Introduction

Micropattern silicon detectors, and in particular silicon strip detectors, are a key component of essentially every modern high energy physics detector (with the exception of most neutrino detectors). Silicon strip detectors have been continuously improved since their first use in the mid 1980's, and since then they have played an ever larger role in experiments. One measure of this development is the area of silicon used in a detector, which has grown exponentially (see figure 1). The first experiments with silicon strip detectors used a handful of planes, each covering less than 10 square centimeters; the CMS tracker will employ more than 200 square meters of silicon strip sensors!

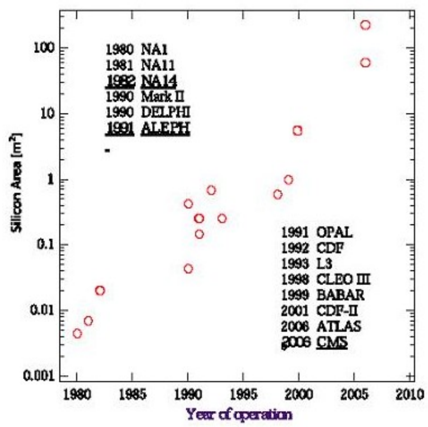

Figure 1: "Livingston plot" for silicon strip detectors… needs revision. Might also include a picture of an early surface barrier strip detector & one of the CMS endplate, or tracker under construction.

This paper reviews the development of silicon strip detectors. It includes a short summary of the relevant properties of silicon and the development of silicon detectors for use in nuclear physics. A summary is also presented of the development of strip detectors for use in fixed target experiments and subsequently in collider experiments. The basic radiation damage mechanisms in silicon sensors and the development of radiation tolerant readout electronics is reviewed. Finally, a short survey of the use of silicon strip detectors outside of high energy physics is presented, as is a review of other types of micropattern silicon devices.

Motivation for the development of SSD's

The development of silicon strip detectors in the early 1980's was motivated by the needs of fixed target charm experiments at CERN and Fermilab. In order to reduce combinatorial background, and to measure decay lifetimes, the proponents of these experiments wished to isolate charm decays from the production vertex. This required a measurement resolution of order 10 μm, significantly better than could be obtained either by wire chambers or by conventional bubble chambers.

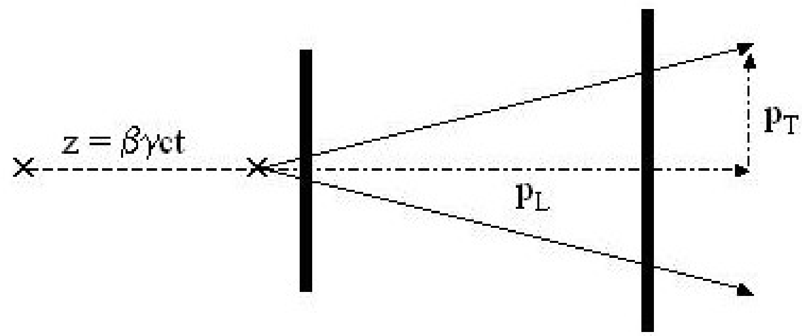

Figure 2: A simplified view of the decay of a heavy particle. The vertical lines in the figure represent measurement stations with position resolution σ_0. The significance of detachment, to first order, depends only on the lifetime of the heavy particle and on σ_0.

The position resolution required to separate the decay vertex of a heavy particle from its production vertex is directly related to the decay lifetime of the heavy particle, and, to first order, is independent of the momentum of the heavy particle. This fact can be illustrated by considering the symmetric two-body decay shown in Fig 1. The error on the determination of the transverse position of the decay vertex is a function of the position and slope errors for each of the decay tracks and of the distance between the first measurement station and the decay vertex. If the effect of multiple scattering can be neglected, and the lever arm between measuring stations is sufficiently large, the error on the transverse position of the decay vertex, σ_x, is approximately equal to the single point measurement error, σ_0. The relationship between σ_x and the error on the determination of the longitudinal position of the decay vertex, σ_Z, depends on the opening angle of the decay. Referring to Fig. 1, $\sigma_Z/\sigma_x = P_L/P_T$, where P_L and P_T are the longitudinal and transverse momenta of one of the daughter particles. In the limit that M, the mass of the heavy particle, is much greater than the mass of the daughters, P_T = ½ M. For the symmetric decay, P_L of each daughter is ½ of the momentum of the heavy particle, ½ $M\beta\gamma$. This implies that $\sigma_Z/\sigma_x = \beta\gamma$, or $\sigma_Z \approx \beta\gamma\sigma_0$. The heavy particle decay length is related to t, the lifetime in its rest frame by $z = \beta\gamma ct$, and the detachment significance $z/\sigma_Z \approx ct/\sigma_0$.

The theoretical expectation in the 1970's was that the lowest mass charm states should all have lifetimes of order 0.1 psec, corresponding to $c\tau \sim 30$ μm [M. K. Gaillard, B. W. Lee, and J. L. Rosner, Rev. Mod. Phys. 47, 277 (1975)]. By 1982, it was established that the neutral D lifetime is closer to 0.5 psec ($c\tau \sim 150$

μm), and the charged D lifetime is closer to 1.0 psec ($c\tau$ ~ 300 μm), [reference PDG], but the D_s lifetime was still believed to be ~0.2 psec ($c\tau$ ~ 60 μm). This provided the primary motivation for the development of tracking detectors with position resolution of 10 μm or better.

Silicon Fundamentals

Silicon was an attractive detector material for use in charm experiments for a number of reasons. The energy required to create free charge that can be collected by an external circuit is small, meaning that sensors can be thin and good energy resolution is achievable. Most importantly for the fixed target charm experiments, silicon has a high stopping power, and almost all of the free charge is created within a few microns of the path of a charged particle. This means that very good position resolution is achievable.

Any photon or charged particle with energy greater than the silicon bandgap of 1.12 eV can create mobile charge in silicon [reference Brown]. For energetic particles passing through silicon, an average of one electron-hole pair is created for every 3.62 eV of energy lost. The energy loss per unit length (dE/dx) depends on the particle energy and type. For a minimum ionizing particle, approximately 30 keV is lost per 100 μm of silicon, corresponding to ~8000 electron-hole pairs. If an electric field is present, these charge carriers will move and can induce a current in an external circuit. In order to sustain an electric field without generating a large standing current, however, it is necessary for the silicon to have very high resistivity. At or near room temperature, this can only be accomplished by fabricating the silicon detector as a diode junction.

One characteristic feature of a silicon diode is the existence at the junction of a very high resistance region that is essentially depleted of mobile charge. An electric field exists in this depletion region that resists the flow of current in one direction but not in the other, giving the diode its rectifying characteristic. This electric field also allows any mobile charge created in the depletion region to be collected by an external circuit[1]. If the diode is reverse biased, the depletion region (from which a signal can be extracted) becomes thicker. The depletion depth at a given bias voltage depends on the resistivity of the bulk silicon, which is a function of the density of charge carriers, and thus of the number of dopant sites per unit volume. Very lightly doped silicon has higher resistivity than heavily doped silicon, and the depletion depth of a diode junction formed in lightly doped silicon is larger for a given bias voltage than is the case for lower resistivity silicon.

[1] The full charge that is created can usually be collected, since both the electron and hole mobilities in silicon are high, and the carrier lifetimes are typically much longer than the time required for a charge carrier to traverse the full thickness of a detector.

A number of different types of silicon diode junctions can be made. One type is formed if n-type and p-type silicon are brought into contact with one another. Free electrons from the n-type silicon diffuse into the p-type silicon where they fill holes in the valance band. Near the junction, the n-type silicon is depleted of mobile electrons and the p-type material is depleted of mobile holes. The n-type material becomes positively charged near the junction, and the p-type material negatively charged, creating an electric field that opposes further polarization of the material. If highly doped silicon of one type contacts lightly doped silicon of the other type, the depletion depth is much greater on the lightly doped side than on the heavily doped side.

A diode junction can also be formed at a metal-silicon boundary. This type of junction is called a surface barrier, or Schottky diode. For n-type silicon, a diode is formed if the metal has a larger work function than silicon, which means that the top of the conduction band of the metal is lower energy than the Fermi level of the silicon (the energy level of the n-donor sites). In this case, mobile electrons will flow out of the silicon into the metal, charging the metal at the junction and depleting the n-type silicon of mobile electrons near the junction. As in the case of the n-p junction, equilibrium is reached when the electrostatic force caused by the buildup of electrons at the junction is large enough to stop further polarization of the silicon. The metal most commonly used to produce silicon Schottky diodes is gold.

An ohmic contact is formed at the silicon-metal boundary with a metal, such as aluminum, that has a smaller work function than n-type silicon. In this case the top of the metallic conduction band is higher energy than the Fermi level of the silicon. When the two materials are brought into contact, electrons flow from the metal into the silicon, charging the silicon at the contact and creating a region of silicon with a higher density of electrons in the conduction band than was the case before the junction was formed. This region of higher-than-normal mobile charge does not resist current flow in either direction, but rather provides a good ohmic contact.

The First Silicon Sensors

The first practical silicon detector was a gold surface barrier device constructed in 1959 at the Chalk River Laboratory in Ontario, Canada [ref. J. M. McKenzie and D. A. Bromley, Bull. Am. Phys. Soc. 4 (1959) 422]. Within months detectors were also made using silicon n-p junctions created by gaseous diffusion of phosphorus into p-bulk silicon. The first semiconductor detectors of any type had been made only about a year earlier using gold-germanium surface barriers.

The primary attraction of semiconductor detectors to the nuclear physicists who developed them was their superior energy resolution compared to the gas and scintillation detectors that were prevalent at the time. This superior energy resolution is made possible by the larger number of primary charge carriers created per unit of energy lost by the particle being detected. As stated above, an average of one electron-hole pair is created in silicon for every 3.62 eV of energy deposited. The rate of mobile charge creation is approximately the same in germanium. By contrast, approximately 26 eV is deposited in Argon, and 22 eV in Xenon, per ion pair created [ref. F. Sauli]. In NaI(Tl), the gold standard of scintillation crystals, more than 50 eV deposition is required per scintillation photon [R. Hofstadter and J. A. McIntyre, Phys. Rev. 80 (1950) 631]. For a phototube with a 20% quantum efficiency, this means 250 eV per photoelectron.

Need a few words here about why both silicon and germanium detectors were developed… or maybe not.

A number of groups were involved in the initial development of silicon detectors. In addition to the Chalk River group these included:

- Bell Telephone Laboratories and Brookhaven National Laboratories; W. L. Brown, P. F. Donovan, W. M. Gibson, and G. L. Miller.
- Oak Ridge National Laboratory; C. J. Borkowski and J. L. Blankenship.
- Hughes Aircraft Research Laboratories; J. W. Mayer and S. S. Friedland.

The development of semiconductor detectors during the 1960's was primarily aimed at making detectors with very thick depletion layers that could provide good energy resolution over a wide range of particle and photon energies. Two development paths were particularly important. The first path focused on changes to the detector design aimed at raising the maximum possible bias voltage (references required), as well as the use of very high resistivity, nearly pure crystal, which translated directly into a thicker depletion layer at a given bias voltage. The second development path involved the use of gold or lithium to compensate for the crystal impurities and create very high resistivity material.

Something here about the use of guard rings to allow high voltage operation. Also very high resistivity bulk material and passivation (SiO2?) to reduce leakage current.

Very thick (several mm) depletion layers were achieved by drifting lithium into p-type silicon [J. H. Elliott, NIM 12 (1961) 60]. Lithium is an electron donor and lithium ions are very small and mobile in silicon. "Lithium drifted" sensors are fabricated starting with p-bulk material. An n-p junction is created using lithium to create the n-type material. The device is then reverse biased in an oven and the temperature raised to approximately 150 °C. Lithium ions diffuse into the bulk silicon, compensating the p-type bulk so that it becomes nearly intrinsic, greatly increasing the depletion depth at a given bias voltage.

Lithium drifted silicon and germanium detectors are made and used to this day. The only real drawback to this method is that the sensor must be kept cold (typically in a liquid nitrogen dewar) for its entire lifetime or the lithium ions will diffuse out of the bulk material and the sensor will no longer fully deplete before reaching its breakdown voltage!

[KEY references:

Development history: “Development of the Semiconductor Radiation Detector” by J.M. McKenzie, NIM 162 (1979) 49-73

Theory of operation: “Introduction to Semiconductor Particle Detectors” by W.L. Brown, IRE Trans. Nucl. Sci. NS-8, #1 (1961) 2]

Development of SSD’s for Fixed Target Charm Experiments

The possibility of making position-sensitive silicon detectors was recognized very soon after the first detectors were made [NS-13 3 (1966) 208]. However, the nuclear physicists who developed silicon detectors had no need for extremely precise position measurements. The development of fine pitch silicon strip detectors came only with the advent of experiments designed to measure charm particle decays. In 1980, two papers were published reporting tests of position sensitive silicon detectors [Heijne, et al. NIM 178 (1980), 331 and Amendolia, et al. NIM 176 (1980) 457], each of which concluded that the technology was well suited to the needs of Fixed Target charm experiments. Both of these groups used small gold surface barrier devices with aluminum ohmic contacts.

The next, and single most important step in the development of fine pitch silicon strip detectors, was the adaptation to detector construction of integrated circuit fabrication procedures collectively referred to as “planar processing.” These include photolithography, which facilitates the manufacturing of micron-scale features, wafer passivation with silicon dioxide, and doping by diffusion and ion implantation. [reference J. Kemmer, “Fabrication of Low Noise Silicon Radiation Detectors by the Planar Process” NIM 169 (1980) 449-502 and J. Kemmer, “Improvement of Detector Fabrication by the Planar Process” NIM 226 (1984) 89-93.] Most of the groups interested in developing practical silicon strip detectors abandoned surface barrier devices in favor of planar processing soon after Kemmer’s initial success. Companies experienced in the manufacture of PIN photodiodes, which are made using similar processing techniques, collaborated with high energy physics groups in much of the subsequent development.

Of special note are the R&D efforts in the period 1981-1985 of the CERN NA11 and NA14 collaborations, and the Fermilab E653 collaboration [references]. The NA11 collaboration, which had developed the first planar devices, continued to fabricate some of their own detectors, and also worked with Enertec Corporation[2] (check name).

E653 collaborators from Nagoya University established a relationship with Hamamatsu in 1981. In 1982, Hamamatsu provided a 3 cm square 1-mm pitch PN junction photodiode (without aluminum electrodes) for testing, and in 1983 Hamamatsu produced 3 cm square fine pitch silicon strip devices suitable for use in E653. The radiation tolerance of these devices was studied in 1984 using the E653 800 GeV/c proton beam, and very fine pitch (12.5 micron) Hamamatsu sensors were used by E653 in 1987 [references].

Also in 1981, NA14 collaborators from Imperial College (London) established an R&D relationship with Centronic, a British PIN photodiode maker. Early in 1983, Colin Wilburn and Tony Lucas left Centronic (which had decided not to proceed with silicon strip detector development) to form Micron Semiconductor Ltd. Wilburn and Lucas secured the necessary bank loans on the basis of contracts from Imperial College and from E653 collaborators from Ohio State University. R&D proceeded with the university groups testing prototype devices and Micron working on fabrication improvements based on the test results. By 1984, Micron had succeeded in making strip detectors with an active area of 5 cm x 5 cm. The E653 and NA14 designs differed slightly, but both had 1000 p+ strips with 50 micron pitch. Every strip was read out and both ends of the sensor were used for readout, with odd number strips being wire bonded to a printed circuit board on one and even number strips wire bonded at the other end.

In 1984, collaborators from the University of California at Santa Barbara purchased NA14-type sensors from Micron for use at Fermilab in the Tagged Photon Laboratory (E691). The experiment collected data during a six month period in 1985. E691 was the first spectacularly successful experiment to use silicon strip detectors. Before E691, the total sample from all experiments of charm particle decays suitable for lifetime measurements was approximately xxx. E691 collected a sample of approximately 10,000 decays, including D^0, $D^\pm$ and D_s. The E691 lifetime measurements [J.C. Anjos, et al. PRL 58, 311 (1987) and PRL 58, 1818 (1987)], published less than two years after the end of data taking, were more precise than the previous world averages by more than a factor of two [reference, 1986 PDG, Phys. Lett. 170B (1986)]. The D_s lifetime, previously thought to be significantly shorter than the D^0 lifetime, was found to be approximately equal to the D^0 lifetime. The success of E691 would not have been possible without the silicon strip detectors developed for NA14, but it also depended on a number of other factors. These included:

1. The Fermilab tagged photon beam, whose high flux and good duty factor were made possible by the 800 GeV/c Tevatron.
2. A working multiparticle spectrometer.

---

[2] Enertec/Schlumberger became Eurisys Mesures by merger. It was later purchased by Canberra and is now known as Canberra Eurisys.

3. An inclusive trigger (based on transverse energy) and a high rate data acquisition system – the experiment wrote an unprecedented volume of data to tape (more than xx GB).
4. First use of a massively parallel computer “farm” based on commodity processors for event reconstruction (the Fermilab Advanced Computer Project).

The success of E691 established silicon strip detectors as a necessary component of the detector builders “kit” for every experiment interested in short lived particles.

The Role of ASIC’s

The first silicon strip detectors all were simple planar devices, typically consisting of a single silicon crystal held in a G10 frame. Signals were fanned out from the fine pitch strips to amplifiers, often held on printed circuit boards mounted on the edges of the frame. This scheme was not well suited to collider experiments in which the vertex detector is the innermost detector element of a number of concentric elements. Fortunately, by the early 1980’s the integrated circuit revolution was well enough advanced that custom “Application Specific Integrated Circuits” (ASIC’s), were becoming available. This allowed the miniaturization of silicon strip amplifiers and associated readout electronics, which could then be located right on the edge of the silicon strip detectors, inside most of the central detector.

Custom integrated circuits developed for silicon strip readout were first fabricated at universities. The first silicon strip readout ASIC was the “microplex” chip designed for use by MARKII at SLAC, and fabricated at Stanford University [reference: Walker, NIM 226…(1984)]. The design was implemented in the Stanford University 5 micron NMOS, single metal, single poly, process. At about the same time, an R&D program to develop a readout chip suitable for use at LEP was started by Buttler, et al. These devices were designed using CMOS and were fabricated at the Fraunhofer Institute in Duisburg. The design used correlated double sampling[3] to reduce the electronic noise associated with sensor leakage current. A 64-channel version of this chip (CAMEX64), fabricated in 3.5 micron CMOS, was used by ALEPH at LEP.

Two other early ASIC design efforts are of note. The first was a CMOS ASIC designed at Rutherford Lab, in the UK [references]. Both OPAL and DELPHI used versions of this chip. The circuit was designed for fine pitch silicon using staggered rows of input bond pads, and it used correlated double sampling.

[3] In correlated double sampling, charge is integrated onto one capacitor before a signal arrives. Charge is integrated onto a second capacitor while the signal is present. The output is the difference between the signals on the two capacitors. This technique eliminates noise due to sensor leakage current, and to “charge injection” associated with closing transistor switches.

Circuits were fabricated in the UK by Plessey, first using 5 micron CMOS, and later 3 micron CMOS. The second was the SVX, designed at LBL for use by CDF at the Tevatron proton-antiproton collider. The SVX design is notable for a number of reasons. It was the first readout chip designed for use in a high rate environment, and for that reason was the first to include pedestal subtraction and zero suppression in the readout. SVX was also the first silicon strip readout chip to be prototyped and produced through MOSIS, the Metal Oxide Semiconductor Implementation Service.

MOSIS facilitates the prototyping and low-volume production of custom integrated circuits. Designers send computer files describing their chip designs to MOSIS using the Internet. MOSIS groups a number of designs together onto one “multiproject wafer.” MOSIS contracts to a number of different vendors for mask generation and chip fabrication. MOSIS grew out of programs supported by DARPA (the Defense Advanced Research Projects Agency[4]) in the mid 1970’s[5]. MOSIS was established in 1981 and was designed primarily to provide fast turn-around IC prototyping for students at American universities. From 1981 to 1985, MOSIS was funded solely by DARPA. From 1985 to 1994, MOSIS received funding from the NSA (National Security Administration) and NSF (National Science Foundation) in addition to DARPA. During this time, MOSIS began accepting designs from commercial customers. In 1994, MOSIS became a non-profit commercial enterprise. It no longer receives any government funds, but rather supports itself by the fees it charges commercial customers. MOSIS provides its services to customers world wide[6]. Two similar services exist in Europe, Circuits Multi-Projects (CMP)[7] and the Europractice IC Service[8].

Collider Detectors

In the late 1980’s and early 1990’s, silicon strip vertex detectors were added to essentially all of the high energy physics collider detectors in operation or being

---

[4] ARPA, the Advanced Research Projects Agency, was established in 1958 by the United States Department of Defense, “for the direction or performance of such advanced projects in the field of research and development as the Secretary of Defense shall, from time to time, designate by individual project or category.” (http://www.darpa.mil/arpa_darpa.html) In 1972, the name was changed to DARPA, the Defense Advanced Research Projects Agency. In 1993, the name was changed back to ARPA (dropping Defense). Finally, in 1996, the name was changed back to DARPA (restoring Defense). Probably the most well known ARPA/DARPA project is the development of protocols for computer networking. The resulting network, ARPANET, evolved into today’s Internet. (http://www.isoc.org/internet/history/brief.shtml).

[5] The first multiproject VLSI prototyping service (called MPC Implementation System) was run by Lynn Conway at Xerox PARC. The first VLSI graphical layout editors were developed and made available to university users at no charge in the context of this and subsequent DARPA-supported programs.

[6] http://www.mosis.org/

[7] http://cmp.imag.fr/

[8] http://www.europractice.imec.be/europractice/europractice.html

commissioned. The number of companies interested in fabricating silicon strip sensors also grew.

. The Mark II vertex detector at the Stanford (e+e-) Linear Collider, was commissioned in 1989 and 1990, and used Hamamatsu sensors and the Microplex readout chip [references]. DELPHI and OPAL at the CERN LEP e+e- collider also commissioned vertex detectors in 1990. Both of these detectors used sensors made by SINTEF (Norway) and the UK readout chip [references]. The ALEPH vertex detector (also at LEP), which used sensors designed at the INFN, Pisa, and fabricated at CSEM (Switzerland), and the CAMEX readout chip [references], was commissioned in 1991. The CDF vertex detector, which used Micron sensors (?) and the SVX readout chip [references] was commissioned in 1992.

Two of the significant improvements designed for this group of detectors and their upgrades were the integration of blocking capacitors, introduced first by Micron Semiconductor in sensors developed for OPAL, and the development of double-sided sensors, used first by ALEPH. AC coupling eliminated the most obvious problems associated with sensor leakage current, and quickly became the norm. The history of double-sided detectors has been much more complex.

Single sided silicon strip detectors are fabricated using p+ strip diode implants in high resistivity n-bulk silicon. Double-sided detectors have p+ strips on one side and n+ strips on the other side. The only implants in the active area of the p-side are the strips, but the n-side requires additional features. This is because the glass passivation layer on covering both sides of the detector becomes positively charged during processing (think of a glass rod and rabbit's fur), and becomes more positively charged with irradiation. This positive charge attracts electrons to the surface of the silicon, making it more n-type than the bulk. This has very little effect on the p-side of the detector, but if the n-side had only n+ strips, the induced n-channels would short all of the strips together. Most double-sided sensors include p+ implants ("p-stops") between the n+ strips[9]. This added complexity adds to the cost of double-sided sensors and limits the strip pitch that is achievable on the n-side.

The other complexity associated with double-sided silicon strip detectors is related to the need to bias the sensor in order to fully deplete it. A single-sided detector can be depleted by applying voltage to the backside and leaving the strips near ground potential. With a double-sided sensor, the strips on at least one side must be biased far from ground. This means that either the readout electronics on that side must also float far from ground [reference to HERA-b], or that blocking capacitors are required. The maximum voltage sustainable across these blocking capacitors has been a limiting factor for a number of double-sided strip detectors [references to CLEO-II, D0, and CDF].

---

[9] Something should be said here about p-spray designs.

Radiation Damage to Silicon Sensors

Silicon sensors were not commonly used in high radiation environments until their use in high energy physics experiments, particularly in those with hadron beams. H.W. Kranner anticipated the most important effects of radiation exposure in a NIM paper published in 1984 [reference "Radiation Damage in Silicon Detectors" H.W. Kranner, NIM 225 (1984) 615-618]. These are greatly increased leakage current and type change of the bulk material (becoming more p-type) "as introduced defects tend to act as acceptors." The same paper also discusses annealing and "reverse annealing."

However, most experimenters did not appreciate the importance of type change, and experiments such as E771 at Fermilab were surprised by the swift degradation in the performance of silicon strip detectors exposed directly to high intensity hadron beams [reference to E771 NIM paper].