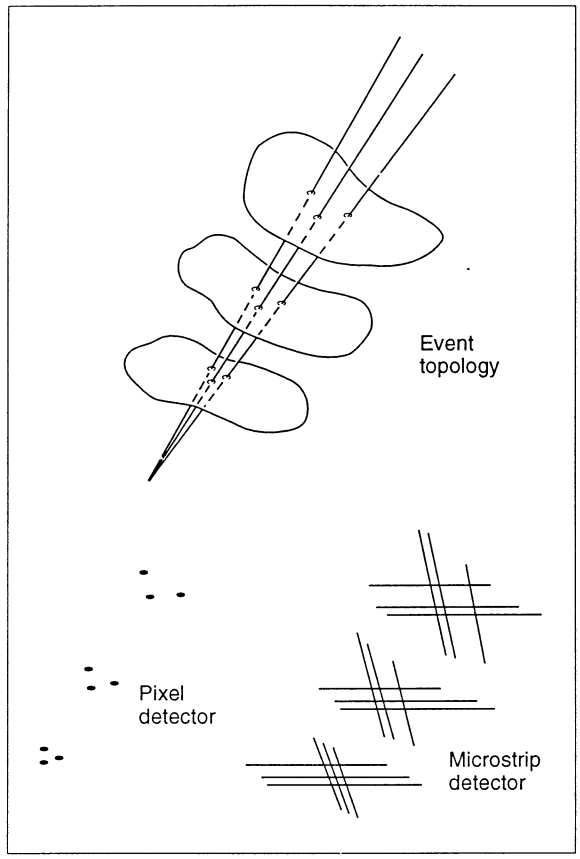

Fig. 31 Upper sketch: a few tracks traversing an unspecified set of three detector plates. Lower sketch: resulting information in case of 1 and 2-D detector types.

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 Section 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 performance. 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 (minimising the problem of δ-electrons) and for angled tracks (minimising 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.

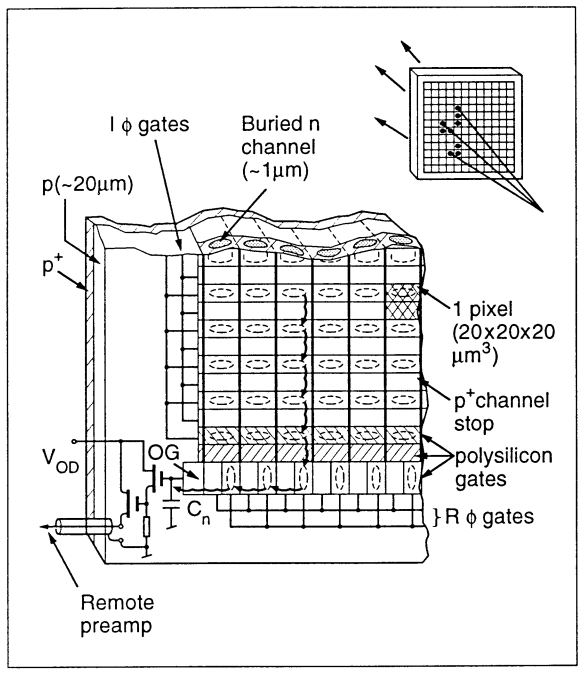

Fig. 32 Upper right: sketch of charge storage in a CCD detector traversed by a number of ionizing particles. Lower left: corner region of CCD showing the principal structural features.

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 (CCDs)

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 2-D charge pattern to a serial train of pulses, well suited to display on a video monitor. The CCD was invented in 1970 [42]. Devices of this pioneering design (so-called surface channel CCDs, 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 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 CCDs, 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 CCDs 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 CCDs. The commercial devices use interline transfer and are typically only about 2 μm thick (active layer). This is excellent for sharp colour 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 specialised 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 CCDs as high precision detectors of ionizing particles was first evaluated theoretically about fifteen 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 Section 3, how such a detector can be built. For more detailed information, there are some excellent books on CCDs [45, 46] as well as CCD conference proceedings and hundreds of published papers.

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 (fig. 33 (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 a *pn* junction by the introduction of a shallow (approximately 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 FETs, the conducting surface layer is termed a gate.

Let us now put some bias voltage onto the structure, as shown in fig. 33 (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 *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

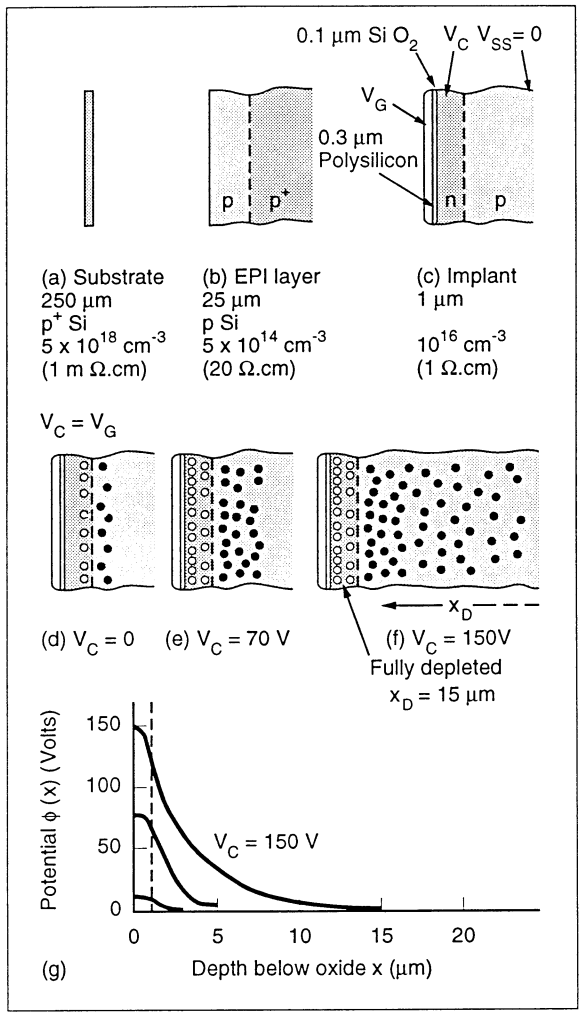

Fig. 33 (a) - (c) The successive stages in making a CCD-like structure (shown with increasing magnification). (d) - (f) The depletion process which would apply if V_c and V_G were increased together. (g) The corresponding potential distributions as a function of depth in the silicon.

fig. 33 (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 fig. 34 (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 *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 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 fig. 34 (c); look initially at the curve for $V_G = 0$. The quadratic

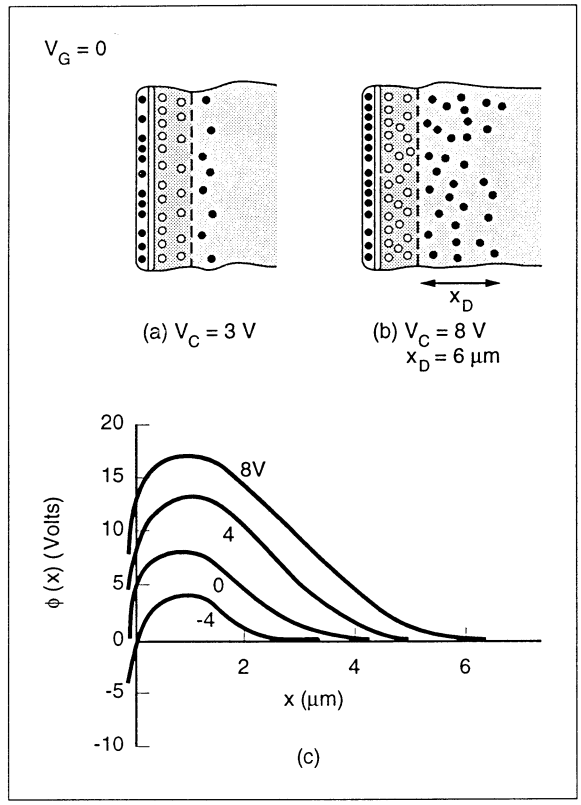

Fig. 34 (a) and (b) The depletion process in normally biased CCD operation with V_G negative with respect to V_C. (c) The corresponding potential distributions after channel pinch-off for various values of V_G.

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 CCDs 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 CCDs, 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 Section 3; see for example [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 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 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 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. 3×7 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.).

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.

Referring back to fig. 32, which shows a 2-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 CCDs. Some, for example, have more complex channel stops, *pnp* structures which can be used for anti-blooming or for fast-clearing the CCDs. Early devices of this type had dead bands between each pixel, a feature which made them unacceptable for most applications as particle detectors.

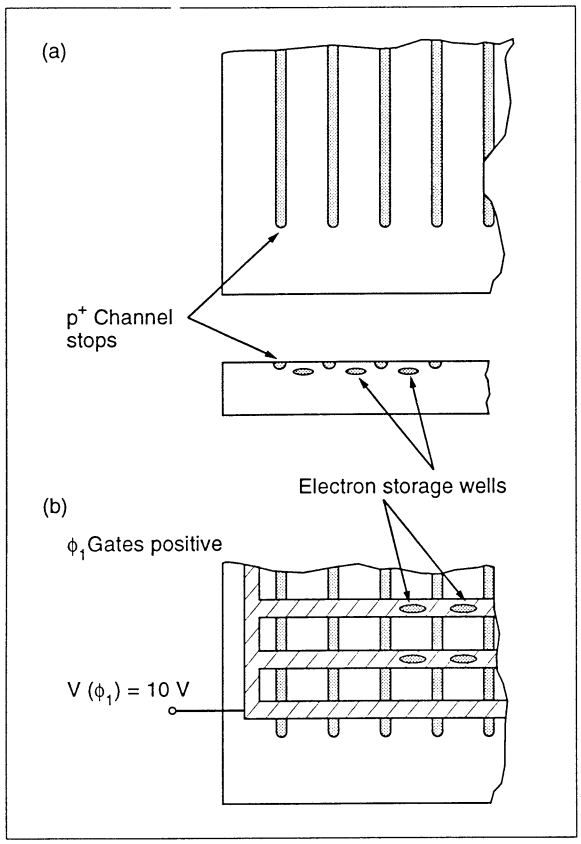

Fig. 35 Establishing the potential well structure: (a) Channel stops create potential barriers running vertically on the device. (b) Gates create horizontal potential barriers. The combined result is a matrix of localised wells, each of which constitutes a pixel.

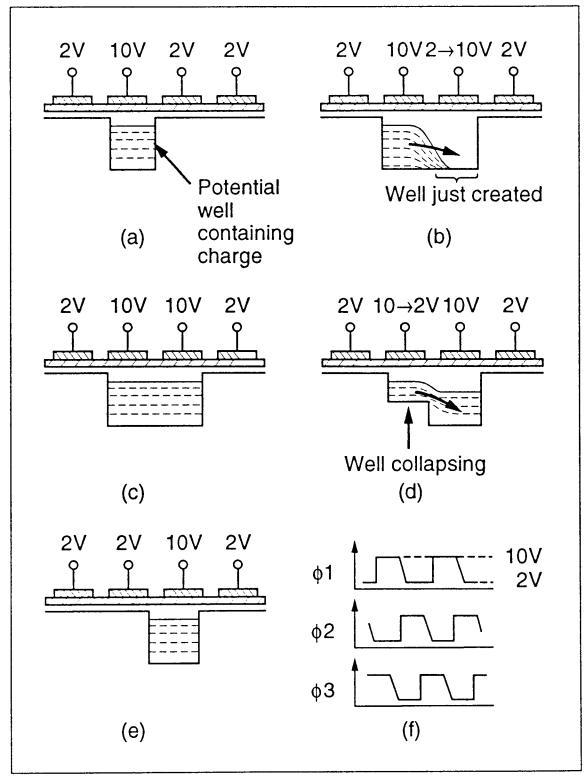

Fig. 36 After [46]. (a) - (e) Movement of potential well and associated charge packet by clocking of gate electrode voltages. (f) Clocking waveforms for a 3-phase CCD.

More recently, designs have been developed which permit fast-clearing while retaining adequate charge collection efficiency of MIP signals over the full active area.

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 interface at the back surface 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 neighbouring 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

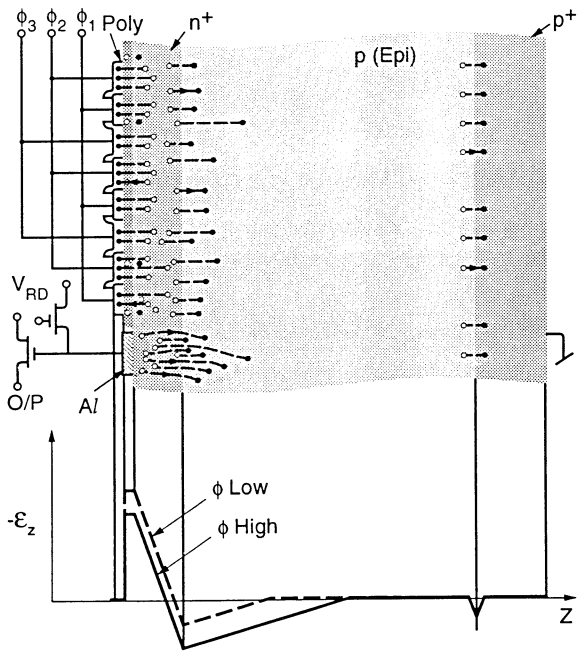

Fig. 37 Sketch of cross-section of a generic 3-phase MOS CCD. As in fig. 25, exposed fixed charges are shown by open circles (positive) and filling circles (negative). Also shown is the electric field distribution in regions of high and low imaging gate voltage.

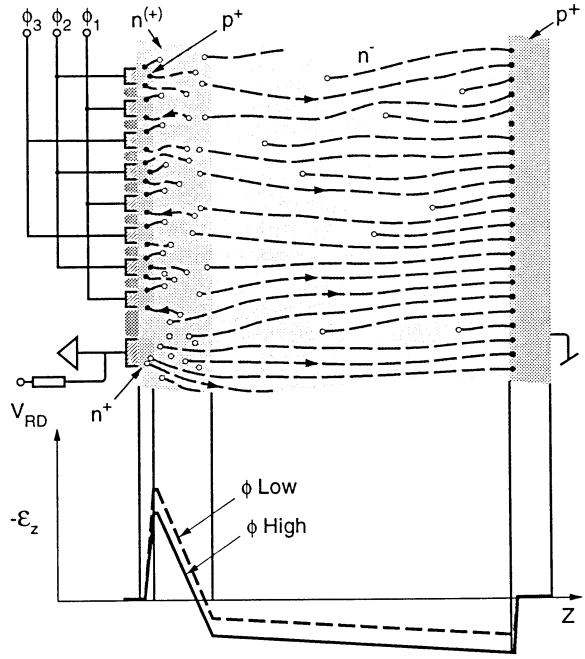

Fig. 38 Sketch of cross-section of a generic *pn* CCD manufactured on high resistivity silicon, and depleted over the full thickness.

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.

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* CCDs 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] CCDs. 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 CCDs) and of companies manufacturing these devices for scientific customers [52].

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 3-phases of a register, which are obviously deposited in separate operations. These 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 CCDs for MIP detection, simply does not work, as we found some years later. Small signals in CCDs 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 Section 6, radiation damage can cause serious growth in the density of these bulk traps. The operating temperature is a very important parameter in minimising this problem, since it profoundly affects the trap emission time constants. The problem of bulk traps affecting CTE in CCDs 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 CCDs operated at low temperature, see [54].

As well as the problem of traps of atomic dimensions, CCDs 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 centre of a large CCD. The large currents induced in the CCD structure by the voltage excursions in the parallel register (which covers nearly all 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 minimised. In an optimised 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 hundred 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 inter-connections). 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 metres distant, and carried in on approximately one thousand 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 2-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 10 to 100 times the 1 mV level. For slow readout systems, one can wait for this to settle down. A major challenge in reading CCDs 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 [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 twenty 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 2- or 3-stage output circuit, as shown in fig. 39. With an optimised design, the noise performance is dominated by the first stage, even in the case where the later stage FETs 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

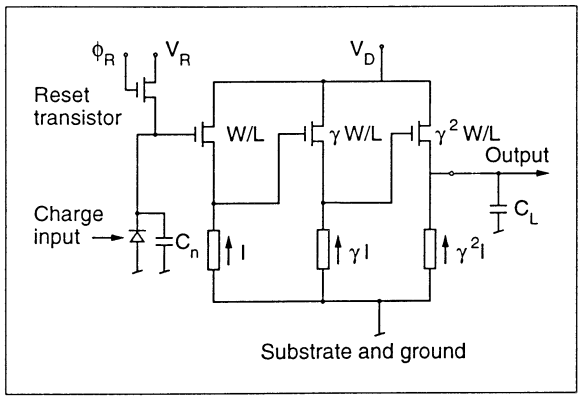

Fig. 39 Schematic diagram of a 3-stage output circuit.

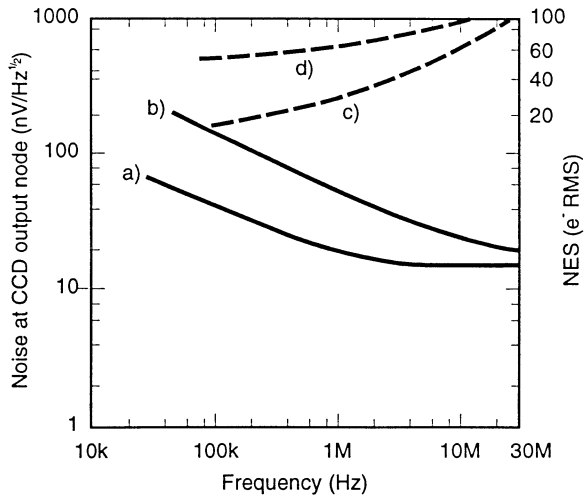

Fig. 40 Noise spectra of (a) buried channel and (b) surface channel first stage MOSFETs in a CCD (left hand axis). (c) and (d) show the corresponding CDS noise equivalent signals in RMS e^- (right hand axis).

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 [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 $\mathrm{Hz}^{1/2}$ at frequency f and $S_F(f)$ is the Fourier transform of the filter sampling function $S_F(t)$

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.

The procedure normally followed in vertex detector readout, where readout speed is to be optimised, 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 is 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 customised 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 CCDs. Inter-line transfer devices, which can offer 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 in conventional CCD designs is to clock the charge out via the output node. During the pre-trigger period, 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 fudicial region, at the detector design stage .

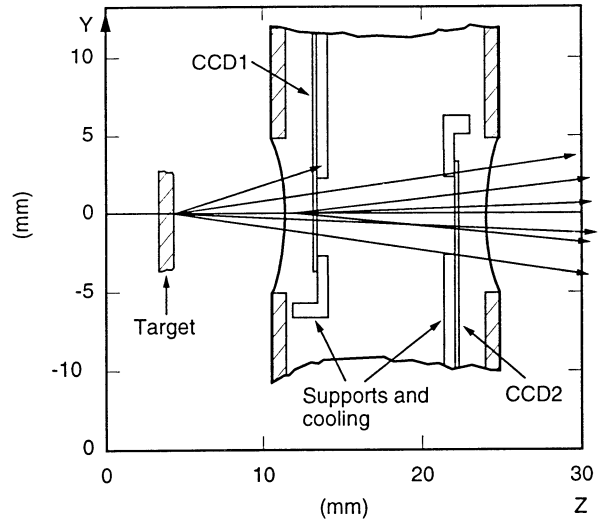

Fig. 41 CCD vertex detector for a fixed target experiment (NA32). Data are fast-shifted into the quiet regions above and below the spectrometer aperture (CCDs 1 and 2, respectively) prior to readout.

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 fudicial 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.

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 dumping the beam during each readout period would cause a significant deadtime loss. 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

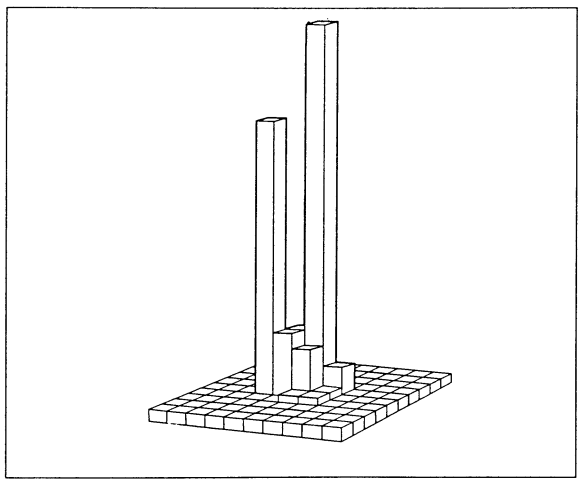

Fig. 42 Two MIP clusters separated in space by 40 μm, well resolved in a single CCD detector plane. Pixel size 20×20 μm^2.

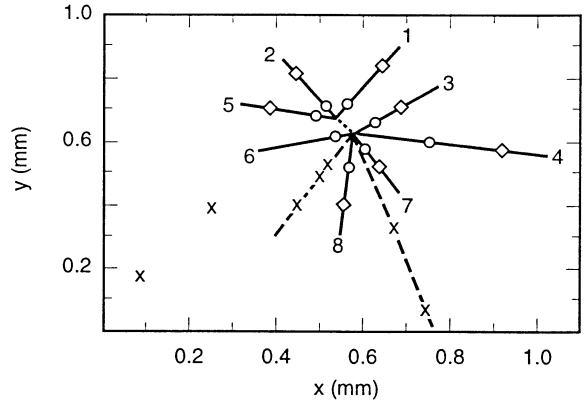

Fig. 43 Tracks from the IP and from a nearby charm decay in the NA32 vertex detector. Frame size 1×1 mm^2.

background. But in this case, one can take advantage of modern CCD design to use a multi-port 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 optimised balance between these extremes.

Most recently, the possibility has emerged for incorporating a fast clear capability in frame transfer CCDs, without loss of MIP efficiency. The idea is to build gated anti-blooming drains, allowing clearing times below 1 μs, in devices with reasonably thick (20-30 μm) epitaxial layers. A MIP traversing the drain region would lose a small fraction of the signal (that in the depletion region) but (by reference to Fig. 24) could still be detected with 100% efficiency by virtue of the signal collected by diffusion from the undepleted epitaxial material.

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 emphasise the opportunity presented to the vertex detector designer by fully customised CCD design.

5.2.3 Physics Performance and Future Trends

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

1. 2-D space point measurement, hence unsurpassed power for track reconstruction.

2. 2-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 reference [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.

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 Section 6. In an environment of high hadronic flux, one either has to exchange CCDs 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 APSs (next section) but one then loses some of the previously listed attributes, as we shall see. Each detector type has its own niche.

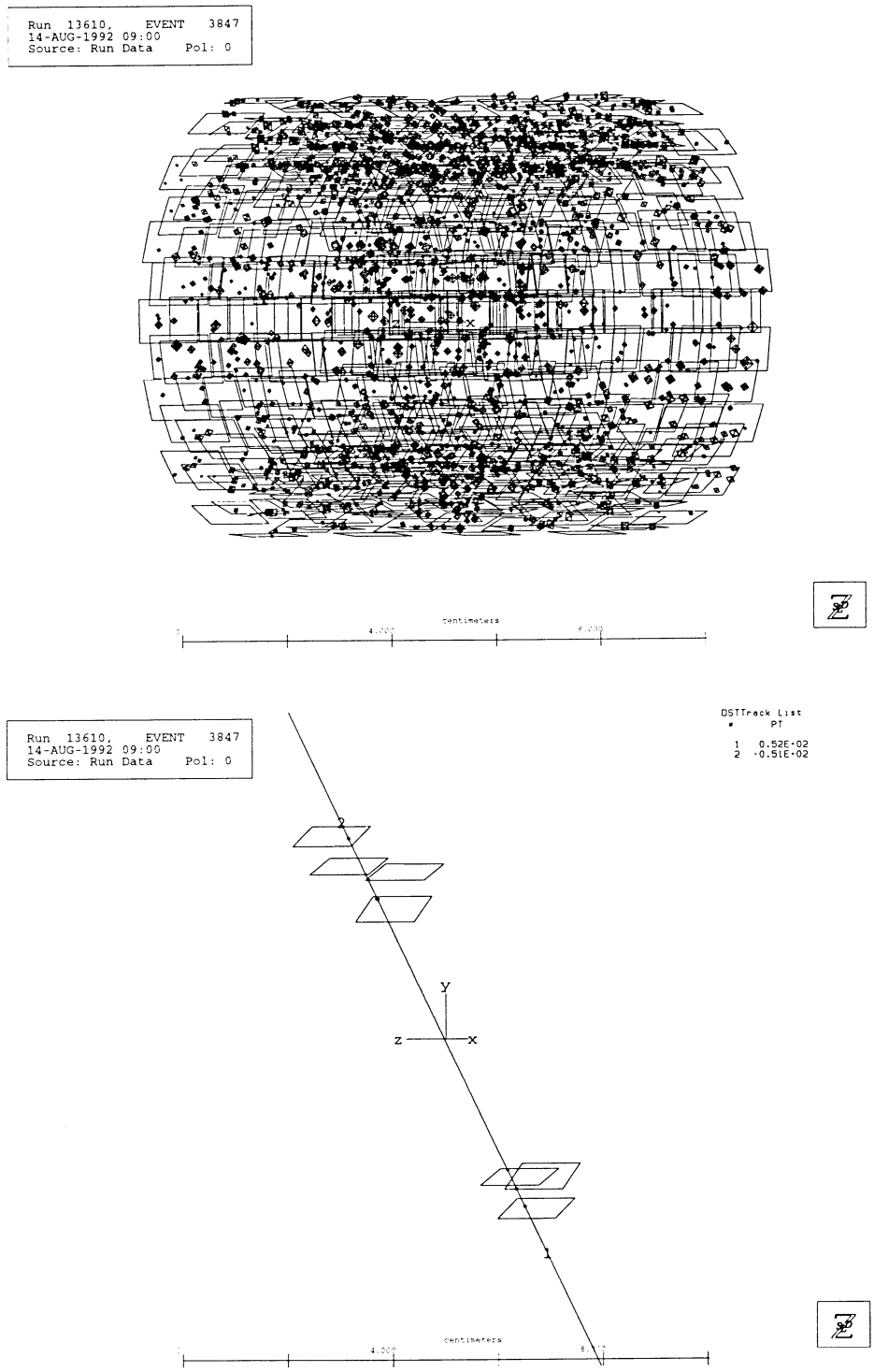

Fig. 44 (Top) Raw data (mostly SR X-ray hits) in the SLD vertex detector. (Bottom) The same event, with background filtered out by a drift chamber/vertex detector track linking algorithm. This proved to be a $Z^0 \to \mu^+\mu^-$ event.

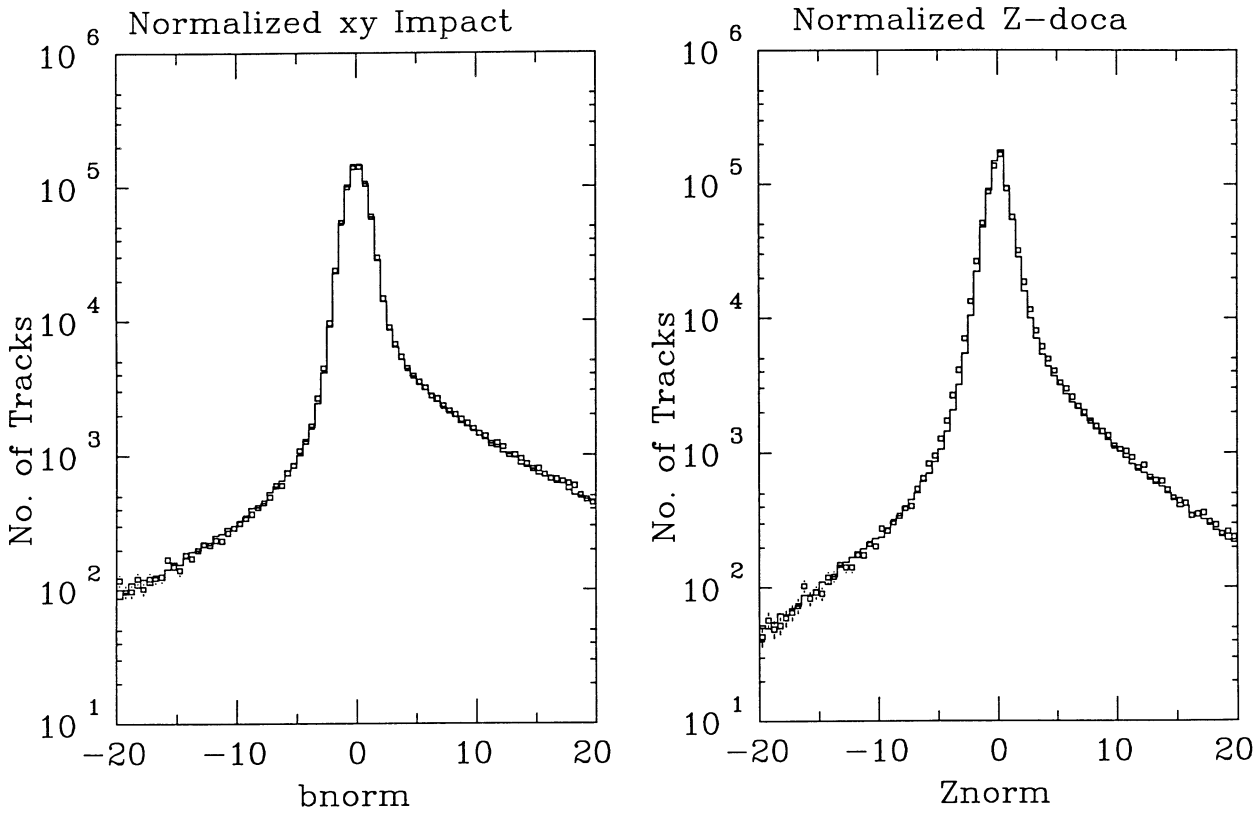

Fig. 45 Data (points) and Monte Carlo (histogram) distributions of impact parameter with respect to the IP in $Z^0 \rightarrow$ hadron decays (SLD experiment). The tails on the positive side are due to heavy flavour decays.

The availability of fully customised large area CCDs has opened the door for very exciting vertex detector development. 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 2-CCD ladders out of which the detector (figs. 47-49) is constructed. See reference [63] for a description of this 307 Mpixel detector.

For the future linear collider, one can be more adventurous. The CCDs 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 reference [64].

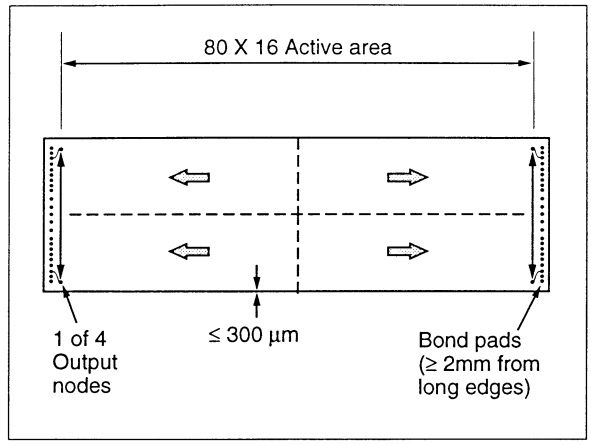

Fig. 46 4-port CCD developed for the SLD upgrade vertex detector. Chip area = 13 cm^2.

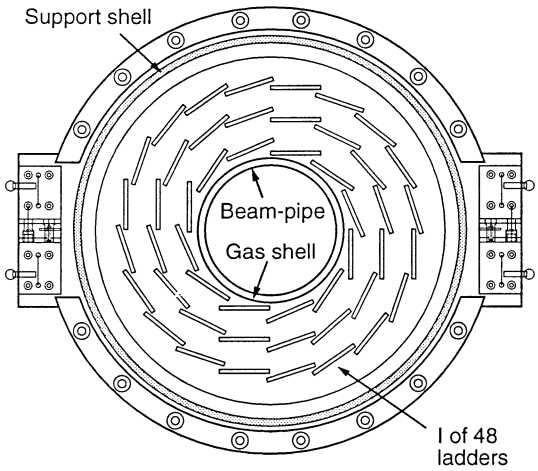

Fig. 47 Cross-section (XY view) of SLD upgrade vertex detector.

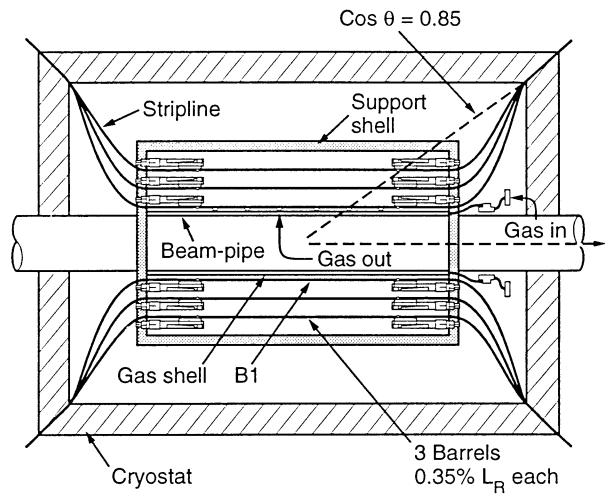

Fig. 48 Cross-section (RZ view) of SLD upgrade vertex detector.

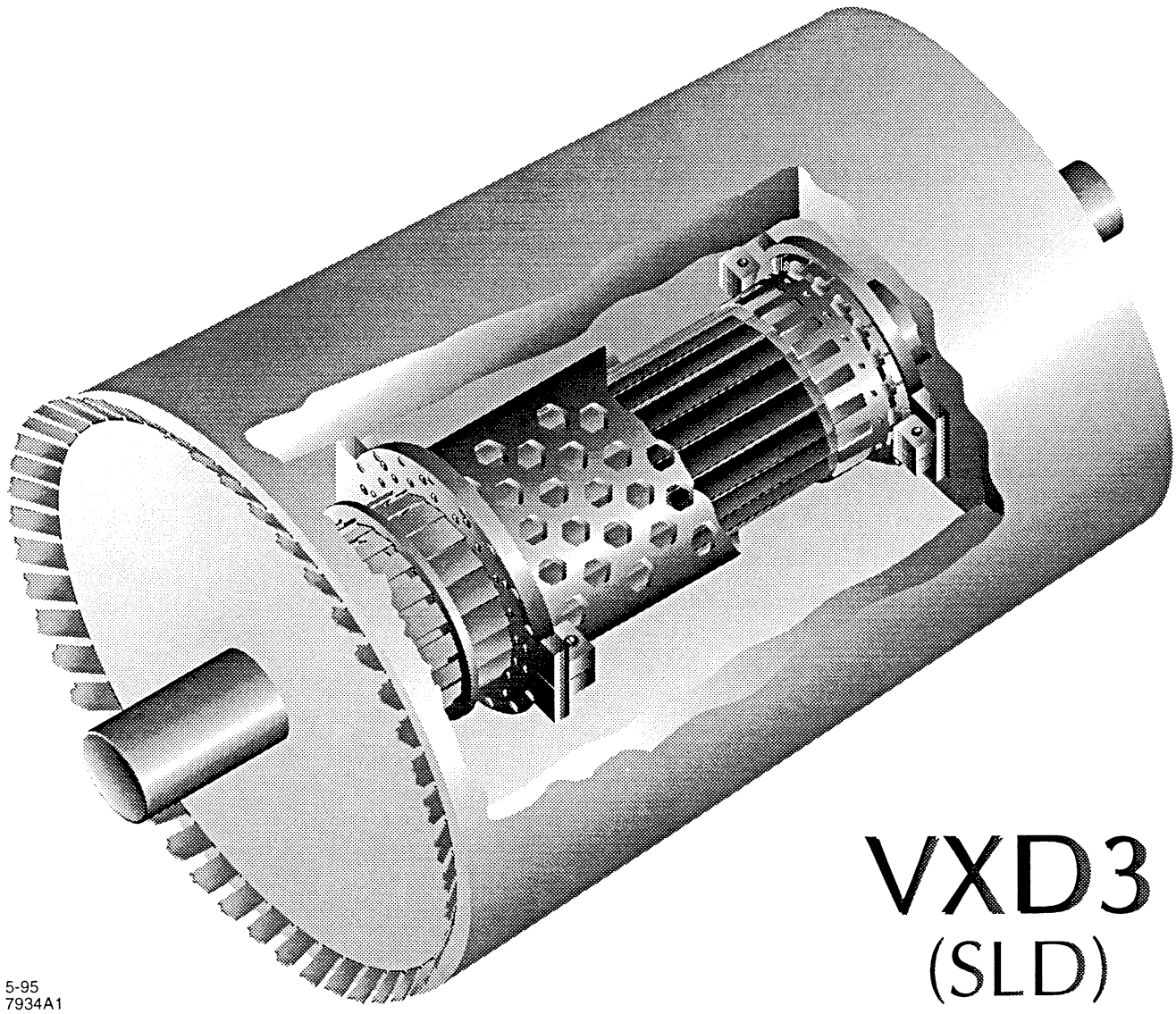

Fig. 49 Isometric drawing of SLD upgrade vertex detector.

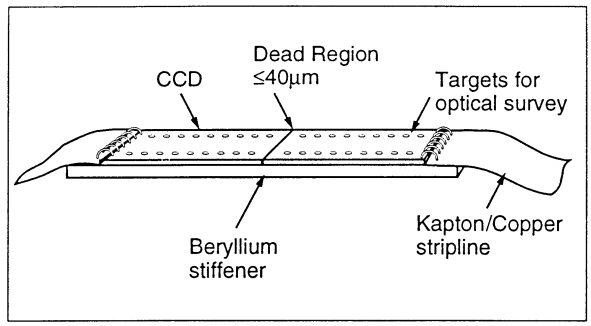

Fig. 50 Advanced 2-CCD ladder design for a vertex detector for the future e^+e^- Linear Collider. Active length ≈ 16 cm

By a combination of larger and thinner CCDs, 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

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

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 emphasised that the low power dissipation in a well-designed CCD detector (approximately 10 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 minimise effects of radiation damage; see Section 6.

5.3 Active Pixel Sensors (APSs)

Both in the wider commercial world, and in the area of scientific imaging, CCDs 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 emphasised. 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 physically fixed matrix over the device area, and each one is

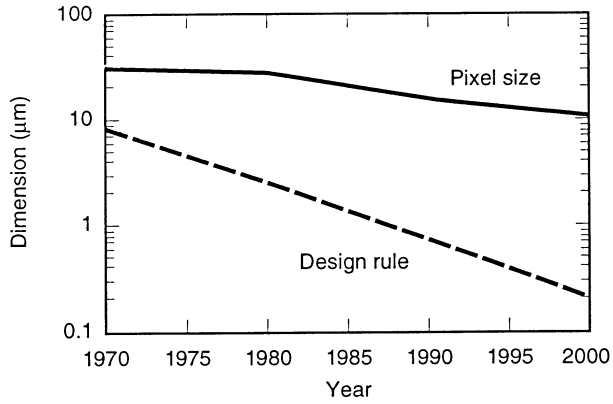

Fig. 51 Evolution of photolithographic feature size versus pixel size.

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 integrated circuit technology. Fig. 51 (after [65]) illustrates this point. A recent review of developments in this field is to be found in [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 CCDs. Since the readout can be non-destructive, 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* ICs. Building sensors of area many square centimetres 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 summarised as follows:

1. high speed gating. In contrast to CCDs, the aim is to latch signals and associate them with specific beam crossovers (BCOs) 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 Section 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 [67].

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 80's 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.

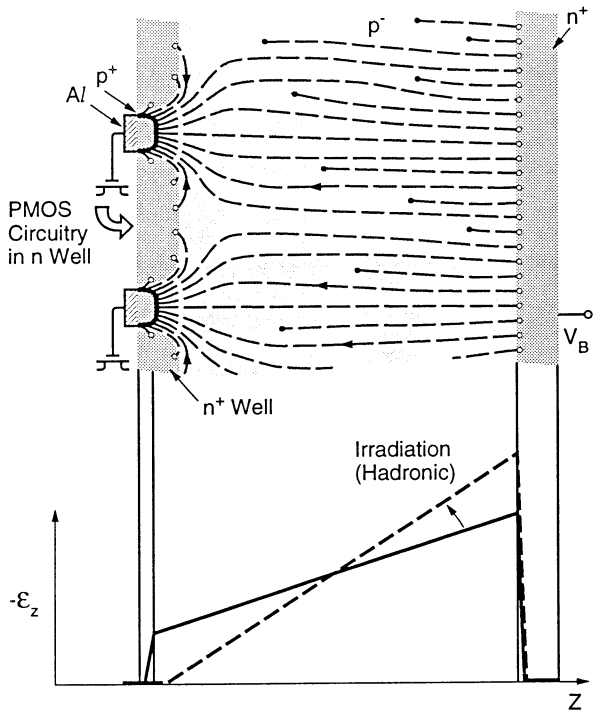

Fig. 52 Pixel structure of the generic monolithic APS. As with the microstrip detector (fig. 25) hadronic irradiation tends to take the detector out of depletion, losing the signal.

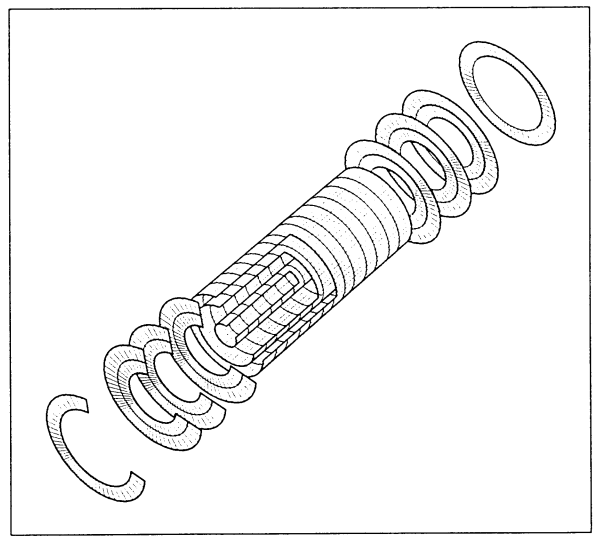

Fig. 53 Conceptual GEANT layout of a pixel vertex detector for ATLAS, consisting of three barrels plus endcaps. The innermost barrel (r = 4 cm) is not expected to survive for long at the full luminosity.

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. 10% 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 sub-micron 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 become possible). Secondly, the CMOS electronics needs to be sufficiently rad-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 Section 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 (reference [72]) involves the use of a *p*-channel JFET on a fully depleted high ohmic substrate (DEPJFET) for use as unit cell for pixel detectors.

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, 74] are very useful books on the subject, [75] provides a valuable current review, and interesting historical reviews can be found in [76, 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. Fig. 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 is proportional to the thickness of the oxide layer 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

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

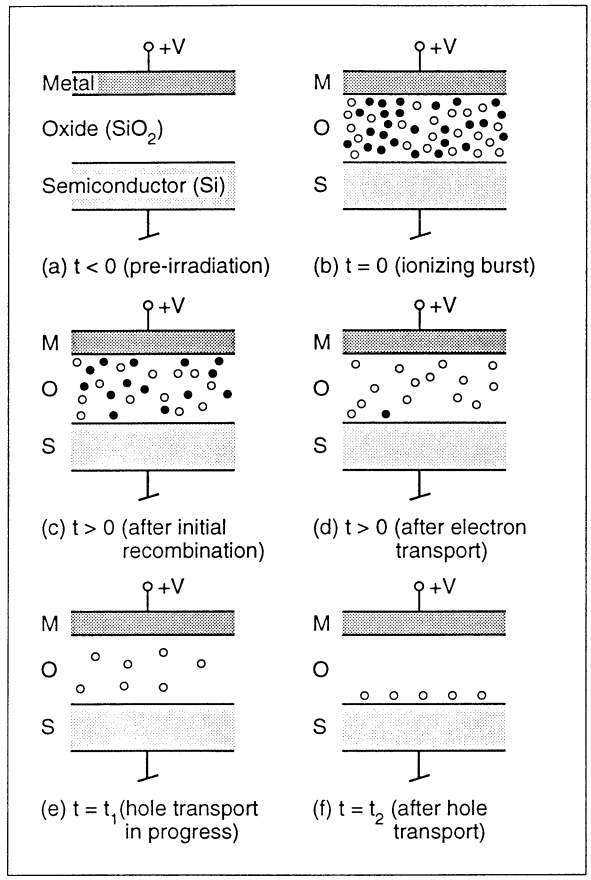

Fig. 54 Time development of charge distributions following a burst of ionizing radiation on a positively biased MOS structure.

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

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

For thin oxide, 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/V s and 20 cm^2/V s 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 immobilised, so there is no performance difference between soft and hard oxide. This however is not a serious 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 non-hardened 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, minimise 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 (JFETs, 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 reference [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 non-ionizing 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 references [80, 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 non-ionizing dose received, except that the effects are much reduced for electromagnetic radiation. In this case, as well as the low specific NEIL value,

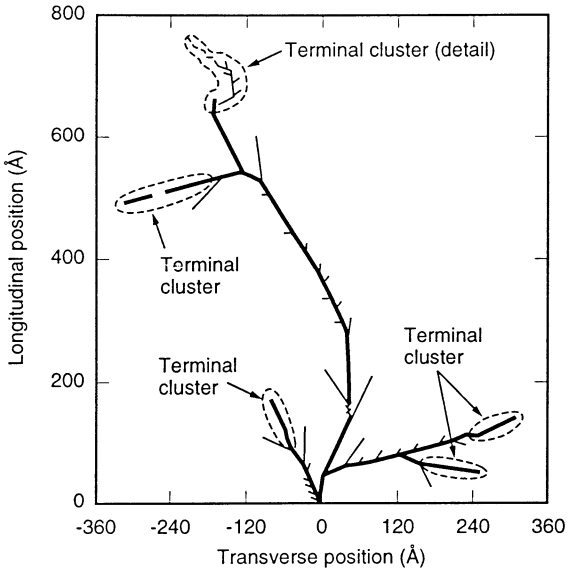

Fig. 55 Development of cluster damage due to a primary knock-on silicon atom of 50 keV, within the bulk material.

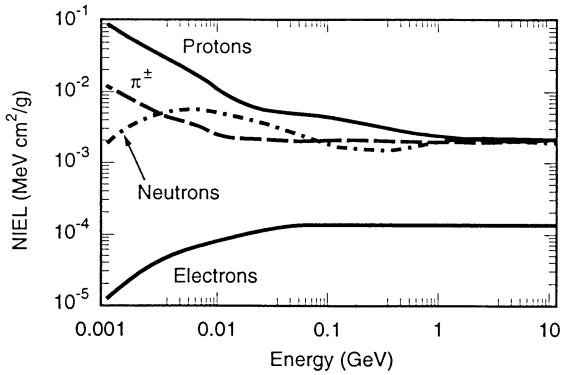

Fig. 56 NIEL for various particle types as a function of energy. A frequently used unit is the NIEL associated with a 1 MeV neutron.

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 ultimate physical and electrical effects, because both vacancies and interstitials are mobile, and can combine stably with atoms other than silicon in the crystal structure.

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 centres and hence seriously degrade the minority carrier lifetime. In addition, these extraneous generation-recombination centres 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, 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

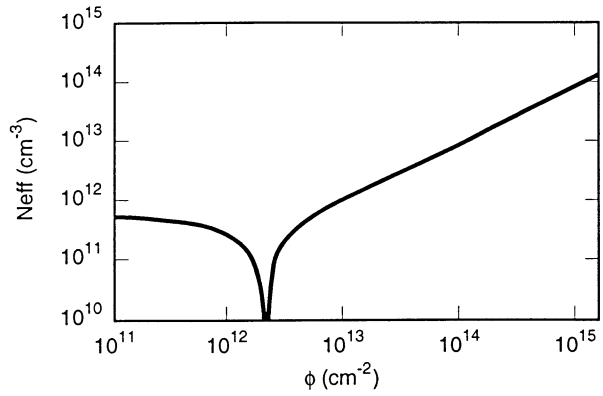

Fig. 57 Dependence of effective dopant concentration N_{eff} on fluence, at room temperature. The material, initially n-type, goes through type inversion for $\phi \approx 5\times10^{12}$ neutrons / cm^2 equivalent dose.

states, populating the entire band gap. Statistically, the Fermi level drifts to approximately mid-gap, 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 inter-gap 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 stabilise 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 behaviour, 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 behaviour 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 behaviour 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 summarised 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:

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

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

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

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

2. The mobile I and V centres 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)

$C_i + C_s \rightarrow CC$

$C_i + O \rightarrow CO$

$O + V \rightarrow VO$ $Si-A$ centre (99.9%)

$P + V \rightarrow VP$ $Si-E$ centre (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 folk-lore regarding the behaviour 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 for this to be a major player by comparison with some of the impurities (notably oxygen).

3. The authors hypothesise 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 hypothesise 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 non-zero offset, but it seems 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 silicon 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} n / cm^2 equivalent. This is presumably related to the high density of trapping centres generated, and probably implies a basic limit to the tolerable radiation dose for such thick detectors, at around the 10^{15} 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 doses as a function of radius are listed in the following table, for a seven year run at $\mathcal{L} = 1.7 \times 10^{34}$ cm^{-2} s^{-1}.

R cm	Ionizing Dose (Mrads)	Fluence 1 MeV n/cm^2 $\times 10^{-14}$
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 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 will be discussed in Section 7.

6.4.2 CCDs

For use as vertex detectors, CCDs 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.

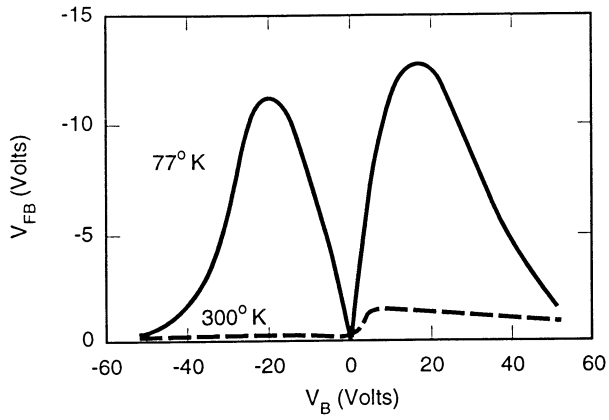

Fig. 58 Flat band voltage shifts after 100 krads of ionizing radiation across a hardened gate oxide.

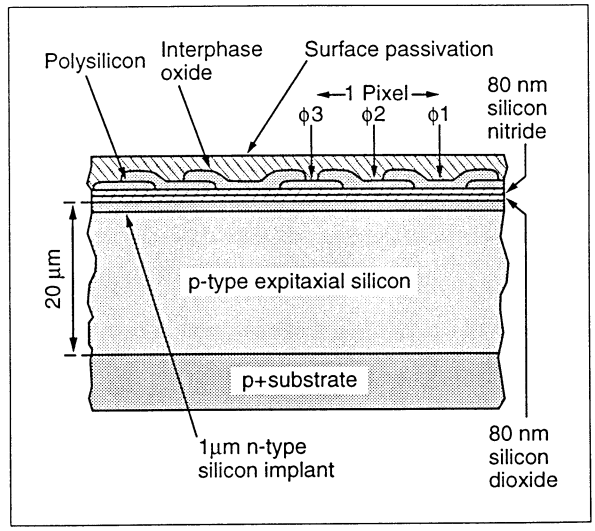

Fig. 59 Gate structure of a modern 3-phase CCD register, designed to avoid potential wells due to radiation-induced charge build-up or other spurious charge in the oxide or surface passivation layers.

Regarding ionizing radiation, the effect to be concerned with in CCDs 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). Fig. 58 (based on reference [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 Fig. 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 CCDs (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 CCDs 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. Fig. 36 is an over-simplification, 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 CCDs, 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 centres (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 first take a qualitative look at the situation. As the charge packet is transported from gate to gate (within a pixel or between neighbouring pixels) *vacant* traps that lie within the storage volume of the charge packet will tend to capture electrons. If the traps are already 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. 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 multi-parameter problem with some room for manoeuvre.

Let us now look at the process quantitatively.

Assuming all traps 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 (3 for a 3-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^- [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 3-dimensional potential well in which they are stored. The volume of this potential well can be reduced (by

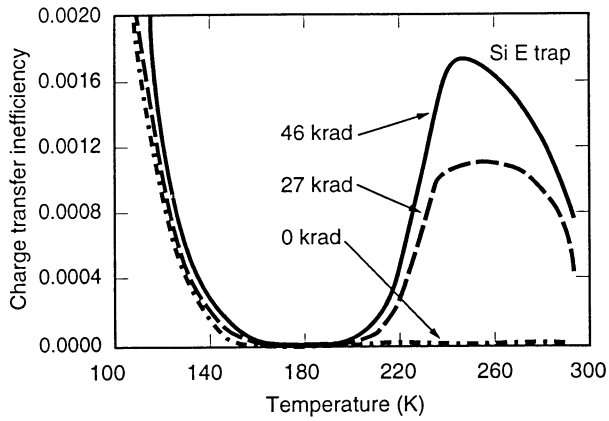

Fig. 60 From reference [90], effect of radiation damage on CTI in a CCD, as function of operating temperature. Irradiated with a Sr^{90} β source.

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

$$\tau_e = \frac{\exp[(E_c - E_{tr}) / kT]}{\sigma_n X_n \nu_n N_c}$$

The terms in the denominator are in turn 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 3-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. Fig. 60 (from reference [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 centres increases, but the effect on CTI can be minimised 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 CCDs.

For hadronic irradiation of CCDs, 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 centre 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 programme. One might select a temperature low enough to 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 a Mrad during its working life.

The issue therefore really only arises in the case of non-multiplexed 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 Section 6.4.1 apply equally to the electronics. For the Central Trackers, (reaching in to $R \approx 30$ cm but not below) radiation resistance up to around 10 Mrads and 2×10^{14} 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 defence 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 non-defence markets are inadequate, and have ceased to offer facilities for rad-hard electronics. As has been mentioned, the trend towards sub-micron processing lends itself incidentally to improved radiation resistance, 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 > 5×10^{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 physics 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 (at 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 challenging 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 readout ICs, bump-bonded together, may be made of different materials, either or both of which may be non-silicon. 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 Section 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 summarised 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) CCDs 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 ICs, 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 it 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 vapour 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 6×10^{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 10 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.

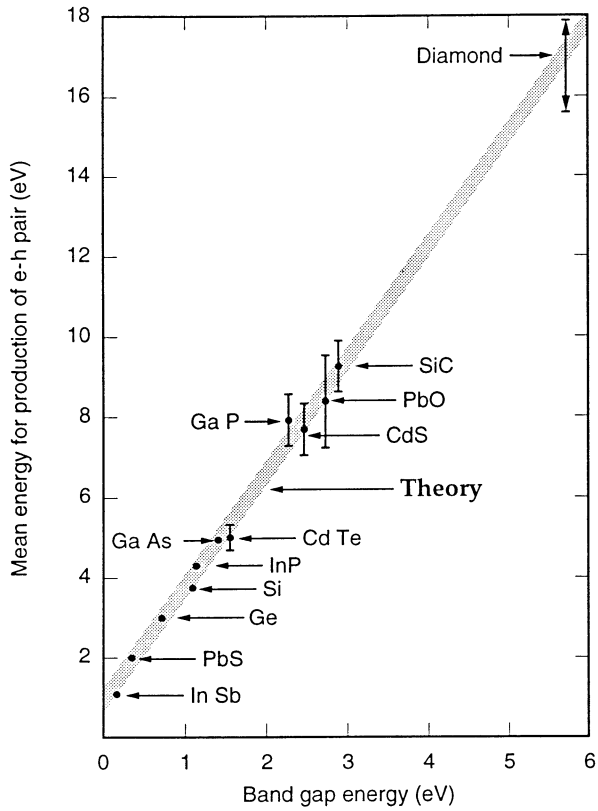

Fig. 61 Band gap and pair-creation energy, for various detector materials.

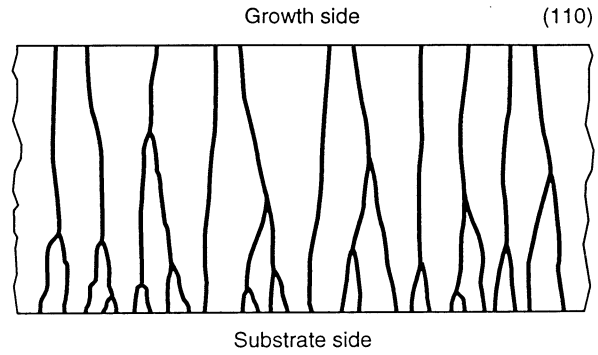

Fig. 62 Evolution of crystalline defects in CVD diamond as a function of thickness of deposited layer.

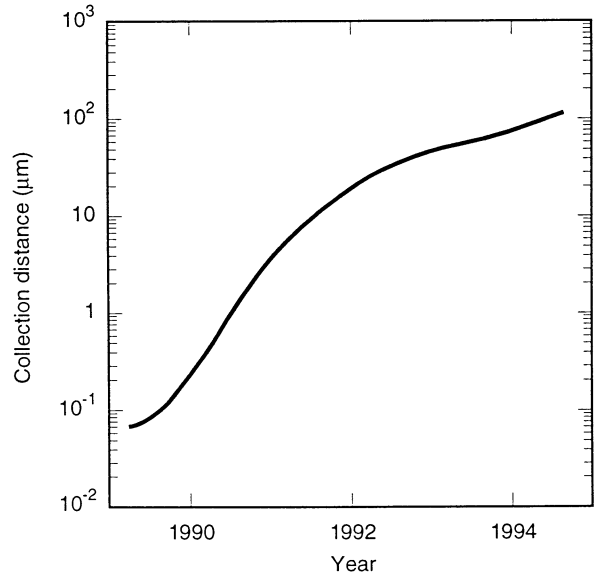

Fig. 63 Time development of collection distance in CVD diamond, from reference [99].

7.3 Local Electronics

For the high radiation vertex detector environments where silicon-based ICs are (probably) ruled out, we are almost certainly in the world of pixels. The basic requirements for the front-end ICs include fast shaping times, low noise at low power, and excellent radiation hardness. The high electron mobility transistor (HEMT) based on heterojunctions between different III-V compounds, offers considerable hope of satisfying these requirements. For a recent review paper, see reference [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-HFETs is stable after irradiation by 100 Mrad gammas and 10^{15} n / cm^2 [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 ICs 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 1-dimensional information (microstrip detectors) and those being pixel-based and providing 2-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 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 ICs 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 learning curve 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 ≈ 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 ($\leq$ 120 Hz bunch crossing frequency). In the case of the superconducting RF design TESLA, the recently conceived CCDs with full area coverage and fast clear capability will permit clearing of the detector between each bunch crossing, if this is required.

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, recognising jets which are devoid of heavy flavours. 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 which focused on 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 summarise the time development of the various types of vertex and tracking detector with respect to some key parameters. Fig. 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 metres) 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 an even greater challenge for them, as we have seen.

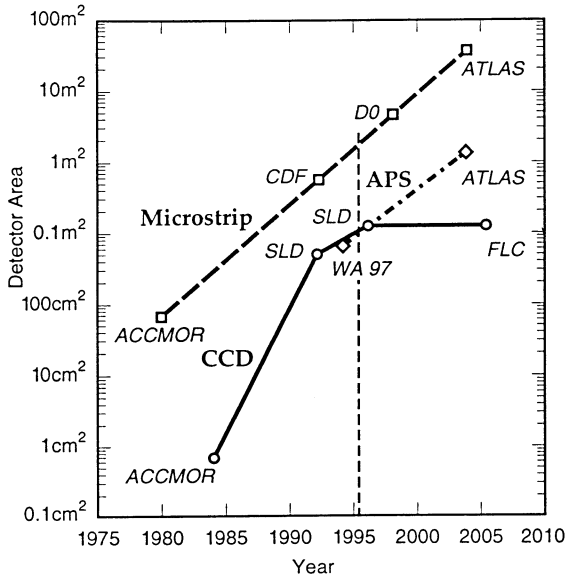

Fig. 64 Time development of area coverage of the leading-edge vertex and tracking detectors according to the main technologies (microstrips and CCDs). The APS is expected to enter the realm of vertex detectors in the LHC environment.

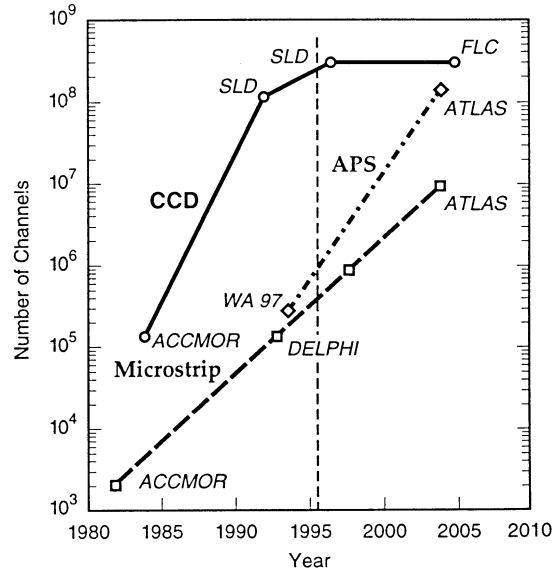

Fig. 65 As fig. 64, but showing the number of channels in leading-edge detectors as function of time.

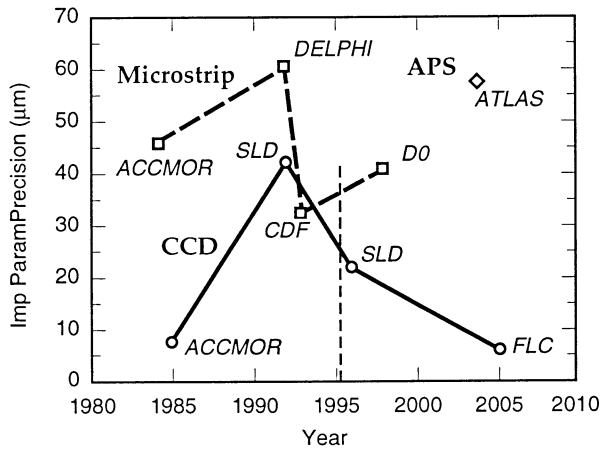

Fig. 66 As fig. 64, but showing the multiple scattering term in the impact parameter precision as function of time.

Fig. 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.

Fig. 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 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 a important to remember the very strong inter-disciplinary 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.

There is one silicon detector type which I have not discussed in these lectures, namely the silicon drift detector. This very interesting device was invented eleven years ago [102] by Gatti and Rehak, and continues to be the subject of intense developments in high multiplicity tracking and other applications [103]. My reason for considering it unlikely to have a role as a vertex detector stems from two intrinsic limitations, namely the coarse 2-particle resolution in the drift direction, and the extreme sensitivity to inhomogenous irradiation. Both of these features make it very difficult to envisage using these detectors close to the interaction point either in collider or fixed target applications. Nevertheless, they continue to be of great interest for a variety of applications as particle and X-ray detectors. The addition of channel stops (as in CCDs, to limit lateral diffusion) and interlaced anodes for improved transverse resolution, are particularly interesting recent developments.

Regarding the specific application to vertex detectors, there is an on-going 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 CCDs for vertex detectors in 1980). The established community of experts on silicon radiation detectors was generally extremely sceptical. 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.

Acknowledgements

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 acknowledgements 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 problems. Finally I would like to thank my wife Joan for her great support, and patience regarding the many lost week-ends that went into the preparation of the lectures and this written report.

Abbreviations

APS	active pixel sensor
BCO	Beam crossover
CCD	charge-coupled device
CCE	charge collection efficiency
CDS	correlated double sampling
C-HFET	complementary high-mobility field-effect transistor
CM	centre-of-mass
CTE	charge transfer efficiency
CTI	charge transfer ineffiency
CVD	chemical vapour deposition method of crystal growth
DEPJFET	*p*-channel FET on a high ohmic substrate
DLTS	deep level transient spectroscopy
EELS	electron energy loss spectroscopy
HDTV	high definition television
HEMT	high electron mobility transistor
HFET	high-mobility field effect transistor
I	interstitial silicon atom
I register	imaging register of CCD
LC	future e^+e^- Linear Collider (eg JLC, NLC, TESLA, CLIC)
LEC	liquid encapsulation method of crystal growth
MIP	minimum-ionizing particle
NIEL	non-ionising energy loss
PKA	primary knock-on atom
R register	readout register of CCD
SCT	silicon central tracker (eg for LHC experiments)
Si-A	oxygen/vacancy defect in bulk silicon
Si-E	phosphorus/vacancy defect in bulk silicon
SIPOS	semi-insulating polycrystalline silicon
SR	synchrotron radiation
V	vacancy in silicon crystal

Symbols

c	speed of light
C_d	node-substrate capacitance
C_g	gate-source capacitance of first stage output transistor
C_{ox}	capacitance per unit area across gate oxide
C_n	detector node capacitance (total)
D	carrier diffusion constant
D_n	electron diffusion constant
D_p	hole diffusion constant
e	exponentiation constant 2.718...
e^+, e^-	positron, electron
e^-	electron charge (as a unit)
$e_n(f)$	RMS noise per $Hz^{\frac{1}{2}}$ at frequency f
e_{nT}	RMS noise following correlated double sampling
E	energy transfer to an atomic electron **OR** energy level within silicon crystal
E_a	acceptor energy level
E_c	energy level of bottom of conduction band
E_d	donor energy level
E_f	Fermi energy level
E_g	the band gap energy $E_c - E_\mathrm{v}$
E_i	intrinsic energy level
E_I	effective ionisation threshold energy for holes or electrons
$E_{\max}$	maximum kinematically permitted energy transfer to an atomic electron
$E_{\min}$	Minimum energy transfer to a bound atomic electron
E_t	energy level at top of conduction band
E_{tr}	energy level of a bulk trap
E_v	energy level of top of valence band
E_γ	photon energy
E_0	energy level of liberated electron (surface or vacuum level)
$\mathcal{E}$	electric field strength
$\mathcal{E}_{\max}$	maximum field strength (at *pn* junction)
f_c	fraction of oxide charge trapped at SiO_2 surface
$f_D(E)$	Fermi-Dirac distribution function
F_j	fill factor for trapping under phase j
$g(E)$	density of states in silicon
g_m	transistor transconductance
G	amplifier voltage gain

$\mathscr{H}$	magnetic field strength
$I\phi$	CCD imaging register gates
k	Boltzmann's constant
ℓ	detector thickness
L_D	diffusion length for minority carriers
$\mathscr{L}$	luminosity of high energy collider
m_e	electron mass
n	electron concentration in conduction band
n_i	electron concentration in intrinsic material
N_a	acceptor concentration
N_a^-	concentration of ionised acceptors
N_c	effective density of states in conduction band
N_d	donor concentration
N_d^+	concentration of ionised donors
N_e	charge packet size (number of electrons)
$N(E)$	density of states
N_{eff}	effective dopant concentration in material under bias (depleted)
N_F	number of phases per pixel
N_s	signal charge density
N_{tr}	density of bulk traps
N_{v}	effective density of states in valence band
p	hole concentration in valence band
p_i	hole concentration in intrinsic material
q_e	charge on electron
Q_g	charge generated in SiO_2 layer
Q_s	signal charge
Q_{tr}	charge trapped at SiO_2 surface
$r_s(t)$	radius of charge distribution expanding by space-charge self-repulsion
R	radius of detector barrel in a collider tracking or vertex detector **OR** radial co-ordinate (polar co-ordinates)
$R\phi$	CCD readout register gates
$S_F(f)$	filter sampling function (frequency domain)
$S_F(t)$	filter sampling function (time domain)
t_d	carrier drift time in silicon
t_{ox}	thickness of SiO_2 layer
T	temperature
v_n	electron drift velocity **OR** electron thermal velocity in field-free material
v_p	hole drift velocity

V_B	applied bias voltage
V_c	channel voltage (of buried channel CCD)
V_g	gate voltage
V_{ss}	substrate voltage
W	mean energy for electron-hole pair creation in the material
x_n	depletion thickness in *n*-type material
x_p	depletion thickness in *p*-type material
X_n	entropy factor for bulk trapping
Z	atomic number
Z^0	neutral vector boson
α	leakage current damage constant **OR** track angle to detector normal
β	v / *c* where v is the speed of the particle **OR** current gain of a bipolar transistor
ΔV_{FB}	flat-band voltage shift induced by interface charge
ε_s	permittivity of silicon
ε_r	relative permittivity or dielectric constant of silicon
ε_0	permittivity of space
γ	energy-mass ratio of projectile
μ	carrier mobility
μ_n	electron mobility
μ_p	hole mobility
μ_n^H	Hall mobility for electrons
μ_p^H	Hall mobility for holes
ϕ	electric potential **OR** azimuthal angle (polar co-ordinates)
ϕ_B	p / p^+ barrier potential difference
ϕ_i	'intrinsic' *pn* potential barrier for an unbiased junction
ϕ_n	electric potential in n-type region
ϕ_p	electric potential in p-type region
$\phi_1, \phi_2, \ldots \phi_N$	separate gates of N-phase CCD register
ρ	resistivity
ρ_i	resistivity of intrinsic material
σ_n	cross-section for electron capture by a bulk trap
σ_R	Rutherford cross-section
θ_p^L	Lorentz angle for electrons
θ_p^L	Lorentz angle for holes
τ_c	trapping time constant
τ_e	emission time from trap
τ_g	dwell time of signal under a gate

τ_m	minority carrier lifetime
τ_r	clock pulse rise/fall time

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 Instr and Methods*, **A350** (1994) 368
4. P Lechner and L Strüder, *Nucl Instr and Methods*, **A354** (1995) 464
5. H Bichsel, *Rev Mod Physics*, **60** (1988) 663
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** (1951) 829
9. J B A England et al, *Nucl Instr and Methods*, **185** (1981) 43
10. J Kemmer, *Nucl Instr and Methods*, **169** (1980) 499
11. R Hofmann et al, *Nucl Instr and Methods*, **225** (1984) 601
12. J T Walker et al, *Nucl Instr and Methods*, **226** (1984) 200
13. P Holl et al, *IEEE Trans Nucl Sci*, **NS-36** (1988) 251
14. P Allport et al, *Nucl Instr and Methods*, **A310** (1991) 155
15. N Bachetta et al, *Nucl Instr and Methods,* **A342** (1994) 39
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** (1990) 1153
18. T Ohsugi et al, *Nucl Instr and Methods*, **A342** (1994) 22
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 Instr and Methods*, A236 (1993) 183
23. P Weilhammer, *Nucl Instr and Methods*, **A342** (1994) 1
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 Instr and Methods*, **214** (1983) 253
27. B Hyams et al, *Nucl Instr and Methods*, **205** (1983) 99
28. L Hubbeling et al, *Nucl Instr and Methods*, **A310** (1991) 197

29. H Tajima et al, *Nucl Instr and Methods*, **A310** (1991) 504

30. H Hanai et al, *Nucl Instr and Methods*, **A314** (1992) 455

31. A S Schwarz, *Physics Reports*, **238** (1994) 1

32. P Seller et al, *IEEE Trans Nucl Sci*, **NS-35** (1988) 176

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

34. E Beauville et al, *Nucl Instr and Methods*, **A288** (1990) 157

35. D E Dorfan, *Nucl Instr and Methods*, **A342** (1994) 143

36. Z Y Chang and W Sansen, *Nucl Instr and Methods*, **A305** (1991) 553

37. S Gadomski et al, *Nucl Instr and Methods*, **A320** (1992) 217

38. C Da Via et al, *Nucl Instr and Methods*, **A344** (1994) 199

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** (1970) 587

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

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

45. M J Howes and D V Morgan (ed), *Charge-Coupled Devices and Systems, 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 Instr and Methods*, **A253** (1987) 386

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

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, *Proc SPIE*, **2172** (1994) 38

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

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

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

56. D J Burt, *Nucl Instr and Methods*, **A305** (1991) 564

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

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

59. D Burt, Private communication

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

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

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 U of Waterloo Pixel Device Conference* (1993)

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

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

68. S Holland, *Nucl Instr and Methods*, **A275** (1989) 537

69. C J Kenney et al, *Nucl Instr and Methods*, **A342** (1994) 59

70. S Parker, Private communication

71 P Middelkamp, *Paper contributed to the 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

72. G Cesura et al, *Paper contributed to the 7th 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 Instr and Methods*, **A288** (1990) 54

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

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

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

79. W Füssel, *Paper contributed to the 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

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

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

82. H J Ziock et al, *Nucl Instr and Methods*, **A342** (1994) 96

83. E Fretwurst et al, *Nucl Instr and Methods*, **A342** (1994) 119

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

85. 3 Papers contributed by G Lutz, Zheng Li and S Watts to the *7th 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 Instr and Methods*, **A335** (1993) 580

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

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

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

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

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

93. M Schieber, *Paper contributed to the 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

94. E Bauser, *Paper contributed to the 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

95. R E Colbeth et al, *Proc SPIE* **1071** (1989) 108

96. A D Holland et al, *Nucl Instr and Methods*, **A346** (1994) 366

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 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

100. G Bertuccio et al, *Paper contributed to the 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

101. W Karpinski, *Paper contributed to the 7th European Symposium on Semiconductor Detectors (1995)* (to be published)

102. E Gatti and P Rehak, Nucl Instr and Methods, **A225** (1984) 608

103. 5 papers contributed by A Castoldi, U Faschingbauer, P Rehak, M Sampietro and L Strüder to the *7th European Symposium on Semiconductor Detectors (1995)* (to be published)