A review of design considerations for the sensor matrix in semiconductor pixel detectors for tracking in particle physics experiments

Sally Seidel*

New Mexico Center for Particle Physics, Department of Physics and Astronomy, University of New Mexico, 800 Yale Blvd NE, Albuquerque, NM 87131 USA

Received 9 December 2000; accepted 26 January 2001

Abstract

Methods have been developed to improve the reliability of silicon sensors, in particular for pixel detectors, and their resistance to radiation damage, as it is encountered in tracking detectors in particle physics experiments. The choice of wafer material, the processing techniques, and the sensor layout are discussed. Alternative semiconductor substrates and variations on the planar hybrid design are mentioned. © 2001 Published by Elsevier Science B.V.

PACS: ■■■■

Keywords: ■■■■

1. Introduction

The principal focus of this paper is the design of the silicon sensor part of a pixel detector. Originally, the included material was part of a full-day course on active pixel detectors. The other lectures treated the electronic readout chips, the hybrid interconnection technologies, and applications.

The development of pixel sensors is an extension to two dimensions of the silicon microstrip sensor technology, many of the features of which are described in Refs. [1,2]. This two-dimensional approach requires innovation in interconnections and electronics signal processing not described here. A silicon pixel sensor is defined here to be the sensing element of a hybridized detector, including a lightly doped substrate (usually n-type), one of whose surfaces is in contact with highly doped silicon of the opposite type (correspondingly, p-type), thereby forming a junction. The opposite side of the silicon wafer is in direct contact with highly doped silicon of the same type as the bulk. The highly doped silicon will be referred to here as “the implants”, although in fact it can be introduced through implantation or diffusion.

The implants on both sides of the device can be electrically contacted. When a reverse bias voltage $V_B$ is placed across them, a region in the bulk silicon is depleted of free charge carriers. The width $W$ of the depletion region in the n-type bulk
The desire for fine granularity makes silicon detectors a natural choice for tracking; however, while the very small feature size available in silicon devices provides low capacitance, low noise, consequently good signal-to-noise ratio, and low occupancy per channel (which reduces event buffering requirements), the radiation damage, which increases capacitance and creates charge traps, must be addressed in the design. Pixels’ small feature size and typically harsher radiation environment have placed constraints upon pixel design beyond those required for strip sensors; these are a subject central to this paper. Specifically, pixel sensor design and development have borrowed what was useful from silicon strip sensor design while focusing on the following issues: (1) engineering for robustness of radiation-damaged sensors designed with proven technologies; (2) maximizing the radiation hardness available through new technologies; (3) minimizing the sensors’ capacitance and maximizing their signal collection; and (4) exploring new design concepts. Because so many aspects of silicon pixel sensor design are influenced by radiation hardness requirements, the first section of the paper briefly reviews the response of silicon to radiation. The first section is not intended to be a complete review of radiation damage effects, but is merely intended to provide foundational information upon which specific design choices described in subsequent sections are based.

2. Radiation damage in silicon

2.1. Introduction

Radiation damage is caused by the passage of particles through the sensor. The main source of charged particles is collisions at the interaction point, so their fluence is proportional to $r^{-2}$. The main source of neutrons is backsplash from the calorimeter, so their fluence depends on the apparatus shielding and design. Bulk and surface damage are induced by different mechanisms, so these are considered separately below. The symbol $\Phi$ is used here to represent fluence. An excellent
recent review of radiation damage effects in silicon may be found in Ref. [3].

2.2. Bulk damage

Particles passing through a silicon substrate can cause dislocations in the lattice that alter the band structure. Following the collision, the displaced atom (or Primary Knock-on Atom, PKA) becomes a silicon interstitial and leaves a vacancy. The combination of vacancy and interstitial atom is known as a Frenkel Pair. In silicon, approximately 25 eV are required to displace the PKA [4]. The semiconductor bulk damage model postulates that the recoiling PKA strikes neighboring lattice atoms, and if its energy is greater than about 2 keV, its action will lead to the formation of clustered damage sites of typical volume $10^{-19}$ cm$^3$ [5]. Interstitial atoms and vacancies that escape a cluster and migrate through the lattice are generally trapped at the impurity atoms and form point defects. The subsequent evolution of the clusters and/or point defects is thought to produce certain macroscopic effects that are described below.

The damage done by radiation to silicon depends upon the type and energy of the radiation. The bulk damage is generally thought to depend exclusively on the non-ionizing energy loss ("NIEL") of the particle. This fact, which has been demonstrated to be the case over 14 orders of magnitude in particle energy, is called the NIEL hypothesis. (Some deviation may be apparent in the case of oxygenated silicon substrates; see Section 5.2.3 below.) It is consequently possible to scale the damage caused by different particle species at various energies by the NIEL, or by an equivalent scale factor known as the displacement damage function. The displacement damage function, which may be calculated by combining the individual reaction cross-section, the energy distribution of recoils produced by that reaction, and information about the partition between ionizing and non-ionizing energy loss of the recoils, and then summing over all reaction channels available to the initial particle at its energy, is shown in Fig. 1 (from Ref. [6]) as a function of particle species and energy. The portion of the spectrum below 190 eV is due to neutron capture and is not expected to be significant for LHC and future Tevatron experiments.

To facilitate comparisons between experiments and radiation sources, fluences are usually expressed in terms of the equivalent damage done by 1 MeV neutrons; in this paper the symbol $\langle n \rangle$ represents the 1 MeV neutron equivalent. Pions cause the worst damage to silicon in nuclear and particle physics experiments through $\Delta$-resonance production in the pion–nucleus interaction.

2.3. Surface damage

Bulk silicon naturally develops a layer of silicon dioxide, SiO$_2$. Bulk damage to the oxide has a negligible effect on its electrical properties because oxides, intrinsically quite disordered by their production process, contain a large number of defects even when unirradiated. In oxides, the most significant damage is caused by ionizing radiation, which generates bound charge in the oxide layer and at the interface between the silicon and the silicon dioxide. Because electrons have significantly higher mobility than holes in SiO$_2$, ionization-induced electrons rapidly diffuse out of the oxide, leaving behind a relatively permanent and immobile population of holes. The oxide charge has been observed [7] to saturate after about 100 krad at a value of about $3 \times 10^{12}$ cm$^{-2}$ in devices with detector-quality oxide. The explanation for this
is thought to be the limited number of permanent trap sites available in the oxide. No saturation of bulk effects has been observed up to fluences of a few times $10^{13} \langle n \rangle \text{cm}^{-2}$ [8].

In general the macroscopic effects of bulk damage are harder to control and more lethal [9–11] to sensors than are the effects of surface damage; they have consequently received more attention.

### 2.4. Macroscopic effects of radiation damage in semiconductors

#### 2.4.1. Introduction

Radiation damage to the bulk of the sensor consists in defects in the crystal lattice. Such defects have associated energy levels in the middle region of the forbidden energy band gap. The defect levels act as generation-recombination centers for positive and negative charge carriers, leading to increase in diode dark current, signal loss by temporary trapping, change in the effective dopant concentration, and increased resistivity of the undepleted part of the diode. Each of these effects is described below.

#### 2.4.2. Leakage current

Empirically

$$J(\Phi) = \alpha \Phi + J_{\text{intrinsic}}$$

where $J$ and $J_{\text{intrinsic}}$ are volume leakage current densities, $\Phi$ is fluence, and $\alpha$ is the current-related damage constant which will be described further below. Current $I_{\text{leakage}}$ increases in response to the development of generation-recombination centers in the band gap. It causes stochastic noise $\text{ENC}$ in the pixel’s amplifier such that

$$\text{ENC} \propto \sqrt{I_{\text{leakage}} \times \tau_{\text{shaping}}}$$

where $\tau_{\text{shaping}}$ is shaping time. If uncontrolled, heat associated with this leakage current can lead to thermal runaway.

The leakage current, which depends on temperature through the damage constant $\alpha$, is observed to change after the irradiation is over through a process called annealing. The relationship between $\alpha$, the temperature $T$ at which the irradiation occurs, and time $t$ can be parameterized as [12]

$$\alpha(T, t) = \alpha_1 e^{-t/\tau_1(T)} + \alpha_0 - \beta \ln(\theta(T)t/t_0)$$

where $t_0$ is the reference time associated with the duration of the irradiation, $\tau_1$ is the characteristic time associated with the annealing, and $\alpha_0, \alpha_1,$ and $\beta$ are annealing functions given in Table 1. The parameter $\theta(T)$ is defined by

$$\theta(T) = \exp\left(\frac{E_I}{k_B T} \left[ \frac{1}{T_R} - \frac{1}{T} \right] \right).$$

In this equation, $k_B$ is Boltzmann’s constant, $T_R$ is the reference temperature to which the measurement is normalized, and $E_I$ is the activation energy. A complete description of the physical processes behind annealing does not yet exist. It is expected to involve multiple interactions between defects and defect complexes, or the dispersal of complexes into point defects, each of which may be activated or deactivated at different temperatures. A useful table of important defects in silicon, and their properties, may be found in Ref. [2]. The empirical formula above for $\alpha$ fits well to data from a variety of processes and irradiation levels, as may be seen from Fig. 2.

#### 2.4.3. Dopant concentration

The effective dopant concentration, $N_{\text{eff}}$, of the substrate reflects the combination of ionized shallow levels and charged deep levels that is present. The effect of radiation is thought to be associated with the removal of shallow levels by creation of defect complexes and introduction of deep donors and acceptors. $N_{\text{eff}}$ has been shown to vary with fluence $\Phi$ over time $t$ for temperature $T$. 

---

**Table 1**

Parameters associated with current annealing at temperature $T_A = 60^\circ\text{C}$ (from Ref. [12])

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>$\times 10^{-17}$ A/cm</td>
<td>1.01 ± 0.38</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>Minutes</td>
<td>93 ± 24</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>$\times 10^{-17}$ A/cm</td>
<td>5.03 ± 0.09</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$\times 10^{-18}$ A/cm</td>
<td>3.34 ± 0.26</td>
</tr>
<tr>
<td>$t_0$</td>
<td>Minutes</td>
<td>1</td>
</tr>
</tbody>
</table>
according to the expression [13]

$$N_{\text{eff}}(\Phi) = N_{\text{eff}0} + N_C + N_a(\Phi, t, T) + N_Y.$$  

Here

$$N_C \equiv N_{C0}(1 - e^{-c\Phi}) + g_C \Phi$$

is known as the stable damage coefficient because it does not depend upon time; $N_a$, the short-term beneficial annealing coefficient, may be parameterized as a sum of exponentials

$$N_a = \Phi \sum_i g_{ai} e^{-t_i / \tau_{ai}(T)}.$$  

Experiments performed at room temperature [14] found this component to be insignificant after 2 days; elevated temperature studies [15] found only one exponential component to be detectable after 5 min.

The $N_Y$ term is the “reverse annealing” or “anti-annealing” coefficient. Formerly parameterized as $g_Y \Phi (1 - e^{-t/\tau_Y})$, it has now been shown [16] to be a first-order effect in defect concentration and is better expressed as

$$N_Y \equiv g_Y \Phi \left(1 - \frac{1}{1 + t/\tau_Y}\right).$$  

Here $\tau_Y$ is the time constant given empirically [17] by

$$\tau_Y = 9140 e^{-0.152 T},$$

where $T$ is temperature in Celsius degrees. This term has been the subject of considerable research because of the property that it can attain values significantly larger than the pre-irradiation dopant density as $t \to \infty$. The parameter $N_{\text{eff}0}$ represents the dopant concentration in the unirradiated substrate, $N_{C0}$ and $c$ are parameters associated with partial donor removal, $g_C$ is the stable acceptor parameter, and $g_Y$ is the anti-annealing coefficient. Table 2 summarizes values from a recent fit [18] for each of the annealing parameters. Fig. 3 illustrates the effect of each of the three annealing terms on the effective dopant concentration; after a period of time on the order of months has elapsed since irradiation, the dopant concentration of an irradiated sensor can be several times what it was both prior to irradiation and immediately after the conclusion of the irradiation. The fluence-dependent change in dopant concentration has

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Activation energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_A$</td>
<td>$(1.92 \pm 0.05) \times 10^{-2}\text{cm}^{-1}$</td>
<td>$1.09 \pm 0.09$</td>
</tr>
<tr>
<td>$g_Y$</td>
<td>$(5.16 \pm 0.09) \times 10^{-2}\text{cm}^{-1}$</td>
<td>$1.31 \pm 0.04$</td>
</tr>
<tr>
<td>$g_C$</td>
<td>$(1.49 \pm 0.03) \times 10^{-2}\text{cm}^{-1}$</td>
<td>—</td>
</tr>
<tr>
<td>$N_{C0}$</td>
<td>$(0.60-0.90) \times N_{\text{eff}0}$</td>
<td>—</td>
</tr>
<tr>
<td>$c$</td>
<td>$(1-3) \times 10^{-13}\text{cm}^{-2}$</td>
<td>—</td>
</tr>
</tbody>
</table>

Fig. 2. Values of $z$ as a function of annealing time at 60°C for diodes. The leakage current was measured at room temperature and normalized to 20°C. The legend indicates the neutron fluence and the manufacturers. Reprinted from Ref. [12] with permission from Elsevier Science.

Fig. 3. An example of the annealing behavior of the radiation-induced change in the effective doping concentration, $\Delta N_{\text{eff}} = N_{\text{eff}} - N_{\text{eff}0}$. The sample was irradiated with a neutron fluence of $1.4 \times 10^{13}\text{cm}^{-2}$ and annealed at a temperature of 60°C. Reprinted from Ref. [18] with permission from Elsevier Science.
significant impact on the behavior of the sensor’s depletion voltage. This connection will be discussed in Section 3.1.

2.4.4. Annealing

“Annealing” is the term used above for the change in both the effective dopant concentration (equivalently, depletion voltage) and the leakage current with time after the irradiation process has stopped. This process occurs in both p- and n-type substrates and is independent of material type (i.e., float zone, Czochralski, or epitaxial silicon) and inversion status (see Section 3.7). Table 3, taken from Ref. [12], illustrates the universality of the annealing parameter $\alpha$.

There is neither universal agreement among experimenters about whether the changes in voltage and current are due to the same microscopic process, nor about exactly what that process is. One opinion holds that the effects are due to deep acceptor creation and possibly donor removal (see, for example, Ref. [14]). Some investigators ascribe them to donor compensation by deep acceptors only [19]. The effort to associate the macroscopic changes in voltage and current with specific defects is a very active field of inquiry and uses a variety of spectroscopic methods. For an introduction to some of these inquiries, see Refs. [20–22]. While there has not yet been an unambiguous connection demonstrated between the presence of a specific defect and the observation of a specific change to the electrical character of a silicon sensor, recent results in Deep Level Transient Spectroscopy and Thermally Stimulated Current measurements support the conjecture that reverse annealing comes from the rearrangement of interstitial defects.

2.4.5. Charge trapping

Trapping occurs when crystal defects produce local energy states within the band gap. A trap’s average capture time increases exponentially with its depth and varies inversely with the capture cross-section. Defects with multiple energy levels can act simultaneously as traps for electrons and holes, in general with different associated trapping times. In systems for which the electron and hole capture probabilities differ, a positional (depth) dependence of the signal amplitude arises. The average time during which a signal charge is trapped in a semiconductor is given by

$$\tau = e^{(E_d - E_i)/k_B T} / \sigma v_{thermal} n_i$$

where $E_d - E_i$ is the difference between the defect and intrinsic energy levels, $k_B$ is Boltzmann’s constant, $T$ is temperature, $\sigma$ is the capture cross-section, $v_{thermal}$ is the thermal velocity of the charge carriers, and $n_i$ is the intrinsic carrier concentration. The relation between trap (defect) concentrations and fluence is given in Section 2.4.3.

Table 3

Measured values of $\alpha$ for a variety of materials. The oxygen and carbon concentrations are both given in units of $10^{16}$ cm$^{-3}$. The units of $\alpha$ are $10^{-13}$/A/cm. Details of the technologies used for manufacturing the diodes may be found in Ref. [12].

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Producer crystal</th>
<th>Producer diode</th>
<th>Guard ring</th>
<th>$\rho$ (k$\Omega$ cm)</th>
<th>[O]</th>
<th>[C]</th>
<th>$\alpha_{(80 \text{ min},60^\circ C)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-FZ</td>
<td>Wacker</td>
<td>MPI</td>
<td>Yes</td>
<td>2.7</td>
<td>&lt;5</td>
<td>&lt;0.5</td>
<td>3.99 ± 0.14</td>
</tr>
<tr>
<td>n-FZ</td>
<td>Wacker</td>
<td>ELMA</td>
<td>Yes</td>
<td>10–20</td>
<td>&lt;5</td>
<td>&lt;0.5</td>
<td>4.01 ± 0.04</td>
</tr>
<tr>
<td>n-FZ</td>
<td>Wacker</td>
<td>ITE</td>
<td>Yes</td>
<td>4.0</td>
<td>&lt;0.02</td>
<td>&lt;3</td>
<td>3.87 ± 0.07</td>
</tr>
<tr>
<td>n-FZ</td>
<td>Wacker</td>
<td>ITE</td>
<td>Yes</td>
<td>0.42</td>
<td>&lt;10</td>
<td>&lt;1.0</td>
<td>4.02 ± 0.11</td>
</tr>
<tr>
<td>n-FZ</td>
<td>Topsil</td>
<td>Sintef</td>
<td>Yes</td>
<td>6.6</td>
<td>&lt;5</td>
<td>&lt;0.5</td>
<td>4.14 ± 0.06</td>
</tr>
<tr>
<td>n-FZ</td>
<td>ITME</td>
<td>ITE</td>
<td>Yes</td>
<td>0.78</td>
<td>17</td>
<td>&lt;2</td>
<td>3.79 ± 0.08</td>
</tr>
<tr>
<td>n-FZ</td>
<td>ITME</td>
<td>ITE</td>
<td>Yes</td>
<td>0.11</td>
<td>&lt;10</td>
<td>2</td>
<td>3.61 ± 0.11</td>
</tr>
<tr>
<td>n-Cz</td>
<td>Polovodice</td>
<td>H</td>
<td>No</td>
<td>0.13</td>
<td>&lt;10</td>
<td>2</td>
<td>3.93 ± 0.13</td>
</tr>
<tr>
<td>p-EPI</td>
<td>ITME</td>
<td>DIOTEC</td>
<td>No</td>
<td>0.4</td>
<td>4–20</td>
<td>1–2</td>
<td>4.41</td>
</tr>
<tr>
<td>p-EPI</td>
<td>ITME</td>
<td>DIOTEC</td>
<td>No</td>
<td>1.6</td>
<td>3–20</td>
<td>1–2</td>
<td>3.92 ± 0.19</td>
</tr>
<tr>
<td>p-EPI</td>
<td>ITME</td>
<td>DIOTEC</td>
<td>No</td>
<td>3.9</td>
<td>4–60</td>
<td>1–2</td>
<td>4.06 ± 0.40</td>
</tr>
</tbody>
</table>
Trapping has implications both for signal loss and detector noise (see Section 3.6).

2.4.6. Conductivity of the undepleted bulk

Measurements [23] of the resistivity of the undepleted bulk of silicon devices show that it increases by more than a factor of 10 (from about 35 kΩ cm to about 400 kΩ cm) during an irradiation to \(10^{13} \langle n \rangle \text{ cm}^{-2}\) (see Fig. 4, which concerns n-type float zone material). This effect has been interpreted [24] as an indication of the relative position of the Fermi level \(E_F\) of the damaged silicon and the silicon intrinsic energy level \(E_i\), which are related to the resistivity \(\rho\) through

\[
\frac{1}{\rho} = qn_i(\mu_n\varepsilon(E_F-E_i)/k_BT) + \mu_p\varepsilon(E_i-E_V)/k_BT,
\]

where \(q\) is the magnitude of the carrier charge, \(\mu_i\) is carrier mobility for type \(i\), \(n_i\) is the intrinsic carrier concentration, \(k_B\) is Boltzmann’s constant, and \(T\) is temperature. Ref. [24] emphasizes that the fact that radiation-induced defects are deep rather than shallow influences the probability of defect ionization and leads to the more complicated expression for resistivity given above rather than the simpler correspondence between \(\rho\) and the voltage-to-current ratio.

3. Consequences of radiation damage for the operation of silicon sensors

3.1. Depletion voltage

Section 2.4.3 introduced the relationship between fluence, \(\Phi\), and effective dopant concentration, \(N_{\text{eff}}\). The depletion voltage of the sensor, \(V_{\text{dep}}(\Phi)\), is related to these through the electrical resistivity, \(\rho\), such that

\[
V_{\text{dep}}(\Phi) = \frac{w^2}{2\varepsilon\rho(\Phi)}
\]

for

\[
\rho(\Phi) = \frac{1}{q\mu N_{\text{eff}}(\Phi)}.
\]

Here \(w\) is sensor thickness, \(\varepsilon\) is electrical permittivity, \(\mu\) is carrier mobility, and \(q\) is electric charge.

If one combines these relations with those in Section 2.4.3, taking care with signs, one finds that when n-type silicon is subjected to radiation, it initially decreases its \(N_{\text{eff}}\) until it becomes quasi-intrinsic, then undergoes an apparent change of type from n to p (this is called type inversion), and subsequently increases its \(N_{\text{eff}}\), and consequently its \(V_{\text{dep}}\), without limit. In the case of a sensor that is initially p-type, the unlimited increase of \(N_{\text{eff}}\) and \(V_{\text{dep}}\) begins immediately with irradiation, and no type inversion occurs. Fig. 5 shows the behavior of \(|N_{\text{eff}}|\) and \(V_{\text{dep}}\) as a function of fluence.

The relationship between \(V_{\text{dep}}\) and fluence means that a detector must be operated partially

Fig. 4. The neutron-induced resistivity change, in n/cm², in the electrically neutral bulk of a high-resistivity silicon sample. Reprinted from Ref. [23] with permission from Elsevier Science.

Fig. 5. The depletion voltage and magnitude of the effective dopant concentration of bulk silicon as a function of fluence, as measured immediately after irradiation. Reprinted from Ref. [14] with permission from Elsevier Science.
depleted once the depletion voltage exceeds the breakdown voltage. Operation in this mode requires attention to several issues. First, in the depleted region, signal collection on the junction side is rapid: the n-side (electron) signal is collected in about 8 ns. The p-side (hole) signal is collected in about 21 ns due to the fact that hole mobility is 2.6 times lower than electron mobility. In a partially depleted sensor, the ohmic side signal (which must propagate through undepleted bulk) is diffused and shows a relatively longer collection time. Secondly, whereas in a fully depleted sensor, one expects the amount of charge collected to be directly proportional to the width of the depleted region, the fraction of charge collected by a partially depleted sensor is considerably less than the fraction of the sensor’s width that is depleted [24]. A half-depleted sensor, for example, will measure only a quarter of the charge of a fully depleted one, when stimulated by identical penetrating ionizing particles. This is because only half as much charge is generated in the depletion region, and half of this charge is unobserved due to induction of charge of the opposite sign in the undepleted region [2].

The undepleted region of a partially depleted sensor demonstrates an interesting effect [25] with respect to definition of the electric field at the sensor cut edge—after type inversion, the high resistivity of the undepleted bulk (see Section 2.4.6 above) along the cut edge of the sensor suppresses current there and consequently suppresses otherwise expected breakdown. Fig. 6 illustrates the effect of the resistive undepleted bulk.

Fig. 6. The distributions of the space charge region, undepleted region, and resistive region of bulk silicon in a single-sided structured p$^+$–n sensor, (a) before type inversion, (b) after type inversion without charge generation, and (c) with charge generation in the cut region. Reprinted from Ref. [25] with permission from Elsevier Science.
3.2. Power

Both depletion voltage and volume leakage current are proportional to the fluence $\Phi$ received by irradiated silicon sensors. Consequently the power dissipated in the devices is proportional to $\Phi^2$. This fact has implications for the cooling requirements. The two-dimensional nature of pixel arrays makes cooling them mechanically more challenging than is typically the case for silicon strip sensors; for a discussion of approaches to cooling pixel sensors, see Ref. [26].

3.3. Implant isolation

Section 2.3 mentioned that the silicon dioxide and the interface between it and the bulk silicon develop a layer of fixed charge. This charge, which is present to some degree even prior to irradiation, is normally positive. The presence of this layer induces an inversion layer of the opposite charge (called an accumulation layer in the case of electrons) which remains permanently attracted to it from the bulk. The accumulation layer can compromise the isolation of implants on the n-side of a pixel device unless special isolation features are included. Ref. [27] reports the decrease in resistance by almost 2 orders of magnitude between adjacent strips on the p-side of a strip sensor, as a function of fluence in the range from zero to about $10^{14} \langle n \rangle$ cm$^{-2}$. Fig. 7, from Ref. [28], shows an even more striking result in which the inter-strip resistance of n-on-n strip sensors is seen to decrease by 3 orders of magnitude, from 10 G$\Omega$ to about 20 M$\Omega$, independent of fluence, for fluences in the range $(0.8–8.3) \times 10^{13} \langle n \rangle$ cm$^{-2}$. Section 4.2 describes design features that can be used to maintain implant isolation.

3.4. Capacitance

The capacitance of a silicon sensor is a sensitive parameter in the design because it directly affects both noise and cross-coupling. The total capacitance presented by a pixel to the front-end electronics includes contributions [29] from the backplane ($10–20$ fF for a 300 $\mu$m thick sensor), the inter-pixel capacitance (approximately 100 fF for a typical design), the bump pad, and the preamplifier input transistor. The total capacitance affects the signal-to-noise ratio ($S/N$) through the relation [30]

$$S/N \approx \frac{Q_{signal}}{\sum_i Q_{noise}^i + V_{noise}^i / C_{total} \sum_i V_{noise}^i}$$

and the ratio, $C_{inter-pixel}/C_{total}$, affects the cross-coupling between channels.

The inter-pixel capacitance dominates the backplane capacitance by a factor of 4–10. Both types of capacitance increase with irradiation [31]. The increased $C_{inter-pixel}$ is thought to be due to the build-up of the accumulation layer: electric field lines in the silicon bulk can terminate on that layer in addition to terminating on the implants themselves—this increases the effective width of the implants and, consequently, the geometrical capacitance. Inter-pixel capacitance of n-type implants in n-type bulk (with p-stop isolation, see Section 4.2) changes by about 10–20% after a fluence of $8 \times 10^{14} \langle n \rangle$ cm$^{-2}$ for a variety of geometries. It can be minimized by appropriate choice of isolation technology and implant dimensions. It can, for example, be parameterized as a...
function of the ratio of width to pitch, \( w/p \), and the size of the unimplanted gaps between charge-collection electrodes on the sensor. The capacitance of silicon sensors is well known to depend upon the frequency of the stimulus once the sensors have been irradiated (see Fig. 8, which is taken from Ref. [32]); attention must consequently be paid by the experimenter to what is the appropriate frequency for a given component or application. Ref. [33] explains the connection between this frequency dependence and the presence of deep levels in the band gap.

The exploitation of large capacitive coupling between pixel cells is being examined by the TESLA collaboration as a way to improve resolution [34]. Noting that the expected resolution for analog devices is directly proportional to pitch, the collaboration seeks to overcome the minimum pitch now achievable for electronics by interleaving read out pixels with ones that are not read out in a manner analogous to that used in the past with strip sensors.

Two groups have recently looked for correlations between strip sensor capacitance and crystal orientation [35,36]. No significant difference in absolute inter-strip or total capacitance was found for signals at the high frequencies most relevant to collider experiments. Some differences in settling times and voltage dependence are reported although these must still be separated from effects associated with processing choices.

### 3.5. Microdischarge

Microdischarge [37], also called microplasma, is a reversible increase in channel noise that grows rapidly and spreads to neighboring channels as bias voltage is increased. This effect has been observed to be associated both with pixel design and with radiation dose and is thought to be due to a tunnelling or avalanche breakdown caused by high fields. It can occur along the junction implant edge inside the silicon bulk or in association with the oxide charge at the silicon–SiO\(_2\) interface. The probability that a sensor will experience microdischarge increases with bias voltage, oxide charge density, and potential difference between an implant and its external readout electronics. Fig. 9, taken from Ref. [38], shows one of the problems that microdischarge poses for silicon sensors: a steep increase in leakage current at relatively low bias voltage. A related problem is...
noise amplitude, which, during microdischarge, increases with bias voltage as well. As the dominant cause of microdischarge is thought to be a MOS effect associated with the implant and its conductive pad, the problem can be reduced if the implant is designed to extend at least 2 μm beyond its conductor in all directions. Additional options for reducing microdischarge are discussed in Refs. [38,39].

3.6. Signal and noise

The signal production by a semiconductor is associated with ionization of the material by through-going charged particles. A review of the subject, including corrections for statistical fluctuations, may be found in Ref. [40]. Fig. 10 shows the rate of energy loss, \(\frac{dE}{dx}\), in silicon, as a function of the kinetic energy of a through-going pion. In semiconductors, only part of the energy lost by the particle subsequently creates electron–hole pairs, as phonon production may not be neglected. The average energy necessary to create a pair in silicon is 3.6 eV; as a minimum ionizing particle loses 1.66 MeV/g/cm² in silicon, its average energy loss along the <111> orientation of the lattice is 390 eV/μm. This translates to production of 108 pairs/μm or \(3.2 \times 10^4\) pairs along a 300 μm track. There is no multiplication of charge in a silicon sensor.

The noise of a silicon detector assembly is typically dominated by the electronics contribution rather than the sensor. Refs. [41,42] review issues associated with the electronics. To minimize the sensor noise, one minimizes the leakage current (hence shot noise) and the capacitive load on the amplifier (see Section 3.4 above). Leakage current is minimized in semiconductors with large band gaps and few mid-gap (defect) states. As will be described further in Section 4.1.1, the leakage current may be further suppressed by operation of the sensor in a low-temperature environment.

It is apparent that both the signal and the noise performance of a sensor are directly related to defect density through trapping and generation. It is because detector grade Group IV semiconductors such as Ge and Si have defect densities that are orders of magnitude lower than typical compound semiconductors that they are frequently chosen as substrates for devices requiring good signal-to-noise ratio.

Radiation-induced lattice defects have been shown to act as trap sites that lead to the loss of up to 15% [43] of the signal in silicon strip sensors after fluences comparable to that received during an LHC lifetime \((2 \times 10^{14} \text{ p/cm}^2)\) and collection times appropriate to LHC electronics (see Fig. 11). Fig. 12 shows trapping probabilities measured separately for electrons and holes in highly irradiated silicon diodes. As irradiation proceeds, the electron signal is found to degrade faster than...
the hole signal [44]. The charge collection efficiency is independent of annealing time [45]. For 300 µm thick sensors irradiated with 24 GeV/c protons to a fluence of $10^{14}$ cm$^{-2}$, a charge collection efficiency of 90% was maintained with 160 V bias voltage and collection time 20 ns. Those irradiated with 300 MeV/c protons to a fluence of $6 \times 10^{14}$ cm$^{-2}$ maintained a 40% efficiency [46]. The presence of trap sites also changes the shape of the electric field distribution in the sensor and consequently alters somewhat the shape of signals to be read out.

3.7. Bulk-type inversion

As was mentioned in Section 3.1 and illustrated in Fig. 5, at a fluence of about $10^{12}$ cm$^{-2}$, the substrate of an initially n-type sensor begins to operate as p-type; this is known as type inversion. An early hypothesis about the process was that the functional form of the effective dopant concentration, $N_{\text{eff}}$, reflected donor removal (by the attachment of radiation-induced vacancies to phosphorus atoms) and shallow acceptor creation [47]. However, subsequent DLTS analysis has indicated that considerably less phosphorus removal occurs than is required, and furthermore, no candidate acceptor state has yet been identified.

A new hypothesis has consequently been proposed that the introduction of deep level acceptor states causes n-type silicon to become effectively p-type when placed under bias [48].
Inversion manifests itself as an abrupt movement of the main junction from the p-side of the sensor to the n-side. Figs. 13 and 14, taken from Ref. [49], are direct evidence of this effect. On each of them, the vertical axis shows the measured pulseheight induced by an infrared LED directed at the segmented (p) and the back (n) sides of some strip sensors fabricated on n-type substrate. The horizontal axis indicates bias voltage. The former figure concerns the sensors prior to irradiation; the latter, after type inversion. One sees that prior to inversion, the signal may be read from the p-side at low voltage, indicating that the junction is there, while the n-side signal does not develop until the voltage is high enough to cause the depletion region to extend to the back side. After inversion, the junction has moved to the n-side, and the situation is reversed: the n-side signal is present at low bias voltages, while the p-side signal appears only after full depletion. Inversion is not a problem for the operation of the sensor as long as the design anticipates it. Design features that are typically required for post-inversion n-side operation (for example channel isolation implants and guard rings) are described in the sections below.

Several investigators have reported a related phenomenon: the development of a second junction which appears on the p-side after inversion. The second junction, which has been observed directly [50,51] and reproduced in simulation [52], is associated with an n-type inversion layer of thickness approximately 15 \( \mu \)m in the effectively p-type bulk. Ref. [52] points out that if more than one defect type is present (for example, a dominant acceptor level and an additional donor level), trapped charge is not distributed uniformly across the bulk: “holes... are more efficiently trapped close to the p+ junction side: such a region is therefore less inverted than the deeper bulk... Therefore, within a certain range of fluences, a depletion layer can simultaneously originate from doping discontinuities at both ends of the detector”. Ref. [53] links the junction to a specific donor-like level below mid-gap and an acceptor-like one above. Fig. 15 is a measurement of TCT current in which the double-peaked structure indicates the presence of both junctions.

4. Techniques for increasing the radiation robustness of proven sensor designs

4.1. Operating temperature minimization

4.1.1. Suppression of annealing

Section 2.4.1 mentioned that radiation damage manifests itself both in increased leakage current and in a change to the effective dopant concentration. The leakage current increase can be con-
trolled if the thermal environment can be controlled; several separate effects are involved. First, the leakage current of any semiconductor device can be thermally suppressed, regardless of whether damage has occurred. The relation between leakage current and temperature is well described by the expression

\[ I_{\text{leakage}} \propto T^2 e^{-E_{\text{gap}}/2k_B T} \]

where \( T \) is Kelvin temperature, \( E_{\text{gap}} \) is the effective band gap [54] (1.12 eV for silicon), and \( k_B \) is Boltzmann’s constant. Fig. 16, taken from Ref. [55], shows the excellent agreement between this formula and the measured temperature dependence of the leakage current in silicon sensors for radiation levels of 0, 0.1, and 2 Mrad from 12 GeV protons. The implication of thermal control for operation of highly irradiated pixel sensors at forward bias (thereby trading high space charge for leakage current) is being investigated [56].

As was indicated in Section 2.4.4, there is a relationship between leakage current and annealing, and this may be associated with mobility of defects in the damaged silicon. Mobility, whose dependence upon fluence has not yet been unambiguously established, appears to saturate with fluence at about 1000 cm² V/s for electrons and 450 cm² V/s for holes at room temperature [57]. The mobility can be thermally suppressed [57,58], leading to a thermal suppression of the...
component of leakage current associated with damage. The effective dopant concentration $N_{\text{eff}}$ of an irradiated silicon sensor is given in Section 2.4.3 by the sum of three terms, each of which corresponds to a type of annealing with its own time constant. Because of the temperature dependence of the annealing coefficients, $a$ and $N_{\text{Y}}$, can be completely suppressed by reduction of the sensor temperature, a fact demonstrated in Fig. 17 (taken from Ref. [59]). To minimize the sensor's depletion voltage, the sensor should be operated at a temperature high enough to activate beneficial annealing but low enough to suppress reverse annealing. The temperature range $-10–0^\circ\text{C}$ is appropriate to achieve this for LHC lifetimes and fluences.

4.1.2. The “Lazarus Effect”

The ability of a highly irradiated silicon sensor to recover its essential pre-irradiation operating characteristics when run at cryogenic temperatures has been demonstrated [60]. A 300 $\mu$m thick silicon strip sensor was irradiated to $2.23 \times 10^{15} \langle n \rangle$ cm$^{-2}$ and then biased to 250 V, it showed no signal at 195 K. With its temperature lowered to 77 K, it recovered a fast, 13,000$e^-$ signal (see Fig. 18). No further improvement was observed when the temperature was lowered to 4.2 K. The temperature range $-10–0^\circ\text{C}$ is appropriate to achieve this for LHC lifetimes and fluences.

The model that has been offered for this “Lazarus Effect” is based on the fact that at cryogenic temperatures, the low thermal energy of the silicon lattice reduces the detrapping rate of carriers, so a large fraction of carriers can be trapped at cryogenic temperatures.
the deep levels is constantly filled and hence deactivated. A small inefficiency which persists in the sensor at low bias voltages even at 4.2 K, where defects are expected to be frozen out, may be explained by the presence of the hexvacancy complex, $V_6$ [61]. The charge collection efficiency is maximized at 130 K and shows some time dependence [62].

4.2. Control of the accumulation layer

In Section 3.3, it was mentioned that as radiation fluence increases, bound positive surface charge develops at the silicon–oxide interface, and that this fixed charge attracts electrons that can ultimately short the n-implants. The p-stop [63] and p-spray [64] techniques have been developed to maintain implant isolation.

p-stops are implanted p$^+$ channels between neighboring n-implants. They have been implemented in some pixel designs after successful application in microstrip sensors. Fig. 19, from Ref. [65], illustrates some of the patterns (ordinary, common, atoll, and combined) that have been examined. Optimization of a p-stop design requires consideration of the effect of these p-implants upon the pixel charge collection efficiency and capacitance as well as on the n-implant isolation. Fig. 20, also from Ref. [65], shows that pixels utilizing the ordinary p-stop typically show the highest charge collection efficiency, followed by those with the combined design. The reduced
efficiency of the atoll design is thought to follow from the fact that the atoll p-stop does not segment all of the accumulation layer. Charge deposited between atolls can be coupled away by the accumulation layer, which is conductive, and this leads to inefficiency. The combined design, on the other hand, has the lowest capacitance (hence, noise) [29,63]. It is clear that decisions about p-stop design must be made in the context of the full detector design including information about other contributors to capacitance (for example, in the electronics).

A p-spray layer is a shallow p-type implant that is applied across the full wafer without mask prior to any other processing. The dopant concentration of the implant is matched to the well-known value at which surface charge saturates, $3 \times 10^{12} \text{ cm}^{-2}$. Subsequent n-implantation then over-compensates the p-spray layer wherever needed. p-spray devices use the growth of the accumulation layer to their advantage: the accumulation layer compensates the dopant acceptors, so that as radiation proceeds, the p-spray layer becomes increasingly closer to intrinsic. The lateral electrical field between implants consequently decreases with fluence, increasing the breakdown voltage.

Fig. 21, from Ref. [64], shows the results of a technology simulation of a p-stop and a p-spray device for various levels of oxide charge density (hence, ionizing radiation). One sees that in the case of the p-spray device, but not in the case of the p-stop, the electric field magnitude decreases (and hence the breakdown voltage increases) with fluence. This improvement of radiation hardness...
with irradiation has been demonstrated with the ATLAS prototypes [66].

Control of the accumulation layer is also a geometrical issue. Studies of surface effects show a clear relationship between the generated surface current of irradiated pixels and the size of the gap between implants [67]. Fig. 22 compares the current after 11 kGy for pixels with large and small gaps. The exponential rise in leakage current in the large gap devices is ascribed to the confinement of accumulation layer electrons in the gap as a consequence of the adjacent depletion.
zones coalescing before the flat-band voltage is achieved. In addition to improving the radiation resistance of the sensor, p-spray has the benefit that since no mask is required for its application, the cost of implant isolation is lowered, and neighboring n-type structures can be placed closer.

4.3. Control of electrical breakdown

Guard rings, typically implanted and metallized structures that surround the active areas of silicon sensors, serve two purposes. (1) As the depletion region develops from the junction, it expands toward the cut edge which, due to its mechanical damage, is conductive. The guard ring serves to drop the voltage from the interior of the sensor face to the cut edge in a controlled manner, so that the voltage gradient across the edge is zero. (2) The accumulation layer induced by the presence of fixed charge at the oxide deforms the depletion region, generating high field points at risk of electrical breakdown. The oxide layer is unstable and sensitive to changes in the environment; consequently, the behavior of the accumulation layer is variable. The guard ring serves to stabilize the oxide and to shape the depletion region. To meet these requirements, typical guard ring structures include metal lines atop the oxide plus one or more ring-shaped p–n junctions that surround the diode array but are not contacted or biased directly.

Fig. 23, from Ref. [68], is an example guard ring layout. (A variety of designs have been proven to be successful; this example is selected merely to illustrate several concepts.) The rings in this design are a serial connection of p-channel MOSFETs, in which the gate only covers half of the distance between the drain and source of the sensor. The gates are connected to the sources rather than the drains. The guard ring is operated by biasing the n-side and grounding the active area and inner guard. As bias voltage rises, the depletion region expands. When it contacts, or “punches through to” the first floating ring, that ring charges up. Increasing the voltage further biases all of the rings sequentially. Each ring’s potential depends upon the bulk dopant concentration and oxide charge (hence on the fluence) as well as on the separation between rings. When charged, the rings distribute the diode’s field beyond the diode’s perimeter, thus reducing $\nabla V$ at every surface point. Fig. 24, from Ref. [69], represents the electrostatic potential at
the sensor surface, as a function of distance from the sensor center, for measurements and simulations of a guard ring structure with a variety of options in surface charge density. One clearly sees that the multi-ring structure steps the voltage by a controlled amount at the location of each ring.

In a particular set of related simulations and designs, the breakdown voltage associated with the guard ring structure was found to increase with distance of the outermost guard to the scribe line up to a distance of 150 \( \mu \text{m} \), and then saturate [70]. The breakdown voltage is maximized for the narrowest achievable inter-ring gaps. The innermost guard must be connected to guarantee that the field is correctly shaped (see Fig. 2) [12]. It is worth emphasizing that n-side guard rings are inactive prior to inversion, and p-side rings, after. Guard ring designs that tolerate 500 V after a fluence of \( 10^{14} \text{cm}^{-2} \) [25] and those that tolerate 900–1000 V before [70] have been demonstrated.

A study of p\(^+\)-on-n devices has also examined the use of an n\(^+\) implanted region along the edge to inhibit avalanche breakdown [71]. It concluded that the n\(^+\) implant should be no closer than 150 \( \mu \text{m} \) to the p\(^+\) and that the p\(^+\) implant should be no closer to the edge than 400 \( \mu \text{m} \). Drive in diffusion steps lead in general to smoother junctions and lower electric fields [72].

### 4.4. Crystal orientation

It has generally been supposed that the \( \langle 100 \rangle \) crystal orientation is more radiation hard than the \( \langle 111 \rangle \) one because its oxide charge density is lower. The \( \langle 111 \rangle \) has nonetheless traditionally been used for silicon sensors because in surface barrier detectors and p–n diodes, the higher oxide charge inhibits breakdown. Furthermore, the \( \langle 111 \rangle \) orientation reduces signal dispersion due to channeling in spectrometry.

Fig. 24. The measured and simulated potential distributions along the surface of a particular multi-guard ring structure. The three plots show the results for different oxide charge densities and substrate doping concentrations. The details of the design may be found in Ref. [69], from which this figure is reprinted with permission from Elsevier Science.
It has been reported [73] that sensors fabricated from epitaxial silicon with the $<111>$ crystal orientation are more radiation hard than are those with $<100>$. The devices about which the report was made have resistivity 630 $\Omega$/cm, considerably less than the resistivity traditionally used for detectors. While it is reasonable to expect that silicon wafers with different growing conditions, including orientation, may have different responses to radiation, the full connection between radiation hardness, crystal orientation, and low resistivity of these devices has, however, not yet been fully sorted out.

4.5. The p-type substrate option

Most silicon sensors fabricated up to this time have used n-type substrates. While p- and n-type silicon substrates have rather similar radiation damage constants [74,75], n-type material has the advantage that its majority carriers, the electrons, have three times higher mobility than holes [54]; the depletion voltage is correspondingly lower. The principal benefit of beginning with p-type substrate is the fact that inversion does not occur. The junction then remains on the n-side of the sensor throughout its lifetime, simplifying quality assurance of the devices and some aspects of the design.

5. Initiatives to improve radiation hardness for future detectors

5.1. Introduction

At present the majority of silicon sensors used in particle physics applications have resulted from planar processing of high-resistivity n-type float zone silicon wafers. While the vast majority have utilized 4-in. wafers, no difference has been observed in those produced on wafers of diameter 6 in. [76]. Several interesting routes are being explored to increase the radiation hardness of detector-quality devices: (1) reduced substrate resistivity, (2) epitaxial or Czochralski substrates, (3) alternatives to planar processing, (4) oxygenation of the silicon, and (5) other semiconductors. This section reports on the status of each of these.

5.2. Wafer fabrication and processing options

5.2.1. Substrate resistivity

The usual classification system identifies detectors of bulk resistivity $\rho$ around 5–10 k$\Omega$/cm as high resistivity, those with $\rho$ around 1 k$\Omega$/cm, medium resistivity, and those with $\rho < 500$ k$\Omega$/cm, low resistivity. While lower resistivity silicon has a higher pre-irradiation depletion voltage than does high, it also has a higher inversion fluence. Inversion fluences $\Phi_{\text{inversion}}$ for the resistivity range $1.5 \leq \rho \leq 20$ k$\Omega$/cm have been shown [77] to be well described by the equation, $\Phi_{\text{inversion}} = 18 \times N_{\text{eff}}$. A low starting resistivity reflects a high density of built-in donor defects.

The use of low-resistivity silicon merits exploration for several reasons [78,79] including the lower substrate cost and the fact that, for applications in which inversion is guaranteed not to occur, single-sided wafer processing, with its associated simplifications and cost reduction, may be used. Full activation, or punchthrough, of all rings in a multi-ring guard structure on such a device is achieved with lower voltage. Lastly, whereas leakage current grows with fluence, depletion voltage decreases with it prior to inversion; consequently power dissipation is balanced throughout the lifetime of a sensor that will be utilized only prior to inversion.

Several low-resistivity sensors have been fabricated, irradiated, and operated in exploratory studies. Fig. 25 shows the effective dopant density of one such 130 $\Omega$/cm demonstration sensor as a function of fluence $\Phi$. One sees that the device is uninvolved up to $\Phi = 9 \times 10^{14}$ $<n>$ $\text{cm}^{-2}$. Detector quality sensors are not yet available with this low resistivity.

Unfortunately, no absolute advantage in depletion voltage can be gained from low-resistivity silicon that has the standard amount of absorbed oxygen: the resistivity must be achieved with highly oxygenated wafers (see Section 5.2.3 below). Extrapolations from existing data (see Fig. 26) predict that after one LHC lifetime (10 years),
standard silicon wafers of all resistivities will require the same depletion voltage [46].

5.2.2. Epitaxial and Czochralski silicon

During production by the float zone method, a polycrystalline ingot is suspended in vacuum or an inert gas and heated to melting in a narrow region along its length. The position of the interface zone between the solid and liquid regions is then slowly moved through the material. Because the solubilities of some impurities are different in solid and liquid silicon, sweeping the liquid zone through the length of the ingot transports the impurities to the end of the ingot, which may be excised. Repeated sweeps leave a highly purified crystal.

The Czochralski method also uses the fact that a moving liquid zone purifies the silicon, but begins instead with a seed crystal lowered into molten silicon. As the seed is raised and rotated, oriented crystals grow upon it. Czochralski-grown ingots have a higher oxygen concentration than do float zone, because the molten silicon is in contact with the SiO2 crucible.

In the epitaxial process, one begins with a substrate (which may be silicon or a material with a similar lattice structure) and exposes it to an environment of free atoms. These deposit on it, preserving the substrate crystal’s aspect. The deposition process for silicon is most commonly chemical vapor deposition, or CVD. The growth rate for silicon is normally between 0.5 and 1 μm per minute.

Epitaxial silicon is known to have more as-grown defects, more crystal mismatch, and consequently larger strain fields and internal stress than float zone silicon [80]. Prior to irradiation, typical samples contain high (≥2 × 10^{12} cm^{-3}) deep level concentrations. It is hypothesized that as-grown deep levels can provide a sink for radiation-induced defects; recently, research has been undertaken to take advantage of this phenomenon [81].

Deep Level Transient Spectroscopy has been applied to samples of non-oxygenated epitaxial silicon to identify the deep levels present. The middle element of Fig. 27 shows the spectrum for an unirradiated epitaxial silicon sample. This sample was irradiated to a fluence of 1.75 × 10^{13} cm^{-2} per year at full luminosity). Reprinted from Ref. [46] with permission from Elsevier Science.

Fig. 26. The calculated depletion voltage as a function of LHC operational years for the first layer of the ATLAS SCT barrel (radius 30 cm, z = 0 cm, fluence 1.75 × 10^{13} cm^{-2} per year at full luminosity). Reprinted from Ref. [46] with permission from Elsevier Science.

The ability of the as-grown defects to act as sinks is limited by their density. For the samples mentioned above, saturation was observed after a fluence of 6 × 10^{15} protons cm^{-2}, at which point the DLTS trap spectrum for the sample was...
similar to that of float zone silicon. Increasing the as-grown defect density of epitaxial silicon requires increasing the growing time for the ingot. The concentration of its defects increases non-linearly with thickness [81].

In other respects epitaxial and float zone material have comparable qualities. Their reverse annealing constants are similar—one can see this in Fig. 28, which shows similar development of the effective dopant concentration, \( N_{\text{eff}} \), for control float zone samples and for several epitaxial samples. Epitaxial and float zone samples of similar initial resistivities have nearly the same inversion fluence [73].

Czochralski silicon can achieve oxygen concentrations up to \( 10^{18} \text{ cm}^{-3} \). While this high oxygenation may eventually prove valuable for radiation hardness (see Section 5.2.3), Czochralski silicon is
not yet available as detector quality wafers. Czochralski material has been used as a substrate for epitaxial deposition [82] with the intent that its oxygen diffuse into the epitaxial material.

5.2.3. Oxygenation of the substrate

It has been hoped for some time that one could improve the radiation tolerance of silicon by defect engineering, which is the deliberate addition of impurities to the silicon in order to form electrically active defects and thereby control the macroscopic behavior of the material. Significant effort has been applied to studies with oxygen and carbon.

Results available in late 1998 first showed that when oxygen is introduced to the silicon wafer above a specific threshold concentration, the silicon is found to be up to 3 times more radiation hard against charged hadrons [83]. The oxygen may be introduced to the silicon by jet injection at the ingot stage or by diffusion at high temperature after oxidation of the wafers. The exact value of the threshold, and optimized parameters for the oxygen’s introduction, are still under investigation, but there are indications that a diffusion of 16 h at 1150°C, such that \([O] = 1.5 \times 10^{17} \text{ cm}^{-3}\) in a 300 µm wafer, may be adequate.¹ Fig. 29 shows the reduction in full depletion voltage (equivalently, \(N_{\text{eff}}\)) as a function of proton fluence, observed for oxygenated wafers.

This discovery is accompanied by two interesting effects that have not yet been fully explained. The first is the fact that the improved radiation resistance applies to charged particles but not to neutrons. This apparent violation of the NIEL scaling hypothesis by the charged particles is receiving considerable attention. It is noted that more point defects are produced by charged particle irradiation than by neutral. A second unexpected consequence of oxygenation is its suppression of reverse annealing. Rather than remaining proportional to the fluence, as is the case for standard silicon, the reverse annealing component of the effective dopant concentration in oxygenated wafers saturates above a fluence of about \(2 \times 10^{14} \langle n \rangle \text{ cm}^{-2}\), leading to a reduction of \(N_{\text{eff}}\) by about a factor of 2. The reverse annealing time constant is, furthermore,

¹Typical high-grade, high-resistivity float zone silicon contains oxygen at a concentration of about \(10^{15} \text{ cm}^{-3}\) without special processing.
enhanced and appears to depend upon the oxygen concentration.

Several suggestions [84,85] have been offered to explain the beneficial effect of the oxygen. One proposes that the defect responsible for the formation of negative space charge in the bulk under bias may be the divacancy–oxygen complex, V$_2$–O. Increasing the concentration enhances the formation of the vacancy–oxygen complex, V–O, and so suppresses V$_2$–O. While correlations between microscopic defects and macroscopic damage parameters have been observed, the naive suppression model does not adequately account for volume current increase due to hadronic radiation. It has been proposed [86] that charge exchange between traps inside clusters may describe a significant portion of the current generation and space charge density associated with neutron irradiation.

The following facts have emerged about beneficial oxygenation. The oxygen must be substitutional; a silicon wafer prepared with a concentration of $2 \times 10^{17} \text{ cm}^{-3}$ interstitial oxygen atoms was shown to be no more radiation hard than normal silicon [87]. Epitaxial wafers with an oxygen concentration of $5 \times 10^{17} \text{ cm}^{-3}$ have demonstrated an inversion fluence two times higher than standard float zone wafers of the same initial resistivity [88]. Oxygen-rich Czochralski wafers show half the generation rate for reverse annealing as do normal Czochralski wafers, although other annealing parameters such as $z$ and $g_C$ are unchanged by oxygen [18]. Like oxygen, tin added to silicon has been shown to act as a vacancy trap [89]; the implications of this for radiation hardness are being explored [90]. Some investigators have also pointed out the potential of nitrogen doping [46]. Germanium introduced to silicon at concentration of $10^{19} \text{ cm}^{-3}$ has thus far not proved beneficial, possibly due to Ge-vacancy complex instability at room temperature [90]. The introduction of carbon into the wafer causes sensors to degrade with irradiation.

5.2.4. Alternatives to planar processing

Planar technology, which was originally invented for microelectronics processing, required adaptation [91] for use in the production of silicon sensors but is now the usual procedure. The planar process generally involves photolithographic structuring, chemical etching, doping, oxidation, deposition of insulating and conducting layers by chemical reaction, deposition of metals by evaporation or sputtering, thermal treatment, and passivation. A general discussion of the process may be found in Ref. [54]. An alternative process, known as mesa, has been applied to the production of $p^+–n–n^+$ diodes. The mesa process involves high-temperature diffusion in a normal atmosphere of boron and phosphor to form a progressive junction and an ohmic contact deep in the bulk. Mesa processing eliminates the oxidation and masking stages and produces devices which, lacking guard rings yet having junctions that extend to the device edge, typically show higher leakage currents. It was invented for single diode pads and is not available at this time for multi-diode arrays. It has, however, produced devices with improved radiation tolerance relative to that observed for comparable planar devices. It is under study in the hope that the essential features that improve radiation hardness may be discovered and transferred to other technologies.

A 1998 study [92] showed that mesa silicon, prepared with or without oxygenation, suppresses proton-induced change in effective dopant concentration by a factor of two relative to planar processed epitaxial or float zone material. A complementary study [93] using neutrons, however, showed no difference between mesa and planar diode full depletion voltages after a fluence of $5 \times 10^{13} \text{ cm}^{-2}$. Oxygenated mesa diodes also show a smaller change in leakage current in response to proton irradiation than do oxygenated planar devices [92]. One group [94] has reported an as-yet unexplained initial decrease in $N_{\text{eff}}$ in $p$-type mesa silicon for low proton fluxes. A very large increase in the oxygen concentration of silicon processed with mesa technology has been observed [73]; the relationship between the benefits that stem from this oxygenation and those associated with oxygenation of planar devices is under study.
5.3. Non-silicon substrates

Several initiatives are underway to identify semiconductors that, like silicon, have relatively large band gaps and so are expected to be radiation hard. The majority of work in this area has been applied to development of GaAs and diamond. Ref. [95], and references therein, provide a recent status report on GaAs. While it typically has a leakage current 10 times that of comparable quality silicon, other properties of GaAs have attracted significant attention to it. These include the fact that it has twice the density of silicon, four times better radiation length, the same pair production energy, and a carrier mobility that is 5 times greater than silicon's: this would imply that a 150 μm GaAs sensor could collect the same charge as a 300 μm silicon one. GaAs devices have been demonstrated to have signal-to-noise ratios of at least 30, and charge collection efficiencies greater than 95%, prior to irradiation. Fabrication by a non-standard technology has produced a “compensated GaAs” with approximately equal concentrations of donors and acceptors and a purity comparable to that obtainable with silicon. The high dopant concentration allows the sensor to collect charge without external bias. Unfortunately, GaAs has not proven to be as radiation hard as was initially hoped [97,98].

An excellent recent review of diamond detectors appears in Ref. [99]. The band gap in diamond is 5.5 eV, approximately five times larger than silicon’s. Consequently bulk currents in diamond are negligible (100 pA cm⁻² for 500 μm thick devices) and no depletion is necessary, so no diode structure is required. This large band gap leads to extreme radiation hardness: diamond sensors exposed to radiation showed no degradation after photon fluence up to 100 Mrad and 2 particle fluence up to 10¹⁵ cm⁻² [100]. After a 300 MeV/c pion fluence of 1.1 × 10¹⁵ cm⁻², the most probable signal decreased by less than 15%. Exposure to 24.2 GeV/c protons produced a measurable effect only after about 2 × 10¹⁵ cm⁻². At 0.75 × 10¹⁵ 1-MeV(\langle n\rangle) cm⁻², the mean value of the signal distribution decreased by about 15%, but the most probable value was unaffected [101]. Furthermore, diamond’s low dielectric constant of 5.6 leads to a relatively low sensor capacitance at the input to the readout electronics.

Diamond crystals generate 13,500 pairs along a 300 μm track, about a factor of two fewer than silicon. The important figure of merit for diamond is its charge collection distance (CCD), which is the average distance an electron and hole separate under the influence of the external electric field before they are trapped. CCD is related to charge collection efficiency (CCE) through the equation, CCD = CCE × thickness. Considerable effort has been devoted to increasing charge collection distance in diamond during the past 10 years, and the improvement has been significant; a typical CCD is now approximately 250 μm. Charge collection distance improves by 50–100% with irradiation up to saturation at 10 krad, through a process called pumping. The model for this proposes that charge traps are reversibly filled by radiation-induced defects, and hence deactivated. Fig. 30 shows the increase in CCD during exposure to a ⁹⁰Sr source. A diamond detector at the LHC would remain pumped throughout its life and would survive for 10 years at 7.5 cm from the interaction point.

A diamond strip sensor has been fabricated with 50 μm pitch. When operated with an analog preamplifier of shaping time 25 ns, it showed signal-to-noise ratio of 7 and position resolution of 18 μm. A 16 × 16 array of 150 μm square pixels
wire bonded to a fanout on a glass substrate and read out with a VA3 chip showed signal-to-noise ratio 27. Present diamond detector R&D is aiming for creation of larger devices (areas of $2 \times 4 \text{ cm}^2$) that have been achieved, increased CCD, lower noise electronics (one goal is a 30% reduction in the noise of LHC strip detector amplifiers), and an optimized metalization for bump bonding to conventional pixel electronics [100].

Another substrate that has received some attention is SiC [102]. Silicon carbide has a band gap three times larger than silicon’s (3.2 eV) and a comparable radiation length. Its leakage current is 1000 times lower than silicon’s, and its capacitance prior to irradiation depends neither on voltage nor frequency, indicating high purity. While its collection time for electrons is short, corresponding to an electron mobility greater than 22 cm$^2$/V/s, the mobility of its holes is low, approximately 3 cm$^2$/V/s. Studies are underway to characterize its radiation hardness fully.

### 6. Other directions

#### 6.1. Introduction

Several interesting silicon-based detectors have been developed in recent years in addition to those described in the preceding sections. This section discusses only two, monolithic pixels and “3D”.

#### 6.2. Monolithic pixel detectors

The subject of monolithic pixel detectors, devices that combine sensing and amplification properties in the same structure, is an extensive one reaching back to the mid-1980s. A review of early developments may be found in Ref. [103]. Only selected highlights will be mentioned here. The benefits of monolithic processing include the possibility of thinner devices (hence reduced multiple scattering), increased reliability of interconnection, lower capacitance, and perhaps, eventually, reduced cost. The principal disadvantage is simply that the sensor and the amplifier cannot be optimized separately.

Different investigators have taken somewhat different approaches to the problem. In 1992, a device with 300 $34 \times 125 \mu \text{m}^2$ pixels in 300 $\mu$ thick high-resistivity p-type silicon was demonstrated [104]. Fig. 31, taken from Ref. [105], illustrates the principle: an n-type phosphorus diffusion creates a junction. Sequential readout circuitry is contained in a two-dimensional array of n-wells surrounded by p$^+$-collection diodes. The n-wells serve as Faraday cages to isolate the collection field from the switching transients in the electronics and shape the field to direct the signal charge to the collection implants. The device showed gain uniformity of $\pm 2.3\%$ within a chip, spatial resolution of $2 \mu$m in the short direction and $5.6 \mu$m in the long, and better than 99.99% of the ionization charge gathered on the collection electrodes.

To address the issue of interference between the two active parts of the detector, a design was undertaken [105] using an isolated buried oxide in the SOI technology. Fig. 32 illustrates this concept. The n–p shield at the interface to the buried...
oxide was shown to be able to reduce coupling between the active layers by a factor of $10^4$ with little contribution to the junction capacitance.

In 1998 a vertical high voltage termination structure was proposed for the backside junction of silicon detectors that require double-sided processing [106]. It has been applied to a monolithic pixel detector and has increased yield. One version of it may be seen in Fig. 33. This robust one-mask structure is a deep vertical etch through the junction into the bulk, etched during processing and passivated with thermally grown oxide to prevent surface generation leakage current. As the etch can be extended all the way through the bulk, the detector can be turned on its side to provide a very deep depletion zone for stopping high energy X-rays or $\gamma$-rays.

A different approach to monolithic detectors was first proposed [107] in 1987 and subsequently built and tested [108]. Fig. 34, from Ref. [103], illustrates this DEPMOS (DEpleted P-channel MOS) concept. A standard MOS transistor is built on top of high-resistivity silicon bulk. The biasing of the MOS gate in such a way as to create an inversion layer at the oxide–semiconductor interface forms a transistor channel connecting two diodes. The conductivity of the channel may be directed by the gate voltage and the bulk potential, leading to a potential well for majority carriers below the transistor. The first amplification stage is in the device itself, as the majority carriers in the well induce charges of roughly the same amount in the channel, increasing the channel conductance and the transistor current.

6.3. “3D” detectors

An interesting recent development is the “3D” detector [109,110], illustrated in Fig. 35. These devices utilize standard silicon wafers with electrodes oriented such that they extend through the full substrate thickness (typically 300 $\mu$m). The small distance between p- and n-type electrodes implies a reduction in depletion voltage of these devices by a factor of about 10 relative to planar electrodes, leading to expectations of excellent radiation hardness. Development of the 3D design is made possible by advances in micro-machining that permit etching of deep, narrow, nearly vertical holes. The holes are coated with polysilicon which is then doped and heated to drive the dopants into

![Fig. 33. A simulated diode termination structure in n-type bulk using a one-mask vertical etch. Reprinted from Ref. [106] with permission. © 1998 IEEE.](image1)

![Fig. 34. The principle of the DEPMOS detector. Reprinted from Ref. [103] with permission.](image2)

![Fig. 35. The principle of the 3D detector, in which electrodes penetrate the substrate. Reprinted from Ref. [110] with permission. © 1999 IEEE.](image3)
the surrounding single-crystal silicon to form the junctions and ohmic contacts.

7. Conclusion

An introduction to silicon pixel sensors is provided, including information about design principles that increase their resistance to radiation damage. Recent developments in wafer fabrication and processing techniques which may improve the radiation hardness of future detectors are also included. Alternatives to silicon substrates and to the planar hybrid design are mentioned.

Acknowledgements

The author wishes to thank the organizers of the 1999 Nuclear Science Symposium and Medical Imaging Conference, in particular Gary Allen and Jeff Appel, for the opportunity to present this information in the Short Course on Pixel Detectors for Nuclear and Particle Physics. Sherwood Parker, Erik Heijne, and Walter Snoeys provided valuable suggestions. Veronica Mata-Bruni assisted with the figures. This work was supported in part by the US Department of Energy.

References