

UNIVERSITY OF LIVERPOOL

T. J.V. Bowcock

Semiconductor Detectors for Particle Physics



PREFACE

These notes have been prepared for graduate students who have an interest in semiconductor devices that are used as detectors in particle physics. They are meant to act as a backup to a set of lectures first delivered in Oct 2004 and build on a series of lectures given at Lausanne University in November 2001. It should be emphasized that they draw very heavily on a number of much more detailed resources, for example the excellent textbook by S.M.Sze, *Semiconductor Devices*, published by J.M.Wiley, from which most of the diagrams in the first three chapters are drawn. There are also excellent books on the Web that have been drawn on, for *example Principles of Semiconductor Devices*, B.v. Zeghbroeck

Semiconductor technology is ubiquitous in particle physics, as in almost any form of modern instrumentation. There are very few, if any, detector elements that do not contain semiconductors either as the active sensing element or as part of the readout. The only examples I can only think of are passive materials, such as emulsions, plastics or mica that are sometimes use to record the passage of ionizing radiation and which may, in principle, be "scanned" using traditional microscopes.

In these notes I have not attempted to provide a complete list of semiconductor devices used in experiments. Instead, I have selected a number of sensors and a few electronic components which are critical to their understanding. The sensors I discuss are those usually associated with traditional tracking and vertexing detectors i.e. strip and pixel based detectors for ionizing radiation.

This is a complex and rapidly moving field. As well as the research journals and textbooks referenced in these lectures I would like to thank my colleagues at the Liverpool Semiconductor Detector Centre (LSDC), especially Dr. J. Velthuis and Dr. G. Casse for allowing me to present their results and insights here. Any mistakes and omissions are purely my fault.

Themis Bowcock

Oliver Lodge Laboratory, University of Liverpool

September, 2004

TABLE OF CONTENTS

PREFA	ACE3
BASIC	S10
1.1	PROPERTIES OF MATERIALS10
1.1.1	Conductivity11
1.1.2	Periodic Table12
1.1.3	<i>Structure13</i>
1.1.4	Valence Bonds16
1.2	ENERGY BANDS16
1.2.1	Effective mass18
1.2.2	Holes19
1.2.3	Conduction-Valence Bands 19
1.2.4	Density of States20
1.3	CARRIER CONCENTRATION
1.3.1	Distributions22
1.3.2	Intrinsic Semiconductors
1.3.3	Acceptors and Donors27
1.4	TRANSITIONS
CARRI	ER PROPERTIES31
2.1	TRANSPORT PHENOMENA31
2.1.1	<i>Mobility32</i>
2.1.2	Resistivity34
2.1.3	Hall Effect35
2.2	CARRIER DIFFUSION
2.2.1	Einstein Relation38
2.2.2	Current Density Relation38
2.3	GENERATION/RECOMBINATION38
2.3.1	Direct Recombination
2.3.2	Indirect Recombination
2.4	CONTINUITY EQUATION
2.5	HIGH FIELD EFFECTS
2.5.1	Velocity Saturation
2.5.2	Avalanche Mode45
2.6	SUMMARY OF PROPERTIES OF SI 46
P-N JU	NCTION49
3.1	THERMAL EQUILIBRIUM49
3.2	DEPLETION REGION
3.2.1	Abrupt Junction52
3.3	JUNCTION CAPACITANCE
3.4	CURRENT-VOLTAGE PROPERTIES56
3.4.1	Ideal Characteristics56
3.4.2	Temperature effects59
3.5	BREAKDOWN60
3.5.1	Tunnelling60
3.5.2	Avalanche60
3.6	BIPOLAR TRANSISTOR ACTION61
3.6.1	Current Gain65
3.6.2	Modes of Operation65
JUNCT	TION DEVICES67
4.1	METAL SEMICONDUCTOR67
4.1.1	Energy Band67
4.1.2	Current Voltage
4.1.3	Schottky Diode70
4.1.4	Ohmic Contact70

42	IFET 71
4.2.1	Principle of Operation
4.2.2	<i>Current Voltage Characteristics</i> 73
13	
4.5	MOS CAFACITOR
4.3.1	Principle of Operation
4.3.2	Charge Distribution
122	$C: O \subseteq C:$ Interfere O
4.5.5	SlO_2 -Si Interjace
4.3.4	Radiation Effects 81
44	MOSEET 81
1.1	
4.4.1	Principle of Operation 81
4.4.2	Types of MOSFET 83
WAFE	R PROCESSING 85
5.1	WAFER PRODUCTION
511	Czochralski Mathod 85
5.1.1	
5.1.2	Float Zone Method 87
5.1.3	Wafer Cutting
511	Enitarial Growth
5.1.4	<i>Приилии Отожии</i>
5.2	WAFER PROPERTIES 89
5.2.1	Resistivity
5 2 2	Thickness
5.2.2	<i>Thickness</i>
5.2.3	Crystal orientation 89
5.3	ION IMPLANTATION AND DIFFUSION90
531	Diffusion Equation 01
5.5.1	
5.3.2	Diffusion Profile
5.3.3	Intrinsic and Extrinsic Diffusion92
534	Ion Implantation 02
5.3.4	
5.3.5	Implantation Range
5.3.6	Channelling
537	Disorder 95
5 2 0	Amagling
5.5.0	Annealing 90
5.4	LITHOGRAPHY AND ETCHING
5.4.1	Optical Lithography
542	Mask Production 98
5.4.2	
5.4.3	Negative Litnography
5.4.4	Positive Lithography 100
545	Photo Resist 100
5 4 6	Etahing 101
5.4.0	
5.4.7	High Resolution Lithography. 102
5.5	POLYSILICON 103
551	Deposition Techniques 103
5.5.1	DEFICIENCE LANERS
5.0	DIELECTRIC LAYERS 103
5.6.1	Thermal Oxidation 104
5.6.2	Chemical Vapour Deposition, 104
57	METALLIZATION 104
5.7	IVIETALLIZATION
5.7.1	<i>Sputtering</i> 104
SIMUL	ATION 107
<pre></pre>	
0.1	SIMPLE MODELS 10/
6.2	COMMERCIAL PACKAGE 108
6.2.1	ISE
6 2 2	MECH 100
0.2.2	МЕЗП 108
623	
0.2.5	DIOS 109
6.2.4	DIOS
6.2.4	DIOS
6.2.4 6.3	DIOS 109 DESSIS 109 APPLICATIONS 110
6.2.4 6.3 6.3.1	DIOS
6.2.4 6.3 6.3.1 6.3.2	DIOS
6.2.4 6.3 6.3.1 6.3.2 6.3.3	DIOS

6.3.4	Process Optimization 11-	4
6.3.5	Charge Sharing 11-	4
6.3.6	Charge Transfer 11-	4
Radiat	ion Damage Effects11	4
6.4	LIMITATIONS 114	4
HEP EX	XPERIMENTS 119	9
7.1	ACCELERATORS 11	9
7.1.1	Accelerator Beam Energy 11	9
7.2	LINAC	0
7.2.1	The Electron Gun12	1
7.2.2	Electron sources12	1
7.2.3	High field emission 12.	3
7.2.4	Photo emission 12	4
7.2.5	The Buncher12	4
7.2.6	Acceleration 12	5
7.3	CYCLOTRONS 12	б
7.3.1	Mathematics of the cyclotron. 12	9
7.3.2	Limitations of the cyclotron 12	9
7.4	SYNCHROTRONS 12	9
7.4.1	Synchrotron Radiation 13	0
7.4.2	Focussing and Optics13	0
7.4.3	Storage Rings 13	1
7.5	ACCELERATOR FACILITIES	2
7.5.1	Large Hadron Collider 13	2
7.6	DETECTORS 13	4
7.6.1	Detector Function13	5
7.6.2	Detector Structure 13	5
7.6.3	Fixed Target13	5
7.6.4	Colliding Beams 13	7
7.7	COMPONENTS OF A DETECTOR 13	7
7.7.1	Tracking Detector	7
7.8	VERTEX RESOLUTION 14	0
7.8.1	Impact Parameter 14	1
РНОТС	DETECTORS14	3
8.1	PHOTONS	3
8.1.1	Absorption14	3
8.2	PHOTOCONDUCTOR 14	5
8.2.1	Operation14	5
8.2.2	Limitations 14	6
8.3	Рноторіоде 14	б
8. <i>3</i> .1	Principle of operation14	5
8.3.2	Construction14	7
8.4	AVALANCHE PHOTODIODES 14	7
8.5	APPLICATIONS 14	7
STRIP	DETECTORS14	3
9.1	SIMPLE SENSORS	9
9.1.1	Charge Collection 14	9
9.1.2	Kamo's Theorem	2
9.1.5	I and au Distribution	5 1
9.1.4	Diada Pull	4 5
9.1.3	<i>Dioue Duik</i> 15. DC Couples State Devices 15	ر د
9.2 0 7 1	Principle of Operation 15	0 6
9.2.1	Geometrical Considerations 15	י 7
9.2.2 0 7 3	Readout Schematic 15	' 7
9.2.3 9 3	AC COUPIED STDID DEVICES 15	٬ ۶
931	Principle of Operation 15	8
/1		-

114	9.3.2	Fabrication	159
114	9.3.3	Geometrical Considerations	160
114	9.3.4	Technical Limitations	160
114	9.4	STRIP BIAS TECHNIQUES	161
114	9.4.1	Polysilicon Resistors	161
4.4.0	9.4.2	Reachthrough Bias	161
119	9.4.3	FOXFET bias	162
110	9.5	SPATIAL RESOLUTION	
119	951	Charge Sharing	163
119	9.5.1	Charge Collection	165
120	0.5.2	Diffusion	105
121	9.5.5	Intermediate String	105
121	9.5.4	Analad Tracks	107
123	9.3.3	Anglea Tracks	108
124	9.6	EDGE EFFECTS	169
124	9.7	BULK AND IMPLANTS	170
125	9.7.1	Strip Implants	170
126	9.7.2	Strip Isolation	171
on. 129	9.8	DOUBLE SIDED DETECTORS	172
129	9.9	READOUT	172
129	9.9.1	Bondpads	173
130	9.9.2	Readout Layers	173
130	9.9.3	Electronic noise	174
131	9.10	SENSOR TESTING	174
132			
122	PIXEL	DETECTORS	177
124	2.1	DADG STRIDG AND DIVELG	177
134	5.1	PADS, STRIPS AND PIXELS	1//
150	3.2	PIXEL DETECTORS	1/8
136	3.2.1	Principle of Operation	1/8
136	3.3	BONDING	179
137	3.3.1	Bump Bonding	180
137	3.4	READOUT	181
137	СНАРС		193
140	CHAR	SE COOFEED DEVICES	
141	4.1	CCD	183
4.40	4.1.1	Principle of Operation	183
143	4.1.2	Geometrical Considerations	186
143	413	Fabrication	186
143	414	Tasting Techniques	186
145	4.1.4	Efficiency and Resolution	186
145	4.1.5	Efficiency und Resolution Technical Limitations	186
145	4.1.0	Lieu Coppo De Lo cum	100
140	4.2	HIGH SPEED KEADOUT	180
146	4.2.1	Clocking	180
140	4.2.2	Column Parallel Readout	186
147		CATIONS	.189
147	/		
147	5.1	TRACKING AND VERTEXING	189
1/0	5.2	SLAC	189
143	5.2.1	BaBar - Strip Detectors	190
149	5.2.2	SLD – CCDs	191
149	5.3	FERMILAB	193
152	531	CDF – Strip Detectors	
s 153	5 3 2	D0 - Strin Detectors	
154	8.5.2 RTøV	- Pirols	100
155	5 1	CEPN	100
155).4 5 / 1	ALICE Divela	100
130	J.4.1 5 4 2	ALICE - Fixels	199 202
130	5.4.2	AILAS – Pixels	203
s 157	5.4.3	ATLAS – Strips	203
157	5.4.4	CMS – Strips	203
158	5.4.5	LHCb – Strips	205
150			

RADIATION TOLERANCE......213 6.1 Low energy photons 213 6.1.1 High energy photons 213 6.1.2 6.1.3 6.1.4 Ionizing Radiation......213 6.1.5 3D PROCESSING217 7.1 7.1.1 Principle of Operation......217 7.1.2 Geometrical Considerations...218 7.1.3 Fabrication......218 7.1.4 Technical Limitations......218 7.2 7.2.1 7.2.2 Geometrical Considerations...219 Fabrication......219 7.2.3 Testing Techniques......219 7.2.4 7.2.5 Efficiency and Resolution 219 Technical Limitations......219 7.2.6 7.3

DRIFT DETECTORS...... 221 SOLID STATE DRIFT 221 8.1 8.1.1 Principle of Operation 221 Geometrical Considerations... 222 8.1.2 Fabrication..... 222 8.1.3 8.1.4 Testing Techniques 222 Efficiency and Resolution...... 222 8.1.5 Technical Limitations 222 8.1.6 8.2 APPLICATIONS 222 9.1 9.2 ACTIVE PIXEL SENSORS 223 9.2.1 Principle of Operation 224 9.2.2 Geometrical Considerations... 225 9.2.3 Fabrication...... 225 9.2.4 Testing Techniques 225 9.2.5 Efficiency and Resolution...... 226 9.2.6 Technical Limitations 226

Part 1 Introduction to Semiconductors



BASICS

The simple properties of semiconductors are compared with those of conductors and insulators. Their electronic and atomic structures are discussed. The concepts of energy bands are developed and the carrier density in intrinsic semiconductors calculated. The effects of introducing donors and acceptors into a semiconductor are introduced.

1.1 Properties of Materials

The vast majority of materials of which we have experience may be categorised as existing in the gaseous, liquid or solid phase at room temperature. The physical properties of these materials are almost entirely governed, at the atomic, microscopic and macroscopic scale, by the electric interactions between the electronics surrounding the atomic nuclei.

In the solid state the atoms are sufficiently closely packed that it is often energetically favourable for the atoms to be arranged in regular repeating patterns – crystals. The proximity of the atoms to each other (of order 1\AA) is such that the more loosely bound identical electrons in the outer shells may no longer be localized to the "parent", but may be shared between the atoms in the crystal. As we shall see this phenomenon gives rise to the most important bulk electrical property of materials, conductivity.

Not all solids are crystalline, most have a structure that is far from being simple and regular. Metals are usually composed of an agglomerate of small crystallites packed together. Other materials, for example glasses and amorphous materials, have a periodic atomic arrangement that is so local that material cannot be considered to have any crystalline structure. Organic solids may exist as polymer chains. Even crystals are far from simple; many will possess impurities, either accidental or deliberate, and most will contain regions where the atoms are displaced or missing within the structure. Within these notes we will be largely be concerned with the electrical properties of relatively pure crystalline solid state materials. We will almost exclusively concentrate on a single element, Silicon, from which most solid state devices are fabricated. We will learn how the control and modify of the crystal, through the introduction of impurity atoms, to make sensors of exceptional sensitivity for particle physics experiment.

1.1.1 Conductivity

The resistivity of a material is measured in Ω -cm. If the resistance, of a sample of material is known, the resistivity is related to *R* through

$$R = \frac{\rho L}{A} \tag{1.1}$$

where we have assumed we have a block of cross-sectional area A and length L. The conductivity is the reciprocal of the resistivity, $\sigma \equiv 1/\rho$, and has units of S/cm.

The conductivity of materials has one of the highest ranges of any known measurable quantity. It varies from about 10^{-22} S/c to about 10^{10} S/cm. This is one of the largest ranges of any measurable property of a material known. Materials that do not conduct, such as quartz, as classed as insulators. In contrast many metals, for example silver and copper, are exceeding good conductors. Between the range of insulators and conductors we have a class of materials



Figure 1:Conductivities of insulators, semiconductors and conductors

which we can generically call semi-conductors.(See Figure 1).

The physical difference between and insulator and a conductor is very strongly related to the number of free charges able to move under the influence of an external electric field. The current density j and field and are related through

$$j = \sigma \mathcal{E} = \mathcal{E} / \rho \tag{1.2}$$

Simple argument from kinetic theory suggests that

$$\rho = mv_F / Ne^2 l \tag{1.3}$$

where *m* is the effective mass of the charge carrier, *e* is the charge of carriers, *l* is the spacing between collisions v_F , is related to velocity of the carriers (not the drift velocity induced by the electric field), and *N* is density of carriers. For a metal *N* may typically be greater than 10^{28} m⁻³ whilst for an insulator *N* may be very low. We will discuss the carrier concentrations in semiconductors in Section 1.2. The conductivity of semiconductors is in general very sensitive to temperature, light, magnetic field and the levels of impurity, It is this sensitivity, not present in metals at room temperature, that makes semiconductors such useful materials for sensors.

1.1.2 Periodic Table

By far the most important semiconductor for detector development is silicon. The discovery of silicon (L. *silex: silicis*, flint) - silicium in French - is generally credited to Berzelius 1824. He prepared amorphous silicon heating potassium with silicon tetrafluoride and purified the product by removing the fluorosilicates by repeated washings. Deville in 1854 first prepared crystalline silicon, the second allotropic form of the element.

Silicon is present in the sun and stars and is a principal component of a class of meteorites known as aerolites. It is also a component of tektites, a natural glass of uncertain origin.

Silicon makes up 25.7% of the earth's crust, by weight, and is the second most abundant element, being exceeded only by oxygen. Silicon is not found free in nature, but occurs chiefly as the oxide and as silicates. Sand, quartz, rock crystal, amethyst, agate, flint, jasper, and opal are some of the forms in which the oxide appears. Granite, hornblende, asbestos, feldspar, clay, mica, etc. are but a few of the numerous silicate minerals.

Silicon is one of man's most useful elements. In the form of sand and clay it is used to make concrete and brick; it is a useful refractory material for hightemperature work, and in the form of silicates it is used in making enamels, pottery, etc. Silica, as sand, is a principal ingredient of glass, one of the most inexpensive of materials with excellent mechanical, optical, thermal, and electrical properties. Glass can be made in a very great variety of shapes, and is used as containers, window glass, insulators, and thousands of other uses. Silicon tetrachloride can be used as iridize glass.

Atomic Number	14	Melting Point	1687 K
Atomic Symbol	Si	Boiling Point	3538 K
Atomic Weight	28.0855	Density	2.3296 g/cm3
Electronic Configuration	$(Ne)(3s)^2(3p)^2$	Ionization Potential	8.1eV

Some of the properties of Silicon are summarized in

Table 1: Properties of Silicon

INTRODUCTION TO SEMICONDUCTORS

Silicon is in period 3 in the periodic table. Note that Silicon occurs close to germanium (Ge) in Column IV. Note also that it is not simply pure elements that are used in semiconductor devices. Many compounds, for example GaAs are may also used.

Period	Column II	III	IV	V	VI
2		В	С	Ν	
		Boron	Carbon	Nitrogen	
3	Mg	Al	Si	Р	S
	Magnesium	Aluminium	Silicon	Phosphorus	Sulphur
4	Zn	Ga	Ge	As	Se
	Zinc	Gallium	Germanium	Arsenic	Selenium
5	Cd	In	Sn	Sb	Te
	Cadmium	Indium	Tin	Antimony	Tellurium
6	Hg		Pb		
	Mercury		lead		

Table 2 Part of the periodic table containing silicon

The electronic structure of silicon is determined by its atomic weight, 14. The ground state of Si will then make it Neon like, with a closed inner shell composed of two electrons and 8 electrons in the next shell.¹ The remaining 4 electrons fill the 3s level and start to fill the 3p level. These last 4 electrons are the easiest to ionize with a typical ionization energy of about 8eV.

1.1.3 Structure

As we discussed above when atoms are close packed into a solid they may form into fixed structures. These lattice structure, if regularly repeated, form the crystals of the material under consideration. In a crystal the atoms do not move far from fixed positions. At normal temperatures the only excursions are small thermal vibrations within the lattice.



Figure 2: Three cubic unit cells. (a) Simple cubic. (b) Body-centred cubic. (c) Face-centred cubic.

¹ The number of electrons that may be put into a particular angular momentum state l is given by 2(2l+1). Thus in the inner shell (l=0 only) we get 2,. For the next shell where both l=0 or 1 we can fit 8 and so on.

The simplest crystal structure that could be envisaged would be a simple cubic crystal structure. See (Figure 2a). Although the simple cubic structure is simple to explain and draw there is only known element that crystallizes in the simple cubic structure, polonium. The open structure makes it much more probable that crystals have a much more close-packed structure.

The simplest way to increase the packing density is to "fill" the centre of the cube. This clearly symmetric structure is shown in Figure 2(b). It may be thought of as being constructed of two interlinked simple cubic structures displaced with respect to one another along the body diagonal. See Figure 3.



Figure 3. The body-centred cubic structure shown as two interlinked simple cubic lattices. Many metals possess this structure including iron, lithium, potassium and sodium.



Figure 4: STM picture showing a fcc structure on a surface. (IBM)

lattice displaced from the other by one quarter of the body diagonal of the fcc. In the diamond lattice each atom is equidistant from four others. The diamond structure is almost as complicated to explain as it is to visualize, especially if one tries to see the interlocking sublattices. It is somewhat easier (Figure 5) to identify the tetrahedrons. Each atom in the diamond structure can be thought of as being at the centre of its own tetrahedron.

Another packing structure is called faced centred cubic (fcc). See Figure 3(c). The fcc structure is very similar to simple cubic structure with an additional atom added at the centre of every face. Fcc structures are typical of many other metals including copper, silver, gold, aluminium and nickel and platinum, Figure 4.

Both silicon and germanium have a much more complicated structure. They share a diamond lattice structure which may be described as two fcc

lattices

with one



Figure 5: Diamond structure

INTRODUCTION TO SEMICONDUCTORS

If we know the approximate size of a Silicon or Carbon atom and the interatomic spacing we can make a classical estimate of the amount of free space to atoms. For silicon this ratio is about 0.3 which may be compared with the value for a closely packed (classical) system which occupies over 0.7. (see Problem 1.1 and 1.2)

Note that distances between atoms can be measured very accurately using X-ray diffraction techniques. The size of an atom remains a much move vague concept -strictly it depends whether the atom is free or in a metal or ionic crystal etc.

Before moving away from this very brief discussion of the crystal structure of silicon it is necessary to define the *Miller* indices. These are used to define planes through the crystal. As some properties of sensors depend on the crystal orientation we will describe how these may be obtained. The recipe is as follows. First determine the intercepts of the planes on the three Cartesian coordinates in terms of the lattice constants. If a plane does not intersect the axis then the intercept is infinite, then take the reciprocals of the numbers and reduce them to the smallest three integers having the same ratio, e.g. (111) rather than (222).



Figure 6: Miller indices of some planes in a cubic crystal



Figure 7 Silicon (111) plane (solid line), and a STM photograph of a Si 111 surface.

Note it is now possible to use STM technology to examine the crystal structure. For example the (111) plane of Si contains a complex set of atoms on which the pattern is far from the simple fcc.

1.1.4 Valence Bonds

The silicon crystal is bound together via valence bonds. Silicon is tetravalent and is able to share its outer four electrons with each of its four neighbours. Each identical electron cannot thus be uniquely associated with a particular atom. The force of attraction between the atoms can only be calculated using complex quantum chemistry techniques and involves a detailed knowledge, or prediction, of the electron wavefunctions in the solid. However the general features of the solid follow the same rules as for ionic crystals or metals i.e. that

- positive ion cores should be kept apart to minimize the Coulomb energy
- Valence electrons should also be kept apart
- Valence electrons should stay close to positive ions to maximize attraction

For covalent crystals e.g. Silicon or carbon the bond is very strong. For carbon the bond strength between atoms is over 7eV. The other important feature of the covalent bond is that it is strongly directional. It is this that gives rise to the tetrahedral diamond structure we have already discussed and hence why the packing ratio, and hence density, of the diamond lattice is so low.

A curiosity associated with the Si diamond structure is that electrons localized within a "bond" have the tendency to be spin aligned. This counter intuitive result is a product of the interaction of the electronic with the Si nuclei; this effect (except for very short distance scales) tends to outweigh the usual Pauli exclusion principle.

1.2 Energy Bands

In order for us to understand the behaviour of basic semiconductor devices we need to introduce some concepts that arise from the study of the quantum Mechanics of solids. Most of these will already be familiar to the reader. The energy bands, the Fermi Surface and energy are particularly important.



Figure 8: The Energy wavevector relation for a) a free electron and b) an electron in a periodic lattice.

For a free electron there exists a simple relationship between its energy and its momentum

$$\varepsilon_k = \frac{\hbar^2}{2m} \vec{k}^2 \tag{1.4}$$

where \vec{k} is related to the momentum by, See

$$\vec{p} = \hbar \vec{k} \tag{1.5}$$

In the free electron case there are no forbidden energy regimes, the electron can freely propagate in its usual way with a wavefunction of the form

$$\psi \propto e^{i\vec{k}.\vec{r}}$$

When the electrons are propagating in a periodic lattice the situation is much more complex. Intuitively when the wavenumber (k) of the electron matches the periodicity of the lattice we expect there to be coherent scattering from each of the atomic centres. This coherent scattering has the effect of creating standing waves within the crystal i.e. the electrons at selected frequencies (and directions) do not easily propagate. This means that certain momentum values are "forbidden", in as much electrons are not observed to propagate with these k values, i.e. there are certain energy "band-gaps". This vision of the origin of the band gaps corresponds to the solutions of the Schrödinger equation and possesses the underlying assumption that we have quasi-free electrons moving in the presence of a lattice of positive ions.

This technique results in the Bloch Theorem which states that the eigenfunctions of the wavefunction in the crystal (or periodic potential) must have the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) \tag{1.6}$$

where the function $u_k(\mathbf{r})$ has the periodicity of the lattice depends on the wavevector k. Note that for a relatively complex crystal the solutions will clearly depend strongly on the wavevector. The wavefunctions, which contain all the information about the electrons in the solid, may then be used to compute the energy-wavevector relations, analogous to the free electron equation above.

In a one dimensional crystal the relation between is the wavefunction and the energy bands is relatively simple to compute Let use assume that the periodic potential in a lattice may be written as

$$V(r) = \sum_{G} V_{G} e^{iGx}$$
(1.7)

Where G are reciprocal lattice vectors. The Schrödinger equation will be doubly degenerate in energy corresponding to opposite travelling waves .Thus from the Bloch equation we can write the general form of the wavefunction as

$$\Psi = aL^{-1/2}e^{-ikx} + bL^{-1/2}e^{ikx}$$
(1.8)

Where the $L^{-1/2}$ terms are simply the free electron normalization factors for a particle in box of length L.

First order perturbation theory tells us that we can compute the energies of two degenerate states (i,j) due to the perturbation V(r) if we can "diagonalize" the

matrix elements V_{ij} Actually we don't have to perform the diagonalization but simply recall that the energy shift(s) are given by solutions to the equation

$$\left(V_{11} - E_{ni}^{1}\right)\left(V_{22} - E_{ni}^{1}\right) - V_{21}V_{12} = 0$$
(1.9)

In our case

$$V_{11} = \frac{1}{L} \int_{0}^{L} \sum_{G} V_{G} e^{iGx} dx = V_{22}$$

$$V_{12} = \frac{1}{L} \int_{0}^{L} \sum_{G} V_{G} e^{iGx} dx e^{-i2kx}$$

$$V_{21} = \frac{1}{L} \int_{0}^{L} \sum_{G} V_{G} e^{iGx} dx e^{i2kx}$$
(1.10)

Now if G is sufficiently large compared to 1/L then both V_{11} and V_{22} are zero as they integrate of a rapidly oscillating function. By the same token V_{12} is only non zero if G=2k. At this value $V_{12}=V_{G}$. This is also true for V_{21} . Thus Eq. (1.9) becomes (putting $\Delta E=E^{l}$)

$$\Delta E^2 = V_G \tag{1.11}$$

which means that the two energies have split, $\Delta E = \pm V_G$. This also means that for any periodic perturbing potential energy gaps appear of width approx $2V_G$. It should also be obvious that at k=G/2 the lower energy state and the higher energy state "bend" away from the free electron energy.

1.2.1 Effective mass

In quantum mechanics the velocity of a particle is measured by its group velocity

$$v = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk}$$
(1.12)

The acceleration of the particle is thus simple

$$\dot{v} = \frac{1}{\hbar} \frac{d}{dt} \frac{dE}{dk} = \frac{1}{\hbar} \frac{d}{dk} \frac{dk}{dt} \frac{dE}{dk} = \frac{1}{\hbar} \frac{dk}{dt} \frac{d^2 E}{dk^2}$$
(1.13)

In the case we apply an external field \mathcal{E} the rate of energy absorption from the external field is force × velocity. Thus

$$-e\mathcal{E}v = \frac{dE}{dt} = \frac{dE}{dk}\frac{dk}{dt}$$
(1.14)

Further from Eq. (1.12) we can substitute for the velocity giving

$$-e\boldsymbol{\mathcal{E}} = \hbar \left(\frac{dk}{dt}\right) \tag{1.15}$$

Thus using Eq. (1.13) we find

$$-e\boldsymbol{\mathcal{E}} = \frac{\hbar}{\left(d^2 E / dk^2\right)} \dot{v}$$
(1.16)

Which defines the effective mass

Figure 9 Band structure of Si for different in 1D directions and values of k INTRODUCTION TO SEMICONDUCTORS

$$m^* = \frac{\hbar}{\left(d^2 E / dk^2\right)} \tag{1.17}$$

For an electron at the top of a band, for example the lower energy state in the previous section, the effective mass is negative, the electron decelerating as the field is applied and the electron exchanges momentum with the lattice. At the bottom of the band the electron behaves much as a free particle having an E-k relation almost the same as that in free space.

1.2.2 Holes

The peculiar behaviour of an electron close to the top of an energy band is usefully understood in terms of the behaviour of vacant states, or *holes*. In a full energy band containing n electrons with no net current flow

$$-e\sum_{n}v_{n}=0$$
(1.18)

If the m^{th} is missing there will, instantaneously, be a net current equal to

$$-e\sum_{n\neq m}v_n = ev_m \tag{1.19}$$

i.e. the net current is the same as that due to a particle with the same velocity as the missing electron but carrying opposite charge.

Thus for a system containing a missing electron close the top of the band, the behaviour of the system may be considered to be the same as that of a *positive* particle moving with an effective mass which is *positive*.

1.2.3 Conduction-Valence Bands

It is usual to call the upper of the two bands in the energy diagram the conduction band and the lower of the two the valence band. The energy gap, if it exists, between the two is the bandgap.

The reader should also be made aware the simple 1D analysis of band structure ignores the complex 3D nature of crystals. The energy diagrams, and bandgap, depends on the wavevector orientation of the electrons. In general the band structure is a three dimensional object.



Figure 10 Band structure of Si for different in 1D directions and values of k



Figure 11: 3D band structure of Si. The radius represents the energy level at $k=6\pi$ and the direction is that of the appropriate k.

1.2.4 Density of States

Apart from the *E-k* relationship for the material it is also vital to know the numbers of available states to the electron. The density of states follows the well known relation from quantum mechanics for a free electron. To make this more explicit let us consider a particle of momentum p inside a cubic volume with sides of length *L*. The allowed values of the momentum (in the *x*-direction) are given by:

$$p_x = \hbar k_x = \frac{2\pi\hbar n_x}{L} \tag{1.20}$$

and likewise for p_y and p_z . The numbers n_x , n_y , n_z are all positive integers. Every combination of these integers represents a different state. The number of states in one dimension (x) in an interval dpx is given by the differential of Eq. (1.20) namely

$$dp_x = \frac{2\pi\hbar}{L} dn_x \tag{1.21}$$

In three dimensions the total numbers of states dN is thus given by:

$$dN = dn_x dn_y dn_z = \left(\frac{L}{2\pi\hbar}\right)^3 dp_x dp_y dp_z$$
(1.22)

where if we set the volume of the confining cube to be V then:

$$dN = dn_x dn_y dn_z = V \left(\frac{1}{2\pi\hbar}\right)^3 dp_x dp_y dp_z$$
(1.23)

we can interpret the above equation as

$$dN = \frac{volume_{phase \ space}}{h^3} \tag{1.24}$$

where Eq. (1.24) is read as stating that the number of states available equals the phase space volume scaled by Plank's constant. In spherical co-ordinates (that we often use because of the spherical symmetry inherent in many problems in nature) we know that

$$d^{3}p = dp_{x}dp_{y}dp_{z} = p^{2}dpd\Omega = p^{2}dpd(\cos\theta)d\phi$$
(1.25)

so that now we may rewrite Eq.(1.24) as:

$$dN = \frac{V}{h^3} p^2 dp d\Omega \tag{1.26}$$

We can convert this to an energy density by dividing through by dE:

$$\rho(E) = \frac{dN}{dE} = \frac{V}{h^3} p^2 \frac{dp}{dE} d\Omega$$
(1.27)

Remembering that classically E=p2/2m then if we let the velocity of the particle be *v*:

$$\frac{dp}{dE} = \frac{1}{\nu} \tag{1.28}$$

then can simply be rewritten as

$$\frac{dp}{dE} = \frac{1}{v} \tag{1.29}$$

Sometimes this is written in the entirely equivalent form

$$\rho(E) = \frac{V}{\left(2\pi\hbar\right)^3} \, pmd\Omega \tag{1.30}$$

Integrating over the solid angle

$$\rho(E) = 4\pi \frac{V}{(2\pi\hbar)^3} \, pm = 4\pi \frac{V}{(\hbar)^3} m^{3/2} \, 2^{1/2} E^{1/2}$$
(1.31)

As there are two electron states (spin up and spin down) for each wavefunction within the lattice the density of states per unit volume is

$$\rho(E) = 4\pi \left(\frac{2m}{h^2}\right)^{3/2} E^{1/2}$$
(1.32)

It is interesting to note (although we do not prove it here) hat the number of states available to a free electron is equal to N, the number of ions in the lattice. Since there are the two spin state this becomes 2N.

The differences between metals, insulators and semiconductors can be understood in terms of both E-k plots and the density of states. For a monovalent metal the first (valence band) will only be half filled (only N of the 2N states) are filled. Given the shape of the E-k curve at this point one can see why the electrons in the monovalent metal do behave as free electrons.



Figure 12 E-k curves and density of states for a) conductor) insulator and c) a semiconductor.

For an insulator the first energy band is completely filled. In order for an electron to move within the crystal when an electric field is applied energy must be supplied (from thermal excitation) to drive an electron across the forbidden region into the conduction band. If the band gap is sufficiently large, compared to thermal energies, i.e. several eV, the transition rarely takes place and the material is an insulator.

If the energy gap is not too large (1.1eV in Si) then thermal excitations will leave some electrons in the conduction band at all time. As the temperature is increased(decreased) the conductivity will increase(decrease). This is the classic behaviour of a semiconductor.

1.3 Carrier Concentration

Before we discuss the density of charge carriers in semiconductors it is important to introduce both the Fermi Energy, Fermi Surface and the Fermi-Dirac distributions.

1.3.1 Distributions

The energy distribution of spin $\frac{1}{2}$ particles at temperature *T* is governed by the Fermi-Dirac distribution function:

$$F_{FD}(E) = \frac{1}{1 + e^{(E - E_F)/kT}}$$
(1.33)

Where EF is the known as the Fermi Energy. The Fermi-Dirac distribution may derived from the same statistical mechanical procedure used to show that systems of indistinguishable particles are distributed according to the Maxwell-Boltzmann distribution

$$F_{MR}(E) \propto e^{-E/kT} \tag{1.34}$$

The difference between the two arises from the requirement that for spin $\frac{1}{2}$ particles no two particles may be in the same state.

The Fermi-Dirac distribution has a very distinctive shape (see Figure 13 a) which varies from a step function at absolute zero to a smoothly varying function at higher values of T.



Figure 13: Shape the (a) the Fermi-Dirac Function (b) Density of states (c) the net number density of electrons as a function of energy

Note that the actual number density of electrons is the product of the density of states and the Fermi-Dirac function. In particular the Fermi-Dirac distribution gives the probability of occupation of a state with energy E, but does not tell us whether those states exist. In the bandgap the number density will be zero in a semiconductor but the *FFD* is still continuous _{Cu} and non-zero.

The Fermi energy, *EF*, may be defined as the energy at which the probability of occupation in *FFD* is 0.5. Another physical definition, entirely equivalent, is to define *EF*, as the energy of the maximum allowed state at T=0.

The Fermi energy, *EF*, is also the energy associated with a particle, which is in thermal equilibrium with the system of interest. This same quantity is called the electro-chemical potential, μ , in most thermodynamics texts.

In 3D, and again at absolute zero, the Fermi



Figure 14: Fermi Surface for Cu

Surface separates the unfilled orbitals from the unfilled orbitals. The Fermi surface has been calculated for many elements. For example Figure 14 show the Fermi Surface of the metal Cu.

1.3.2 Intrinsic Semiconductors

In semiconductors thermal energy leads to the excitation for electrons through the bandgap from the valence band to the conduction band. If there are no, or very few impurities, in a material compared with the density of the electrons/holes the semiconductor is called *intrinsic*. The number of holes in an intrinsic semiconductor equals the numbers of electrons in the conduction band.



Figure 15: Intrinsic semiconductor (a) Schematic band diagram b) density of states c) Fermi distribution function d) Carrier concentration

In Figure 15 the band diagram, density of states as well as the Fermi-Dirac distribution and carrier concentrations are shown schematically. Note that in the conduction band there are very few carriers (electrons) and similarly there are only a few vacancies (holes) in the valence band.

Remembering that the Fermi energy is defined as the energy of the state at which the probability of occupation is $\frac{1}{2}$ we would like to understand where the *EF* lies in Figure 14.

Let us assume, and we will show this is a good assumption, that the Fermi energy lies approximately in the middle of the bandgap for an intrinsic semiconductor. Then the since the bandgap is much greater than kT (in Silicon bandgap more than a factor of 10 than the thermal energy)) we can approximate the Fermi-Dirac distribution by

$$F_{FD}(E) \cong e^{-(E-E_F)/kT} \qquad \text{for } (E - E_F) \gg kT$$

$$F_{FD}(E) \cong 1 - e^{-(E_F - E)/kT} \qquad \text{for } (E - E_F) \gg kT$$
(1.35)

Thus the number of electrons in the conduction band may be computed

$$n = 4\pi \left(\frac{2m_n}{h^2}\right)^{3/2} \int_0^\infty E^{1/2} e^{-\frac{(E-E_F)}{kT}} dE = 4\pi \left(\frac{2m_n kT}{h^2}\right)^{3/2} e^{E_F/kT} \int_0^\infty y^{1/2} e^{-y} dy$$
$$= 2 \left(\frac{2\pi m_n kT}{h^2}\right)^{3/2} e^{E_F/kT}$$
(1.36)

Where we have used the substitution y = E/kT and use the result

$$\int_{0}^{\infty} y^{1/2} e^{-y} dy = \sqrt{\pi} / 2$$
 (1.37)

and have also taken (initially) the energy of the bottom of the conduction band to be defined as EC=0.

In the general case where we refer to the bottom of the conduction band as EC we may write

$$n = 2 \left(\frac{2\pi n_n kT}{h^2}\right)^{3/2} e^{-(E_C - E_F)/kT} = N_C e^{-(E_C - E_F)/kT}$$
(1.38)

where NC

$$N_C \equiv 2 \left(\frac{2\pi m_n kT}{h^2}\right)^{3/2} \tag{1.39}$$

is the effective density of states in the conduction band. At room temperature for Si $N_C = 2.8 \times 10^{19} \text{ cm}^{-3}$.

The expression for the hole density (p) may be derived in a similar and yields

$$p = 2 \left(\frac{2\pi n_p kT}{h^2}\right)^{3/2} e^{-(E_c - E_F)/kT} = N_c e^{-(E_F - E_V)/kT}$$
(1.40)

Where

$$N_V \equiv 2 \left(\frac{2\pi n_p kT}{h^2}\right)^{3/2} \tag{1.41}$$

Equation (1.41) is the effective density of states in the valence band. At room temperature for Si. $N_V = 10^{19} \text{ cm}^3$

Since $n=p=n_i$. for an intrinsic material we may equate our carriers per unit volume in the conduction and valence bands. Thus

$$E_F = E_i = \frac{E_V + E_C}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right)$$

$$= \frac{E_V + E_C}{2} + \frac{3kT}{4} \ln\left(\frac{m_p}{m_n}\right)$$
(1.42)

The intrinsic Fermi energy E_i at room temperature is approximately half way between the bandgap as assumed at the beginning of this section, (since kT is much less than the bandgap).

The density of carriers is obtained by eliminating EF from the above equations

$$np = n_i^2 \tag{1.43}$$

So that

$$n_i^2 = N_C N_V e^{-E_g / kT}$$
(1.44)

where we have called $Eg = E_C - E_V$ the bandgap. The intrinsic carrier density for silicon is shown as a function of temperature in Figure 16

Since the level of carriers is so low this means that to observe intrinsic behaviour in Si (at room temperature) we must work with very pure samples in which the ionized impurity level must be less than about 10^{-10} of the atomic density at room temperature.

Note that in Eq.(1.43) there is no reference (by design) to the Fermi Energy, nor

most importantly to the intrinsic nature of the material. In fact Eq. (1.43) is known as the law of mass action. The only important assumption in its derivation is that the Fermi level is far (in terms of kT) from either top of the valence or the bottom of the conduction band.

It may be seen from Figure 16 that the exponential form is very closely followed for a sample of Si, and this shape will be followed for almost all samples of pure or doped Si.

One of the most important implications of the mass action law is that because the product of the hole and densities electron is a constant, the addition of even a small amount of impurity we can drastically control the carriers density in the material.



Figure 16: Intrinsic carrier density in Si

One could imagine detecting

the numbers of electrons(holes) created in the valence(conduction) band in an intrinsic semiconductor as a potential device for detecting ionizing radiation. A highly energetic cosmic ray will create $O(10^5)$ eh pairs on passing through a

mm³ of Si (as we shall see later). If we could sense the eh pairs in a volume of about 1mm³ the number of free carriers at $T=0^{\circ}C$ is approx 10^{7} i.e. extra mobile carriers would only be 1% of the thermally excited number. From the Figure 16 cryogenic temperatures (~-100°C) would be required to make the increased number of carriers readily detectible. Alternatively a way must be found to "remove" the mobile carriers "masking" our signal.

1.3.3 Acceptors and Donors

If the semiconductor is doped then impurity energy levels are introduced. For example if an arsenic atom is introduced into the lattice it may displace a silicon atom. The arsenic has an additional valence electron to silicon and "donates" this to the conduction band making the silicon more "n" (negative) type because of its carrier type. Similarly if a boron atom displaces a Si atom then as it only possesses three valence electrons then a hole is created in the crystal. This then makes the crystal more "p"-type (positive charge carrier). The arsenic atom is a donor(of electrons), the boron an acceptor(of electrons).

When the crystal properties are dominated by the impurities it is said to have become an extrinsic semiconductor to differentiate it from the intrinsic semiconductors discussed above.



Figure 17: n-type semiconductors showing a) the band diagram b) the density of states c) the Fermi distribution and d) the carrier density.

In very many cases there is enough thermal energy to completely ionize the donors or acceptors. In the case of donors this would make the electron density approximately

$$n = N_D \tag{1.45}$$

The Fermi level under these conditions is determined from Eq. (1.38) and becomes

$$E_{c} - E_{F} = kT \ln(N_{c} / N_{D})$$
(1.46)

We can see that as the numbers of donors is increase towards the Fermi energy moves towards the conduction band. An example of this is shown in Figure 17

Similarly for (shallow) acceptors we find that

$$p = N_A \tag{1.47}$$

and this that the Fermi Energy is given by

$$E_F - E_V = kT \ln\left(\frac{N_V}{N_A}\right) \tag{1.48}$$

From inspection as the numbers of acceptors is increase the Fermi level moves towards the valence band.



Figure 18: Fermi Energy as a function of temperature and doping for n-type and p-type Si

We may also express the electron and hole densities in terms of the intrinsic carrier concentrations using the formulae we have derived above. From Eq.(1.38) and Eq.(1.40) we can write

$$n = N_c e^{-(E_c - E_F)/kT} = N_c e^{-(E_c - E_i)/kT} e^{(E_F - E_i)/kT} = n_i e^{(E_F - E_i)/kT}$$
(1.49)

and similarly

$$p = n_i e^{(E_i - E_F)/kT}$$
(1.50)

These are the Boltzmann Relations. One very important case to consider is when both donor and acceptors are present at the same time. The material must maintain electrical neutrality thus the total number of electrons and ionized donors must equal the number of holes and ionized acceptors

$$n + N_A = p + N_D \tag{1.51}$$

Using the mass action law this may be solved for n (for example for an n type semiconductor where the number of donors is larger than the number of acceptors) as

$$n_n = \frac{1}{2} \left(\left[N_D - N_A \right] + \sqrt{\left(N_D - N_A \right)^2 + 4n_i^2} \right)$$
(1.52)

In this case if the intrinsic carrier density may be ignored, for example in an ntype material where the electrons are the majority carrier

$$n_n \sim N_D - N_A \tag{1.53}$$

And similarly for a p-type material where the holes are a majority carrier

$$p_p \sim N_A - N_D \tag{1.54}$$

The Fermi-Energy for both p and n-type Silicon as a function of doping and temperature is shown.

In a later chapter we shall see how by the introduction of impurities into a semiconductor enables us to create a depletion region, i.e. one depleted of mobile carriers, which is thus capable of acting as a radiation sensing detector.

1.4 Transitions

Transitions between an electron in the valence band to conduction band in semiconductors can be typified by a change or wavevector or not. If the wavevector does not change direction this is known as a direct transition and corresponds to the transition taking place without any resultant change of energy in the lattice. In this kind of transition the atom is simply ionized. For indirect transitions for the transition to take place energy must be exchanged with the lattice.



Figure 19: General schematic for direct and indirect transitions

Figure 20 Shows the band diagrams for Si and GaAs. By inspection Si will be dominated by indirect transitions as is thus known as an indirect semiconductor



Figure 20: Band diagram for Si and GaAs

References

- 1 C. Kittel, Introduction to Solid State Physics, Fourth Edition, John Wiley and Sons
- 2 S.M.Sze, Semiconductor Devices, John Wiley and Sons
- 3 H.M.Rosenberg, The Solid State, Second Edition, Oxford Physics Series
- 4 http://www.hpc.susx.ac.uk/~venables/jony/
- 5 Principles of Semiconductor Devices, B.v. Zeghbroeck, http://ecewww.colorado.edu/~bart/book/
- 6 http://www.ini.unizh.ch/avlsi/winterlabs/lec1semiphy.pdf

Problems

- 1 Calculate the fraction of "free" space in a bcc lattice given the atom at the centre touches the atoms at the corners of the lattice
- 2 Calculate the fraction of "free" space in a silicon tetrahedral (diamond lattice) if the interatomic spacing is twice the nominal covalent radius of Si 1.17Å. The nominal ionic radius of free silicon is 0.38Å. Comment on the fraction of free space in a tetrahedral silicon if the atoms in the crystal were the same size as the free crystals.
- 3 The periodic potential in a one dimensional lattice of spacing a can be approximated by a box-shaped wavefunction which has a value at each atom of -V. and width a/10 and which is zero elsewhere. Estimate the width of the bandgap in the electron energy spectrum.



CARRIER PROPERTIES

The properties of carriers within a semiconductor are discussed. Fundamental properties if the materials, for example, resistivity are discussed in terms of the carrier properties. Finally the continuity equation, which governs the motions of the carriers, is developed

2.1 Transport Phenomena

ur purpose in studying the properties of semiconductors is to learn how they may be used to build useful detectors for sensing radiation. In the last chapter we touched on the possibility of using a small volume of an intrinsic semiconductor to detect the creation of electron-hole pairs. In order to progress further it is necessary to understand the properties of the mobile carriers, electrons and holes within the semiconductors.

One of the classic problems that is set for undergraduate students is to calculate the kinetic energy of electrons within a material, and from that expression to derive an expectation value for the average thermal velocity for the electrons. This is simply done recalling that each degree of motional freedom possesses, on average, rise to $\frac{1}{2} kT$ of energy thus

$$\frac{1}{2}mv_{thermal}^2 = \frac{3}{2}kT \tag{2.1}$$

At room temperature the student is often surprised to find that the average thermal velocity is approximately 10^7 cm/s for Si. This type of motion is random and is rapidly reversed as the particle hits many scattering centres per second, typically the time between interactions, τ_c , being as small as a ps.

This random motion is to be compared with the drift velocity of the electron. When a small electric field is applied to the material a free electron gains momentum in time proportional the time the force is applied.

$$m_n v_d = -q \mathcal{E} \tau_c \tag{2.2}$$

Thus the drift velocity is simply proportional to the applied electric field

$$v_d = -\frac{q\tau_c}{m_n} \mathcal{E}$$
(2.3)

2.1.1 Mobility

The mobility of the electron μ_n , is simply the proportionality constant in Eq. (2.3) so that the for the electron

$$\mu_n \equiv q \,\tau_c \,/\, m_n \tag{2.4}$$

The mobility of the holes may be similarly defined using

$$v_p = \mu_p \mathcal{E} \tag{2.5}$$



Figure 21: Mobility for electrons in Si as a function of temperature. For (1) High purity Si (Nd< 10-12 cm-3); time-of-flight technique (2) High purity Si (Nd< 4.10-13 cm-3): photo-Hall effect (3) Nd= 1.75.1016 cm-3; Na = 1.48.1015 cm-3; Hall effect(4) Nd= 1.3.1017 cm-3; Na = 2.2.1015 cm-3; Hall effect. The straight line on the plot give has slope T-3/

The mobility of a carrier is a direct function of the amount of scattering (i.e. of the mean free time between collisions) that takes place in the material. The more scattering that takes place the lower the mobility we expect. Thus the introduction of impurities into the crystal should reduce the mobility. Moreover as the lattice vibration increases with temperature we expect the cross-section for scattering to increase and thus the mobility to decrease with temperature. However impurity interactions have a completely different T dependence.

The scattering between the carrier and the ionized impurity depends on the interaction time and the number of impurities. Large impurity concentrations will result in a lower mobility. The more careful study of interaction time explains the



Figure 22: Transition between impurity and lattice/phonon scattering

Rutherford scattering.

temperature dependence of impurity scattering

The interaction time is directly linked to the relative velocity of the carrier and the impurity. This thermal velocity increases with the ambient temperature so that the interaction time decreases. Thereby the amount of scattering decreases, resulting in a mobility increase with temperature. To first order, the mobility due to impurity scattering is proportional to $T^{3/2}/N_{\rm I}$, where $N_{\rm I}$ is the density of charged impurities. For ionized impurities the scattering mechanism is similar to

This is to be compared with the scattering by lattice waves which includes the absorption or emission of either acoustical or optical phonons. Since the density of phonons in a solid increases with temperature, the scattering time due to this mechanism will decrease with temperature as will the mobility. Theoretical calculations reveal that the mobility in non-polar semiconductors, such as silicon and germanium, is dominated by acoustic phonon interaction. The resulting mobility is expected to be proportional to $T^{-3/2}$, while the mobility due to optical phonon scattering only is expected to be proportional to $T^{-1/2}$.

The mobility may be observed experimentally and note that at low temperature for heavily doped materials that it deviates from the simple $T^{3/2}$ power law. This is because at low temperatures the impurity scattering dominates. The transition may be seen clearly in Figure 21. The form of the transition is shown schematically in Figure 22.



Figure 23: Optical and acoustic modes for phonons in a lattice.

33



Figure 24: Mobility electrons and holes as a function of donor and acceptor density. Note the electron mobility is higher than that of the holes for equivalent doping.

The probability of a collision taking place depends on the reciprocal of the mean free time between collisions i.e.

$$\frac{1}{\tau_c} = \frac{1}{\tau_{phonon}} + \frac{1}{\tau_{impurity}}$$
(2.6)

Or equivalently

$$\frac{1}{\mu} = \frac{1}{\mu_{phonon}} + \frac{1}{\mu_{impurity}}$$
(2.7)

For a fixed temperature the mobility of both electrons and holes may be studied in more detail as a function of doping. The mobility reaches a maximum when the lattice is very pure and the scattering is entirely dominated by phonon interactions. At high impurity levels the mobility reaches a minimum value, or saturates.

2.1.2 Resistivity

When an electron moves in the presence of an electric field, \mathcal{E} the electron experiences a force that is proportional to the strength of the electric field. The force is equal to the negative gradient of the potential energy *Ei*. In 1D

$$-q\boldsymbol{\mathcal{E}} = -\frac{dE_i}{dx} \tag{2.8}$$

With the electric field there is defined a potential which has the property

$$\mathcal{E} = -\frac{d\psi}{dx} \tag{2.9}$$

In this scheme the electrostatic potential and the potential energy of the electron are simply related by

$$\psi = -\frac{E_i}{q} \tag{2.10}$$

The resistivity of semiconductors may be very simply related to the mobility of the carriers. For electrons the current density J_n flowing in the crystal can be found by multiplying the electron density in the material by the drift velocity

$$J_n = -qnv_n = qn\mu_n \mathcal{E}$$
(2.11)

Where ${m {\cal E}}$ is the local electric field. Similarly for holes

$$J_{p} = -qpv_{p} = qp\mu_{p}\boldsymbol{\mathcal{E}}$$
(2.12)

If both holes and electrons are present

$$J = J_n + J_p = (qn\mu_n + qp\mu_p)\varepsilon$$
(2.13)

The conductivity is

$$\sigma = (qn\mu_n + qp\mu_p) \tag{2.14}$$

where the resistivity is the reciprocal of the conductivity i.e.

$$\rho = \left(\frac{1}{qn\mu_n + qp\mu_p}\right) \tag{2.15}$$

For extrinsic semiconductors it is usual that there is a much higher density of the majority carrier i.e. $\rho = 1/qn\mu_n$ for n-type Si and $\rho = 1/qp\mu_n$ for p-type.



Figure 25; Resistivity of Si for various Doping concentrations

2.1.3 Hall Effect

The Hall effect is used to measure the carrier concentration directly in a semiconductor. The basic setup is shown below.



Figure 26:Schematic of setup used to demonstrate the Hall Effect

For a hole moving to the right (x-direction) the Lorentz force

$$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$$

tends to accelerate the carrier upwards if the magnetic field is oriented along the +'ve z direction. As current cannot flow out of the sample in the y direction holes will accumulate in the y direction and eventually an electric field will be established which precisely balances out the Lorentz force

$$q\boldsymbol{\mathcal{E}}_{y} = q\boldsymbol{v}_{x}\boldsymbol{B}_{z} \tag{2.16}$$

By measuring the *Hall voltage* across the detector in the y-direction, where the thickness is W, we can calculate the field

$$\boldsymbol{\mathcal{E}}_{v} = V_{H} / W$$

Substituting for the drift velocity in terms of the current density in the x-direction we obtain

$$\varepsilon_{y} = \left(\frac{1}{qp}\right) J_{p} B_{z} = R_{H} J_{p} B_{z}$$
(2.17)

Where the Hall coefficient

$$R_H = 1/qp \tag{2.18}$$

For an n-type semiconductor

$$R_{H} = -1/qn \tag{2.19}$$

If we know the current drawn and the magnetic field we can extract

$$p = \frac{1}{qR_{H}} = \frac{J_{p}B_{z}}{q\mathcal{E}_{y}} = \frac{(I/A)B_{z}}{q(V_{H}/W)} = \frac{IB_{a}W}{qV_{H}A}$$
(2.20)

where all the terms may be measured. This enables us to evaluate the carrier density (and type) for a particular sample of semiconductor.

2.2 Carrier Diffusion

All our calculations to this point have been based on the movement of charges under the influence of electric (or magnetic) fields. There is another process that
gives rise to the movement of carriers, and that is diffusion. In diffusion carriers tend to move from a region of high carrier concentration towards a region of low carrier concentration. Any currents due to this carrier movement are known as



diffusion currents.

Figure 27 One dimension example the diffusion of carriers. The number N flows from region of higher density to lower. For electrons the current is on the opposite direction

The diffusion arises from the stochastic motion of carriers (at the same temperature) between to adjacent, but differently doped samples. If both slices are of thickness 1, which is small and approximately equal to the mean free path, and the doping concentration is varying smoothly and then if on average (due to thermal excitations) the velocity in the +x direction for the n(x) sample is v_x^{th} (an equal number are leaving to the left) then the number N (related to the charge current Jn)crossing the unit plane (to the right) between the two samples per time interval $dt = l/v_x^{th}$

Must be given by

$$N = \frac{J_n}{-qn} = (n(x)l - n(x+l)l)v_{th}^x / l \sim -dn / dx l v_{th}^x$$
(2.21)

Another way of writing this equation is

$$J_n = qD_n dn / dx \tag{2.22}$$

Where D_n is the diffusion coefficient. This is a one dimension example of Fick's law. In 3D we may write which state that

$$\mathbf{J}_n = q D_n \nabla n \tag{2.23}$$

2.2.1 Einstein Relation

It is sometimes convenient to express the diffusion coefficient in terms of the mobility of the of the carriers. We note that from Eq. (2.21) that

$$D_n = v_{th}^x l \tag{2.24}$$

.Now from equipartition of energy we know that

$$\frac{1}{2}m_n \left(v_{th}^x\right)^2 = 1/2kT$$
(2.25)

Thus using our expression for the mean free path in terms of its mobility we find that

$$D_n = v_{th}^x v_{th}^x \tau_c = \frac{kT}{m_n} \frac{m_n \mu_n}{q} = \frac{kT}{q} \mu_n$$
(2.26)

Which is known as the Einstein relation. This relates diffusivity and mobility.

2.2.2 Current Density Relation

When an electric field is present in addition to a density gradient we can write

$$J_{n} = q\mu_{n}n\mathcal{E} + qD_{n}\frac{dn}{dx}$$

$$J_{p} = q\mu_{p}p\mathcal{E} - qD_{p}\frac{dp}{dx}$$
(2.27)

Note that the expression for the hole current contains a negative sign as for a positive hole gradient, holes flow min the -x direction.

2.3 Generation/Recombination

A very important property of a semiconductor is its behaviour when exposed to ionizing radiation which, in general, gives rise to electron hole pairs, or the generation of carriers. Other phenomena such as the transfer of energy from the lattice to create an electron-hole pair is also a common source of carrier generation. The recombination of carriers, in which electron-holes pairs "annihilate" is a separate but equally important phenomenon. The recombination may take several forms. The energy that is released when the electron-hole pair recombine may manifest itself either as a photon or energy transferred directly to the lattice (as heat). When the photon is emitted the process is a radiative recombination.

A further distinction is drawn between the recombination phenomena in direct and indirect semiconductors (see section 1.3). For direct semiconductors direct recombination dominates whereas in indirect semiconductors (e.g. Si) indirect recombination dominates.

2.3.1 Direct Recombination

In direct semiconductors the recombination process is relatively easy to understand. An electron makes a transition from the conduction band down to the valence band annihilating a hole state. It will be relatively easy for them to annihilate as there is not need to invoke an interaction with the crystal lattice in order to conserve energy and momentum.

2.3.2 Indirect Recombination.

In indirect the recombination the situation is substantially more complicated. The direct recombination process is heavily suppressed due to the difficulty of simultaneously transferring the appropriate amount of energy and momentum to the crystal to conserve these quantities. The dominant recombination process has then to proceed via a localized energy state, often in the bandgap. These states are called recombination centres or traps.

The prediction of the behaviour of these recombination centres is unfortunately quite messy, but as it has an vital bearing on the radiation properties of Si sensors we are obligated to make an attempt to describe the underlying processes. Following Sze's prescription we note that there are four basic processes that can take place from the intermediate states to the conduction and valence bands. Capture and emission of holes and capture and emission of electrons. (see Figure 28)



Figure 28: Four processes involved in recombination in an indirect semiconductor. The recombination centre has a single energy level.

Let us assume that the electron capture centre have a density of N_t in the semiconductor. Let us further assume that the capture centres may only be occupied by one electron at a time. We know the Fermi-Dirac distribution F_{FD} so the rate of capture R_a must be proportional to the number of unoccupied centres $(1-F_{FD})$

$$R_a \propto nN_t (1 - F_{FD}) \tag{2.28}$$

We allow the constant of proportionality to be represented by a capture crosssection times the thermal velocity

$$R_a = v_{th} \sigma_n n N_t (1 - F_{FD}) \tag{2.29}$$

The emission probability of electrons of the electrons for the capture centres will be proportional to the number of occupied states. As in thermal equilibrium the number of captured electrons must equal the number of emitted electrons this gives

$$e_n = v_{th} \sigma_n n_i e^{(E_t - E_i)/kT}$$
(2.30)

Where we have used Eq. (1.49). Exactly the same process may be followed for hole capture and emission yielding

$$R_c = v_{th} \sigma_p p N_t F_{FD} \tag{2.31}$$

and

$$R_d = e_p N_t (1 - F_{FD})$$
(2.32)

The equilibrium conditions yield

$$e_p = v_{th} \sigma_p n_i e^{(E_i - E_t)/kT}$$
(2.33)

The capture cross-sections, energy levels and the concentrations levels may all be measured experimentally. Note in the equilibrium case with no external injection of charge $R_a=R_b$ and $R_c=R_d$. If extra charge is being generated, through a mechanism such as illumination, a number of electrons and holes will be generated per second G_L .

Under these conditions as the numbers of electrons and holes will still be constant but now

$$G_{L} - (R_{a} - R_{b}) = 0$$

$$G_{L} - (R_{c} - R_{d}) = 0$$
(2.34)

We can this eliminate our need to know both GL and in fact eliminate FFD yielding a net recombination rate U as

$$U \equiv R_a - R_b = \frac{v_{th}\sigma_n\sigma_p N_t (pn - n_i^2)}{\sigma_p [p + n_i \exp(E_i - E_t)/kT] + \sigma_n [n + n_i \exp(E_t - E_i)/kT]}$$
(2.35)

Note that under these conditions the product pn may be $> n_i^2$ as the our mass action law does not account for the injection of charges into the material.

In a simple case where we have strongly n-type material then $n \gg p$. Also if we assume that the system is such that the centres are close to the centre of the bandgap then (for strong doping)

$$n \gg n_i e^{(E_t - E_i)/kT}$$

Under these conditions we see that if the material is in a low charge injection mode $(n \sim no)$ where *no* is the thermal equilibrium condition

$$U \sim v_{th} \sigma_p N_t (p - n_i^2 / n) = v_{th} \sigma_p N_t (p - p_o)$$
(2.36)

This tells us the recombination rate is dominated by the minority (holes) carriers and the typical lifetime is given by

$$\tau_p = \frac{1}{v_{th}\sigma_p N_t} \tag{2.37}$$

Thus we get faster recombination in a hotter system, with "larger" trapping cross-section and more recombination centres.

A similar expression is obtained in a p-type semiconductor. Note that in n-type silicon $\tau_p=0.3\mu$ s and in p-type Si 1.0 μ s. Thus the recombination lifetime echoes the higher mobility of the electrons.

If the cross-section for electron and hole capture are the same more general forms for the recombination rate may be calculated. (See Sze).

$$U = v_{th} \sigma N_t \frac{(pn - n_i)^2}{p + n + 2n_i \cosh \frac{E_t - E_i}{kT}}$$
(2.38)

Under these conditions the recombination lifetime for n-type Si is given by

$$\tau_r = \frac{1 + \left(\frac{2n_i}{n_o + p_o}\right) \cosh\left(\frac{E_t - E_i}{kT}\right)}{v_{th}\sigma N_t}$$
(2.39)



Figure 29: Recombination and generation lifetime versus the energy level of the centres

In a reverse biased diode carriers are generated from the generation/recombination centres. Under the condition that p < ni and n < ni we find the generation lifetime in n-type Si has the form

$$\tau_{g} = \frac{2\cosh\left(\frac{E_{t} - E_{i}}{kT}\right)}{v_{th}\sigma N_{t}}$$
(2.40)

Comparison of the generation and recombination lifetimes shows that (Figure 29) the generation may be much slower than the recombination especially if Et is far away from Ei.

For most of our applications for sensor design the above argument will be especially important when we consider impurities, either added during production time, or induced during irradiation. For example "poisoning" a silicon sensor with gold can reduce the minority carrier lifetime by many orders of magnitude from about 1.0μ s to 0.1ns.

Irradiation, to which we will return in Part 3 of these notes, has the effect of introducing mid-bandgap energy levels through damage to the lattice. These will become recombination and generation centres. Note the electron irradiation does have a different effect to neutron irradiation. Neutron irradiation creates acceptor levels very close to the effective Fermi level, whereas electrons create energy bands just above the valence band.

2.4 Continuity Equation

The movement of charge within a semiconductor is governed by the continuity equation. As charge must be conserved the amount of charge flowing into an infinitesimal box must equal the amount flowing out corrected by the amount of combination and generation taking place.

In 1D the derivation of the continuity equation is relatively simple.



Figure 30: Diagram showing the 1D flow of current in an infinitesimal slice of semiconductor with both generation and recombination taking place.

If the amount of charge flowing in (out) through an area A per unit time is given by J then we have an infinitesimal slice containing n electrons

$$\frac{J_n(x)A}{-q} + \frac{J_n(x+dx)A}{-q} + (G_n - R_n)Adx = \frac{\partial n}{\partial t}Adx$$
(2.41)

where G,R are the generation and recombination densities and n is the number density if the carriers

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G_n - R_n)$$
(2.42)

where there is also an equivalent expression for holes given by

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + (G_n - R_n)$$
(2.43)

These are the continuity equations in 1D. In 3D the continuity equations are:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla J_n + (G_n - R_n)$$

$$\frac{\partial p}{\partial t} = \frac{1}{q} \nabla J_p + (G_p - R_p)$$
(2.44)

In 1D substituting for J from Eq. (2.27) we find

$$\frac{\partial n_p}{\partial t} = n_p \mu_n \frac{\partial \boldsymbol{\mathcal{E}}}{\partial x} + \mu_n \boldsymbol{\mathcal{E}} \frac{\partial n_p}{\partial x} + D_n \frac{\partial^2 n_p}{\partial x^2} + G_n - \frac{(n_p - n_{p0})}{\tau_n}$$

$$\frac{\partial p_n}{\partial t} = -p_n \mu_p \frac{\partial \boldsymbol{\mathcal{E}}}{\partial x} - \mu_p \boldsymbol{\mathcal{E}} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} + G_p - \frac{(p_n - p_{n0})}{\tau_p}$$
(2.45)

2.5 High Field Effects

2.5.1 Velocity Saturation

One of the most important high electric field effects in a semiconductor is the relationship between the mobility and the field. As the field is increased from low values the mobility (velocity) increase linearly. However as the carrier energy increases beyond the optical phonon energy the probability of the carrier emitting an optical phonon increases dramatically. This mechanism causes the carrier velocity to saturate with increasing electric field. For carriers in silicon and other materials which do not contain accessible higher bands, the velocity versus field relation can be described by:

$$v(E) = \frac{\mu E}{1 + \frac{\mu E}{v_{sat}}}$$
(2.46)

The maximum obtainable velocity, *vsat*, is referred to as the saturation velocity.



Figure 31 Drift velocity of electrons in Si for different temperatures as a function of applied field.



Figure 32:Drift velocity of holes in Si for different temperatures as a function of applied field.



Figure 33: Optical phonon interaction as a function of energy. Note the sharp turn on of the scattering cross section.

2.5.2 Avalanche Mode

If the electric field is sufficiently high then the carriers (electrons or holes) will have enough energy to create additional electron hole pairs. For example a highly energetic electron in n-type Si in the conduction band may have enough energy to interact with an electron in the valence band creating another conduction band electron and a hole in the valence band. The new pair of carriers will themselves begin to accelerate in the applied field. If the field is high enough then this pair will themselves create other electron-holes pairs triggering an "avalanche". The number of electron-holes pairs generated by an electron per unit distance travelled is called the *ionization rate* for the electron, α_n . The same definition holds for holes. Avalanches are typified by values of α of order 10⁴/cm or greater.



Figure 34: Ionization rate versus applied field

2.6 Summary of Properties of Si

Below we include a table of some the most important properties of Si.

Atoms/cm	5.0×10^{22}
Atomic Weight	28.09
Breakdown field (V/cm)	~3×10 ⁵
Crystal structure	Diamond
density (g/cm ³)	2.328
Dielectric Constant	11.9
$N_{c} (cm^{-3})$	2.8×10 ¹⁹

INTRODUCTION TO SEMICONDUCTORS

$N_{v} (cm^{-3})$	1.04×10 ¹⁹
Effective Mass, m*/m0 Electrons	
m_l	0.98
m* ₁	0.19
Holes	
m_{eh}^{*}	0.16
m_{hh}^{*}	0.49
Electron affinity, x(V)	4.05
Energy gap (eV) at 300K	1.12
Intrinsic carrier conc. (cm ⁻³)	1.45×10^{10}
Intrinsic Debye Length (µm)	24
Intrinsic resistivity (-cm)	2.3×10 ⁵
Lattice constant (Å)	5.43095
Linear coefficient of thermal expansion, L/LT(C ⁻¹)	2.6×10 ⁻⁶
Melting point (C)	1415
Minority carrier lifetime (s)	2.5×10 ⁻³
Mobility (drift) (cm ² /V)	1500 450
Optical-phonon energy (eV)	0.063
Phonon mean free path (Å)	76(electron) 55(hole)
Specific heat (J/g C)	0.7
Thermal conductivity (W/cmC)	1.5
Thermal diffusivity (cm ² /s)	0.9
Vapour pressure (Pa)	1 at 1650C 1× ¹⁰ -6 at 900 C

Table 3:Summary of the Properties of Si

References

- 1 Jacoboni and Reggiani, Rev Mod. Phys, Vol55, No3, 1983
- 2 http://www.ioffe.rssi.ru/SVA/NSM/Semicond/Si/electric.html
- 3 C. Kittel, Introduction to Solid State Physics, Fourth Edition, John Wiley and Sons
- 4 S.M.Sze, Semiconductor Devices, John Wiley and Sons
- 5 Principles of Semiconductor Devices, B.v. Zeghbroeck ,http://ecewww.colorado.edu/~bart/book/contents.htm

Problems

- 1 Find the room temperature resistivity of n-type silicon doped with 1015 phosphorous atoms/cm3
- 2 Fin the electron and hole concentration, mobility and resistivity of Si at 300K when doped with 1016 Boron atoms/cm3.



P-N JUNCTION

The most important semiconductor devices in particle physics over the last two decades have been based on the reversed biased p-n junction. A simple model of the behaviour of the p-n junction is provided. The creation of a depletion region, which forms the basis of many particle physics devices, is discussed For completeness a description of bipolar transistor action is given. The presentation closely follows that in Sze.

3.1 Thermal Equilibrium

The basis of the p-n is the joining of two pieces of doped Si, one predominantly p-type and the other n-type. There are a variety of different doping schemes and approximation. Below its is assumed that the system

under consideration S composed of uniformly doped *p*-type and *n*-type materials. In most cases a voltage will applied to the p-n junction. A forward biased junction is one where a +'ve potential is applied to the "anode" (pside) compared with the "cathode" (n-side). In this configuration the most obvious and important feature of the junction is manifest. Forward biased a large current may be passed, reverse biased very little current is passed. It is this rectifying feature of the junction that makes it so



Figure 35: Schematic of p-n junction

useful.

It is instructive to imagine that the two types of semiconductor can be brought together without an external voltage being applied. Before joining the Fermi levels of each of the samples remains the same until contact is made. On contact note that the density of free electrons is higher in the n-type side and the density of holes higher on the p-type side, Figure 36b.



Figure 36: The p-n junction before being joined (a) and before reaching thermal equilibrium (b).

Of course on "contact" at the metallurgical junction neither the n-type nor p-type sides are in thermal equilibrium as electrons will diffuse to the p-type side and holes to the n-type side.



Figure 37: p-n junction in thermal equilibrium without h external applied voltage. The positive Energy axis corresponds to increasing electron energy.

As the electrons diffuse towards the *p*-side and holes diffuse the *n*-side they leave behind ionized donors or acceptors. On the *n*-side a region of depth x_n

develops which is depleted of electrons. On the *p*-side a region of depth x_p develops which is depleted of holes.

As the electrons drift towards the p-side leaving behind a positive net charge electrons will be attracted from the p-side towards the n-side i.e. a field is set up so the drift of electrons is towards the n-side. The holes on the p-side also diffuse so as to set up a field that causes drift of holes towards the p-side.

The movement continues until the material reaches a thermal equilibrium. In thermal equilibrium the Fermi Energy is uniform within the material. Under this condition an internal potential φ_i is set up as shown in Figure 37. The potential φ_i is sometimes called the built in potential, and labelled V_{bi} .

The equalizing of the Fermi energy is easy to derive. In thermal equilibrium the drift current (of electrons or holes) must equal the diffusion current (of electrons or holes). For holes

$$J_{p} = J_{p}(\operatorname{drift}) + J_{p}(\operatorname{diffusion}) = q\mu_{p} p \mathcal{E} - qD_{p} \frac{dp}{dx}$$

$$= q\mu_{p} p \frac{1}{a} \frac{dE_{i}}{dx} - kT\mu_{p} \frac{dp}{dx} = 0$$
(3.1)

Using the expression for the concentration of holes (Eq. (1.50))

$$p = n_{i} e^{(E_{i} - E_{F})/kT}$$
(3.2)

We note

$$\frac{dp}{dx} = \frac{n_i}{kT} e^{(E_i - E_F)/kT} \left(\frac{dE_i}{dx} - \frac{dE_F}{dx}\right) = \frac{p}{kT} \left(\frac{dE_i}{dx} - \frac{dE_F}{dx}\right)$$
(3.3)

Thus

$$J_{p} = q\mu_{p}p\frac{1}{q}\frac{dE_{i}}{dx} - kT\mu_{p}\frac{dp}{dx} =$$

$$q\mu_{p}p\frac{1}{q}\frac{dE_{i}}{dx} - kT\mu_{p}\frac{p}{kT}(\frac{dE_{i}}{dx} - \frac{dE_{F}}{dx}) = 0$$
(3.4)

As the first two terms cancel each other this means that

$$J_p = \mu_p p \frac{dE_F}{dx} = 0 \tag{3.5}$$

Exactly the same derivation holes for the holes leading to

$$J_n = \mu_n n \frac{dE_F}{dx} = 0 \tag{3.6}$$

This in thermal equilibrium (for both holes and electron) the Fermi energy is equalized. See again Figure 37.

To extract the built in potential it is necessary to solve Poisson's equation in the material. Poisson's equation relates the potential field to the charge density within the material. In 1D assuming all acceptors and donors are ionized:

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon} \left(N_D - N_A + p - n \right) \tag{3.7}$$

In the region far away from the metallurgical junction it is reasonable to assume that the potential is not changing i.e. the system has retained electrical neutrality.

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon} \left(N_D - N_A + p - n \right) = 0 \tag{3.8}$$

Considering the *p*-type volume, if ND=0 and p>>n then NA=p. Thus again using Eq. (1.50)

$$N_{A} = n_{i} e^{(E_{i} - E_{F})/kT}$$
(3.9)

So

$$E_i - E_F = kT \ln \frac{N_A}{n_i} \tag{3.10}$$

From Eq.(2.10) we deduce that well away from the metallurgical junction in the p-type neutral region

$$\Psi_p = -\frac{1}{q} (E_i - E_F) = -\frac{kT}{q} \ln \frac{N_A}{n_i}$$
 (3.11)

And likewise in the *n*-type neutral region

$$\psi_n = \frac{1}{q} \left(E_i - E_F \right) = \frac{kT}{q} \ln \frac{N_D}{n_i} \tag{3.12}$$

Thus subtracting the two potentials, the built in potential becomes

$$\phi_{i} = V_{bi} = \frac{kT}{q} \ln \frac{N_{A}N_{D}}{n_{i}^{2}}$$
(3.13)

Note the p-side has a higher electron energy therefore the built in potential is +'ve towards the n-side. For typical Si devices the built in potential is of order 0.6V.

In the region close to the metallurgical junction the holes and electrons which are normally present have been neutralized by the diffusion of the opposite sign carriers. This depletion region is of extreme interest for those building particle physics sensors. In between the neutral region and the depletion region are narrow transition zones which, for the moment, we shall ignore.

3.2 Depletion Region

For a known doping profile Poisson's equation may be solved. The simplest, and most important example is where it may be assumed the doping profile is constant for both n and p-type materials up to the metallurgical junction at x=0.

3.2.1 Abrupt Junction

In this region

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon} \left(N_D - N_A \right) \tag{3.14}$$

as both *p* and *n* are neutralized. In the *p*-type region if the depletion width is between x = xp and 0:

$$\frac{d^2\psi}{dx^2} = \frac{q}{\varepsilon} (N_A) \text{ for } -x_p \le x \le 0$$
(3.15)

and in the *n*-type region between x=0 and x=xn

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon} (N_D) \text{ for } 0 \le x \le x_n$$
(3.16)

The charge density in these two regions is dominated by that of the fully ionized donors (+'ve charge) and the fully ionized acceptors (-ve charge).



Figure 38: Diagram of the volume round a 1D metallurgical junction showing the distribution of charges and the electric field.

Note also that as the material must retain its overall electrical neutrality thus:

$$N_A x_p = N_D x_n \tag{3.17}$$

The total depletion region is assumed to have a width

$$W = x_n + x_p \tag{3.18}$$

Thus we obtain the electric fields in both the p region and n region

$$\mathcal{E}(x) = -\frac{qN_A(x+x_p)}{\varepsilon} \quad \text{for} \ -x_p \le x \le 0 \tag{3.19}$$

and

$$\boldsymbol{\mathcal{E}}(x) = \frac{qN_D(x - x_n)}{\varepsilon} \quad \text{for } 0 \le x \le x_n \tag{3.20}$$

The maximum field that is attained \mathcal{E}_m is given by

$$\mathcal{E}_m = \frac{qN_D x_n}{\varepsilon} = \frac{qN_A x_p}{\varepsilon}$$
(3.21)

These equations may be integrated to yield the total potential

$$\phi_i = V_{bi} = \frac{qN_A x_p^2}{2\varepsilon} + \frac{qN_D x_n^2}{2\varepsilon} = \frac{1}{2} \boldsymbol{\mathcal{E}}_m W$$
(3.22)

This equation may be solved for W whilst eliminating xp and xn explicitly by noting

$$N_A x_p = N_D x_n$$

so

$$V_{bi} = \frac{qN_A x_p^2}{2\varepsilon} + \frac{qN_D x_n^2}{2\varepsilon} = \frac{q}{2\varepsilon} \left(N_A x_p^2 + \frac{N_A^2 x_p^2}{N_D} \right)$$

$$= \frac{q}{2\varepsilon} \left(\frac{N_A N_D + N_A^2}{N_D} \right) x_p^2$$
(3.23)

i.e.

$$\left|x_{p}\right| = \sqrt{\frac{2\varepsilon}{q} \frac{N_{D}}{N_{A}(N_{D} + N_{A})}} V_{bi}$$
(3.24)

and

$$\left|x_{n}\right| = \sqrt{\frac{2\varepsilon}{q}} \frac{N_{A}}{N_{D}(N_{D} + N_{A})} V_{bi}}$$
(3.25)

so

$$W = \sqrt{\frac{2\varepsilon}{q} \left(\frac{N_A + N_D}{N_A N_D}\right)} V_{bi}$$
(3.26)

If the material is strongly doped on one side e.g. p+n ($N_A >> N_D$) then the depletion then the depletion layer on the *p*-side is very narrow

$$W \sim \sqrt{\frac{2\varepsilon V_{bi}}{qN_D}} \tag{3.27}$$

In this case the width of the depletion zone depends on the lighter doping of the "bulk". Note also the dependence on the depletion depth on the square root of the internal potential.



Figure 39: Diagram showing the effect of reverse and forward biasing a pn junction

A positive potential (forward bias - Vf) applied to the anode (the p-side in this case) reduces the electrostatic potential across the metallurgical junction and we can repeat the same analysis replacing V_{bi} by V_{bi} - V_f . For the heavily doped example just discussed if N_D or N_A (the bulk doping) is denoted by N_B then

$$W = \sqrt{\frac{2\varepsilon(V_{bi} - V_f)}{qN_B}}$$
(3.28)

Which is one of the most important equations for Si detector construction.

For a reverse biased junction where the applied voltage, V_{bias} , is much greater than V_{bi} we obtain

$$W = \sqrt{\frac{2\varepsilon V_{bias}}{qN_B}} \tag{3.29}$$

3.3 Junction Capacitance

The capacitance of any system is defined by the increase (or decrease) dQ due to a small change in an applied voltage, dV. For the *p*-*n* junction if we increase the charge on either side of the depletion zone then we know that the electric field depends on the total charge through Equations (3.19) and (3.20) as

$$d\mathcal{E} = dQ/\varepsilon$$

The approximate change in the voltage is thus

$$dV \sim W dQ / \varepsilon \tag{3.30}$$

i.e. the capacitance per unit are is given by

$$C \equiv \frac{dQ}{dV} \sim \frac{dQ}{WdQ/\varepsilon} = \varepsilon/W \tag{3.31}$$

This is identical to the formula for the capacitance of parallel plate separated by a distance *W*.

For a reverse biased diode we then obtain

$$C^2 \sim \frac{qN_B}{2V_{bias}} \tag{3.32}$$

In a practice the $1/C^2$ of a diode (as measured with a capacitance bridge or meter) may be measured as a function of the applied voltage. This should be linear until W exceeds the thickness of the diode at the full depletion voltage. The slope of the linear region may be used to directly infer N_B .



Figure 40: Capacitance versus voltage for an ATLAS (strip) diode. This device is depleted at about 50-60V.

3.4 Current-Voltage Properties

The rectifying behaviour of the p-n junction which is evident from studying simple p-n devices has a relatively simple source. Assuming that the junction may be approximated by:

- A system in which outside of the depletion zone the material is neutral
- The depletion zone has sharp boundaries i.e. the transition regions are very small
- The majority carrier density is much larger than the minority carrier currents
- No recombination nor generation is taking place in the depleted zone i.e. all the currents are due to diffusion
- The electron and holes currents are uniform throughout depletion zone

It is possible to calculate the current through the p-n junction as a function of applied voltage. Here we follow the derivation by Sze very closely.

3.4.1 Ideal Characteristics

Under the ideal conditions mentioned above we introduce the following nomenclature

• *nn0* is the equilibrium distribution of electrons in the n-side

- *np0* is the equilibrium distribution of holes in the p side
- *pn0* is the equilibrium distribution of holes in the n-side
- *pp0* is the equilibrium distribution of holes in the n-side

In terms of these variables the built in potential Eq. (3.13) becomes:

$$V_{bi} = \frac{kT}{q} \ln \frac{p_{p0} n_{n0}}{n_i^2} = \frac{kT}{q} \ln \frac{n_{n0}}{n_{p0}}$$
(3.33)

where we have used the mass action law. From this we obtain

$$n_{n0} = n_{p0} e^{qV_{bu}/kT} \tag{3.34}$$

and similarly

$$p_{p0} = p_{n0} e^{qV_{bu} / kT}$$
(3.35)

In the case we apply a forward bias V the potential difference is reduced, when a reverse bias is applied the potential difference is increased. We write for the general case

$$n_n = n_p e^{q(V_{bi} - V)/kT}$$
(3.36)

where we have replaced the equilibrium densities with non-equilibrium values. If there is little current then the minority carrier density is much smaller than the majority carrier density thus $n_n \sim n_{n0}$. Thus

$$n_p e^{q(V_{bi} - V)/kT} = n_{p0} e^{qV_{bi}/kT}$$
(3.37)

i.e at x = -xp

$$n_p = n_{p0} e^{qV/kT} \tag{3.38}$$

Similarly we get at x = xn

$$p_n = p_{n0} e^{qV/kT} \tag{3.39}$$

at the n-side boundary of the junction. These distributions are shown in Figure 41. As no current is being generated within the depletion zone, this is the ideal approximation, all currents must come from the neutral regions. Using Eq. (2.45) we obtain for the density of holes in the n-side

$$\frac{d^2 p_n}{dx^2} - \left(\frac{p_n - p_{n0}}{D_p \tau_p}\right) = 0$$
(3.40)

This can be solved assuming that as we get very far away from the metallurgical junction that $p_n = p_{n0}$ as

$$p_n - p_{n0} = p_{n0} \left(e^{qV/kT} - 1 \right) e^{1(x - x_n)/L_p}$$
(3.41)



Figure 41: Energy bands and carrier distribution in ideal p-n junction for a) forward biased and b) reverse biased case.

Where

$$L_p = \sqrt{D_p \tau_p}$$

is the diffusion length of holes in the *n*-region. Similarly

$$n_p - n_{p0} = n_{p0} \left(e^{qV/kT} - 1 \right) e^{(x + x_p)/L_n}$$
(3.42)

In the neutral *p*-region. From these equations we can calculate the currents at xn and -xp

$$J_{p}(x_{n}) = -qD_{p} \frac{dp_{n}}{dx} \bigg|_{x=x_{n}} = \frac{qD_{p}p_{n0}}{L_{p}} \left(e^{qV/kT} - 1\right)$$

$$J_{n}(-x_{p}) = qD_{n} \frac{dn_{p}}{dx} \bigg|_{x=-x_{p}} = \frac{qD_{n}n_{p0}}{L_{n}} \left(e^{qV/kT} - 1\right)$$
(3.43)

Thus the total diffusion current through the junction (and the device) is

$$J = J_s \left(e^{qV/kT} - 1 \right)$$

$$J_s \equiv \frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n}$$
(3.44)

This gives rise to the classic diode (rectifying behaviour) of the p-n junction as shown in Figure 42. Typically one of terms in the diffusion current equation will

dominate. However note that the current on the p-side is controlled by the electron density (np0) and the current on the n-side is controlled by the holes i.e. the control is by the *minority* carriers on either side of the junction.



Figure 42: Current in an ideal diode as a function of bias

3.4.2 Temperature effects

For Si based devices the ideal diode equation falls short of accurately describing reality. This is due to the fact that in (for example) a p^+n Si semiconductor the diffusion current which is approximately given by

$$J = J_s \left(e^{qV/kT} - 1 \right)$$

$$J_s \equiv \frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n}$$
(3.45)

may not be very much greater then the rate of generation inside the depletion zone (width W)

$$J_G \sim \frac{q n_i W}{\tau_g} \tag{3.46}$$

In general for the reverse bias (which we are most interested in)

$$\frac{J_{diff}}{J_G} = \frac{n_i L_p}{N_D W} \frac{\tau_g}{\tau_p}$$
(3.47)

The term varying the fastest with temperature is ni and at very low temperatures it may be possible, as the diffusion freezes out, for the current to be dominated by the generated carriers

3.5 Breakdown

If a large reverse voltage is applied to the junction then, as the magnitude of voltage is increased, the junction will eventually breakdown. There are two mechanisms, tunnelling and avalanche breakdown.

3.5.1 Tunnelling

In tunnelling a valence electron may make the jump to the conduction band through the bandgap. This is similar to the classical quantum mechanical effect of a particle tunnelling through a potential barrier. Typically the fields required for tunnelling to occur in a *p*-*n* junction are enormous – over 10^7 V/m. At the these fields the electrons (or holes) will pick up an energy many times the bandgap between collisions. This process was first suggested by Zener. For a normal diode, such as one can find in the lab, the tunnelling process is accompanied by a break down at a voltage of order 5V.

3.5.2 Avalanche

If sufficient energy is put into the electron then it may have enough energy to break lattice bonds and create an electron hole pair. This electron hole pair may, instead of recombining with the ionized atom, be accelerated and in turn create further electron hole pairs. This process is the avalanche breakdown.

This process is very similar to that in a gaseous material. If the "gain", the number of electron-holes pairs created per unit length, is sufficiently high the avalanche process will proceed and amplify the original electron. The avalanche process dominates in diodes that have a breakdown voltage of more than about \sim 7V.

Although it is possible to calculate the form of behaviour for both processes it is not particularly instructive. However what is important is to see the level of the doping and the field necessary to produce breakdown for reverse biased Si (see Figure 43)),



Figure 43: breakdown voltage for Si and GaAs as a function of the background doping.

For diodes with different radii of curvature of the implant (many of these are used in particle physics applications) a typical breakdown voltage diagram versus background doping is shown below.



Figure 44: Breakdown voltage versus impurity and geometry

3.6 Bipolar Transistor Action

The *pn* junction can be used to build one of the transforming technological devices of the last century. The transistor. First operated in 1940's at Bell Labs, the first transistor was made of germanium with metal contacts. The transistors discussed below are composed of two back to back *pn* diodes in an *npn* or *pnp* configuration. There are many closely related devices, for example the Thyristor, which will not be described here. The reader is recommended to consult Sze for more details on this and similar devices.



Figure 45: The first transistor (Bell Labs)

A simple one dimensional view of a pnp transistor is shown below in Figure 46. The idealized view contains a heavily doped *p*-region (p+) which is in the diagram below to the left of the base *n*-type material to which a contact may be made. The p+ implant is known as the emitter. The other contact is a more lightly doped *p*-volume called the collector. In this idealized case all the volumes are assumed uniformly doped. In the configuration shown the emitter-base is forward biased 3 whereas the collector current is reverse biased.

In thermal equilibrium where we apply no voltage across either the emitter-base or the base-collector the terminals of the transistor are effectively shorted. As in the case of the p-n junction under thermal equilibrium the Fermi energies are equalized, see Figure 47.

If voltage is now applied so as to forward bias the emitter-base and reverse bias the collector-base we it can be shown the transistor will behave as a rudimentary amplifier.



Figure 46: Idealized view of the pnp transistor

Under the application of forward bias the depletion width in the emitter-base will be narrower than the depletion width in the collector-base. As the emitter-base is forward biased holes are emitted from the p+ emitter into the base and electrons injected into the emitter from the (n) base. The reversed biased collector-base junction will only be passing a small saturated reverse bias current.

If the emitter-base junction is sufficiently narrow/small (order of a micron) the holes from the emitter may diffuse through the base directly into the collector with very little recombination. In this case the collector hole current will be very close to the emitter hole current. Any of the few holes arriving onto the base will be neutralized by a current of electrons into base. As we know from a forward biased diode, a small change in the base-emitter voltage will cause a large change in the emitter hole current. This large current flows onto the collector but as the collector-base is reverse biased a small change in the base-collector voltage makes little difference to the current.

Although this may not at first sight look like amplifying behaviour it most certainly is. What has been achieved is the ability to drive a large current without loading the base/collector circuit. Thus if we need to amplify the input voltage on the emitter all we need to do is to put a large resistor in series with the collector.



Figure 47: pnp junction in thermal equilibrium with no applied voltages.



Figure 48: pnp transistor in active mode with bias applied



Figure 49: Current flow in an actively biased pnp transistor.

3.6.1 Current Gain

In Figure 49 the current flow in the *pnp* transistor is shown. The holes are injected from the emitter are labelled I_{Ep} . Most of these reach the collector and are labelled I_{Cp} . There are three base currents. I_{En} the current of holes from the base to the emitter, I_{BB} which is the small electron current to neutralize the holes from the emitter that do end up on the base ($I_{BB}=I_{Ep}-I_{Cp}$). I_{Cn} are the electrons that are thermally generated at the collector-base junction. The terminal currents may now be written

$$I_{E} = I_{Ep} + I_{En}$$

$$I_{C} = I_{Cp} + I_{Cn}$$

$$I_{B} = I_{E} - I_{C} = I_{En} + (I_{Ep} - I_{Cp}) - I_{Cn}$$
(3.48)

The most important parameter for a *pnp* transistor is the common-base current gain

$$\alpha_0 = \frac{I_{Cp}}{I_E} = \frac{I_{Cp}}{I_{Ep} + I_{En}} = \frac{I_{Ep}}{I_{Ep} + I_{En}} \frac{I_{Cp}}{I_{Ep}}$$
(3.49)

The first term on the right hand side is known as the emitter efficiency γ which measures hole current in the emitter compared with total current time a transport factor

$$\alpha_T = \frac{I_{Cp}}{I_{Ep}} \tag{3.50}$$

Which measures what fraction of the emitted holes get to the collector. In a good transistor both the emitter efficiency and transport factor are close to 1.

3.6.2 Modes of Operation

Before we finally leave the bipolar transistor we simply note that there are four modes of operation.

- 1. Active mode: the emitter-base is forward biased and the collector-base is reversed biased.
- 2. Saturation mode: in this mode both junctions are forward biased . In this mode the transitory behaves as a switch in the on position.
- 3. Cutoff mode: both junctions are reversed biased. Very little current flows. This is the transitory acting as a switch in the off mode.
- 4. The inverted mode: the emitter base is reverse biased and collector-base forward biased. In this mode, because the collector generally has a lower doping than the emitter the system acts as a transistor with poor emitter efficiency.

References

- 1 S.M.Sze, Semiconductor Devices, John Wiley and Sons
- 2 Principles of Semiconductor Devices, B.v. Zeghbroeck, http://ecewww.colorado.edu/~bart/book/book/contents.htm
- 3 H.M.Rosenberg, The Solid State, Second Edition, Oxford Physics Series

Problems

- 1 Show that the value of the capacity of a *n*-junction is the same as that of a parallel-plate condenser with a plate separation equal to the width of the depletion layers and filled with a dielectric with the same value .of ε .A p- n junction of area 1 mm2 is made of germanium with carrier densities 10^{21} m⁻³ acceptors and donors respectively. Estimate the width of the depletion layers and the capacity of the junction for zero applied voltage. For Ge the energy gap = 0.75eV. ε =16.)
- 2 Estimate the ratio of the forward and the reverse biased currents in a p-n junction diode when the applied voltage is 0.5 V.



JUNCTION DEVICES

In this chapter we deal with other junction devices that are important in the design of particle detectors. Unlike the p-n junction these are all unipolar, in which only one carrier type participates in the in the conduction. These devices are the metal-semiconductor junction, the field-effect transistor (FET), the metal-oxide-semiconductor diode (MOS diode) and the MOS field-effect transitory (MOSFET).

4.1 Metal Semiconductor

The first transistor built was not actually composed of back-to-back *pn* junction but contained two metal-semiconductor junctions, these semiconductor devices were point-contact rectifiers which were direct metal semiconductor contact. Today all solid state devices for particle detection

contain areas of the device that provide for this kind of contact. Thus it is imperative that we understand how this simple junction functions. Metal semiconductor junctions can also be made that conduct current, without rectification, in or out of the semiconductor. In this case the contact is said to be *ohmic*.

4.1.1 Energy Band

As the two materials are brought close, but not yet in contact, the energy bands can be sketched as shown in Figure 51a. The work function of the metal is labelled Φ_M and the electron affinity χ . Note the



Figure 50: Schematic of metal semiconductor contact

vacuum state is the reference state in these diagrams. Figure 51 (b) represents the system just after contact before thermal equilibrium is established. Note that the Fermi-Levels are not yet equalized. This diagram is known as the flatband diagram as non of the bands are "bent".



Figure 51: Diagram showing energy bands immediately before and after contact



Figure 52: Diagram showing the energy bands in thermal equilibrium.

Before thermal equilibrium the barrier height for electrons leaving the metal to enter the semiconductor is simply the difference between the electron work function in the metal and the affinity in the semiconductor

$$\phi_{\rm B} = \phi_{\rm M} - \chi \tag{4.1}$$

will always be a barrier as long as the Fermi level in the metal lies between the conduction and valence bands in the semiconductor. This barrier was first proposed by Schottky as a mechanism for rectification.

As the two materials attempt to reach thermal equilibrium electrons in the semiconductor may reach a state of lowed energy by diffusing across the junction into the metal. As the negative charge leave the semiconductor the positive ionized donors remain, lowering the potential of the bulk. The extra positive

charges make it "harder" for the electrons to be moved to the vacuum state, i.e. the whole band diagram for the bulk is lowered. Electrons will continue to flow into the metal until they are compensated by a flow of "drift" electrons attracted by the positive potential in the bulk. When the flows have equalised the Fermi level in both materials is equal, Figure 52. Note the "band bending" typical of this sort of process.

It is of interest to note that in thermal equilibrium, i.e. with no external voltage applied, there is a region in the semiconductor close to the junction $(0 < x < x_d)$, which is depleted of mobile carriers i.e. a depletion region. The potential across the semiconductor equals the built-in potential φ_{bi} . The built in potential

$$\phi_{bi} = \phi_B - V_n \tag{4.2}$$

Where V_n is the difference in energy between the conduction band and the Fermi Level in the semiconductor.

The calculation of the depletion width in the metal semiconductor junction is very similar to that in the *p*-*n* junction. (See Zeghbroeck). For an applied voltage V_a the total width *W* is given by



Figure 53: Energy bands in an n-type metal semiconductor (a) forward biased and (b) reversed biased junction.

If a positive potential is applied to the metal (the forward bias condition) this will reduce the metal Fermi Level (decrease the energy of the electrons) relative to the vacuum and the semiconductor. This will tend to reduce the barrier height for diffusion and more electrons will diffuse into the metal than electrons will drift into the semiconductor. A positive current therefore flows into the semiconductor. In the case of the a negative potential applied to the metal this has the opposite effect. It raises the Fermi energy of the electrons in the metal and it is now substantially harder for the electron to drift into the metal. Note the barrier preventing the electrons in the metal from drifting into the semiconductor remains approximately unchanged. Another way of thinking about this process is that the negative potential applied at the metal has repelled electrons from the junction region.

4.1.2 Current Voltage

In the metal-semiconductor junction most of the current is transported by the majority carriers. In most cases current is due to thermionic emission of charge over the barrier. Using this it may be shown (see for example Zeghbroeck) that the current in the junction is

$$J = J_s \left(e^{qV/kT} - 1 \right) \tag{4.4}$$

Where the saturation current is given by

$$J_{c} \equiv AT^{2}e^{-q\phi_{Bn}/kT} \tag{4.5}$$

Where A is called the effective Richardson constant.

4.1.3 Schottky Diode

The main difference between the Schottky diode and the forward biased p-n junction is its response to high frequency transients. As there is a negligible diffusion current from the minority carriers there is correspondingly little minority carrier injection or storage and hence capacitance. This gives rise to very fast response time.

4.1.4 Ohmic Contact

If the workfunction of the metal is smaller, or close to the electron affinity, there is no Schottky barrier between the semiconductor and the metal and a direct ohmic contact may be made between the metal and the semiconductor. The contact is called ohmic if there is negligible voltage drop across the junction for an applied voltage.

The specific contact resistance, which should be small for a good contact, is defined by

$$R_{c} = \frac{\partial V}{\partial J}\Big|_{V=0}$$
(4.6)

From Eq (4.4) this shows the specific contact resistance

$$R_c = \frac{k}{qAT} e^{\frac{q\phi_{Bn}}{kT}}$$
(4.7)

This shows, as we may expect, that food resistance is obtained if the barrier height is low.

If the doping is very high in the semiconductor the width of the barrier becomes very narrow Eq. (4.3). under these conditions the tunnelling probability (and

hence the current) is simply a function of the tunnel height as we expect from basic quantum mechanics. The current can be shown (see Sze) to be

$$I \sim e^{-\frac{\sqrt{m_n \varepsilon_s}}{\hbar} \frac{(\phi_{Bn} - V)}{\sqrt{N_D}}}$$
(4.8)

Thus the specific contact resistance for high doping will be

$$R_c \sim e^{\frac{\sqrt{m_n \varepsilon_s} \left(\phi_{Bn}\right)}{\hbar} \sqrt{N_D}}$$
(4.9)

i.e that when tunnelling dominates the specific resistance depends exponentially on

$$\phi_{Bn} / \sqrt{N_D}$$

In order to produce an ohmic contact on Si a local high implant of a donor is often made.

4.2 JFET

Unlike the p-n junction the junction field-effect transistor has a relatively simple mode of operation. It is occasionally suggested, somewhat speculatively, that had the FET been invented before the p-n junction, that FET's rather than bipolar transistors would have been the primary source of the electronics revolution. The simplicity of its operation, like the Schottky diode it is essentially unipolar, makes it much easier to understand than other devices. Until the advent of the high speed MOSFET it was very common in high speed electronics and has far outstripped the bipolar transistor in common usage.

4.2.1 Principle of Operation

The JFET is composed of two contacts which are ohmic (the source and the drain) connected by a normally conducting channel. When a positive voltage is applied to the drain electrons will flow from the source to the drain.



Figure 54: Schematic of a JFET in (a) 3D and (b) a simplified cut through.



The key component of the JFET structure is the gate, which is composed of two parts, an upper and lower gate. The gates are "tied together".

Figure 55: Behaviour of the channel width as a function of different conditions. a) V_G mall and V_D small b) At pinch-off c) $V_D > V_{Dsat}$, d) Small (negative) gate voltage applied.

In normal conditions where the gate voltage is small and the drain voltage is relatively small and with respect to the source then current will flow. If the bulk is n-type then the majority carriers will be electrons. For the n-type JFET a small positive voltage applied to the drain will cause a drift of electrons from the source to the drain.

The gates in the JFET are an ohmic contact to a p+ implants. Thus a negative voltage applied to the gates, relative to the drain voltage will produce a depletion zone around each gate. If both gates are connected to the source then as V_D is increased the depletion zones in the JFET will become deeper and reach out across the n-bulk.

The channel behaves as a resistance of value

$$R = \frac{L}{2q\mu_n N_D Z(a-W)} \tag{4.10}$$

Thus at low applied voltage the current is expected to rise approximately linearly. As the drain voltage is increased then the depletion layers grow. The decrease of the depth of the conduction channel implies an increase in the resistance and hence a reduction in the rate of conduction until a pinch-off condition is met. Figure 55 (b). As the drain voltage is increased further Figure 55 (b) the pinch off moves towards the source, but as no current can flow to the gates, the drain current is largely controlled by the applied drain voltage. At this point the current is saturated.
The saturation voltage is reached when the depletion depth for each gate becomes, a, i.e.

$$V_{Dsat} = \frac{qN_D a^2}{2\varepsilon} - V_{bi} \tag{4.11}$$

where V_{bi} is the built in potential of the gate junction.

Applying a small gate voltage to the system will clearly modify the saturation voltage, in this case for a JFET and a negative gate voltage by

$$V_{Dsat} = \frac{qN_D a^2}{2\varepsilon} - V_{bi} - |V_G|$$
(4.12)

4.2.2 Current Voltage Characteristics

A more detailed analysis of the pinch-off point of a JFET can be gained from calculating the voltage drop across the channel as it thins down and then using this value to calculate both the pinch-off current and voltage.



Figure 56: View of the channel region (a) and the voltage drop along the channel (b). The voltage drop across a small section is given by

$$dV = I_D dR = \frac{I_D dy}{2q\mu_n N_D Z(a - W(y))}$$
(4.13)

The width of the channel is simply given by the depletion depth formula

$$W = \sqrt{\frac{2\varepsilon(V(y) + V_G + V_{bi})}{qN_D}}$$
(4.14)

Assuming the current is independent of y we extract from (4.13) that

$$I_D dy = 2q\mu_n N_D Z(a-W)dV$$
(4.15)

and also obtain from (4.14) that

$$dV = \frac{qN_D}{\varepsilon} W dW \tag{4.16}$$

There we can extract an expression for the drain current

$$I_D dy = 2q\mu_n N_D Z(a-W) \frac{qN_D}{\varepsilon} W dW$$
(4.17)

Which may be integrated along the length of the JFET to yield

$$I_D = I_P \left[\frac{V_D}{V_P} - \frac{2}{3} \left(\frac{V_D + V_G + V_{bi}}{V_P} \right)^{3/2} + \frac{2}{3} \left(\frac{V_G + V_{bi}}{V_P} \right)^{3/2} \right]$$
(4.18)

Where

$$I_P = \frac{Z\mu_n q^2 N_D^2 a^3}{\varepsilon L} \tag{4.19}$$

and

$$V_P \equiv \frac{qN_D a^2}{2\varepsilon_s} \tag{4.20}$$

This behaviour may be seen in Figure 57.



Figure 57: IV characteristics for a JFET

4.3 MOS Capacitor

The MOS capacitor receives its names from its design as Metal-Oxide-Semiconductor junction. The diode is very important device in VLSI integration and in devices used in particle physics.

4.3.1 Principle of Operation

A schematic of the MOS capacitor is shown is Figure 58. The difference between the MOS diode and Schottky diode is the introduction of the insulator (usually SiO_2) in between the metal and the semiconductor. The reverse side of the diode is an ohmic junction. One of the most common MOS capacitor is the Al-SiO₂-Si capacitor which is a structure can easily be built into many particle physics detectors.



Figure 58: Schematic of a MOS capacitor



Figure 59: Flatband diagram for an ideal MOS capacitor not under bias

To liberate an electron in the Al we need 4.10eV. In the insulator, SiO₂, we require about 9 eV to raise an electron into the conduction band and an additional 0.95eV to allow the electron escape to the vacuum. In the Si an electron must

gain a kinetic energy of more than about 4.05eV above the conduction band to escape the crystal. This value as we have already discussed in the electron affinity, χ . The potential barrier between the Si and SiO₂ with reference to the vacuum energy is thus approximately 3.1eV.

In a metal the work function is simply the difference between the Fermi Energy and the vacuum level. In silicon the workfunction is more complicated and given by

$$\phi_s = \chi + \frac{E_g}{2q} + \psi_B \tag{4.21}$$

Where Ψ_B is given by the Boltzmann relations (1.50) and is the difference between the Fermi Potential and the intrinsic potential.



Figure 60: Energy band in an ideal MOS capacitor for n-type and p-type Si for the four different modes. See text.

In an ideal MOS capacitor² the energy difference between the metal and the semiconductor work function is zero. Thus if there is no voltage applied to the metal the both the Fermi and work functions line up. This the flatband condition for the MOS capacitor.



Figure 61: Bandbending diagram for semiconductor oxide interface in MOS capacitor. The figure is for a p-bulk semiconductor. Note that: if $\psi_s < 0$ the device is in accumulation mode, $\psi_s = 0$ is the flatband condition, $\psi_B > \psi_s > 0$, $\psi_B = \psi_s$ yields the intrinsic concentration and $\psi_s > \psi_B$ puts the device into inversion.



Figure 62: Charge density in a p-type MOS capacitor (after Sze)

 $^{^2}$ Two other conditions are required for an ideal MOS capacitor (See Sze). a) That there are no charges in the structure except for those in the semiconductor and on the metal surface. b) That the resistivity of the oxide is infinite.

If a negative voltage is now applied to the gate of the p-type MOS capacitor this raises the metal Fermi level, i.e. it raises the electron energy with respect to the semiconductor. This creates an electric field in the oxide that would tend to move the electrons towards the p-type Si. This induces a field in the Si in the same direction as in the oxide. This will bend the bands upwards towards the oxide interface. Since there is no net flow of charge the Fermi levels in the Si and SiO₂ remain equal. The net effect is that the Fermi level in the semiconductor is much closer to the valence band in the interface than in the bulk i.e. the hole density is greater at the interface than in the bulk. This positive charge accumulation is referred to as the accumulation condition. It is easy to think of this as the holes being attracted to the surface by the negative metal voltage.

If a positive voltage is applied the gate a similar argument applies. The bands bend downwards towards the new metal Fermi energy. Since at this point the Fermi level is still some way away from the bent conduction band there is no great flow of charge but the region becomes relatively depleted of charge compared with the bulk. This is known as the depletion condition.

If a greater positive voltage is applied then the conduction band can bend sufficiently down that electrons are attracted from the bulk to the surface region. In this region the bulk (at the surface) has become effectively inverted due to the applied electric field.

4.3.2 Charge Distribution

The charge density within an ideal MOS capacitor may be calculated (We follow the proof by Sze).We consider a MOS structure where the bulk is p-type.

We assume that the potential ψ is defined to be zero within the bulk of the detector and that the potential at the surface of the semiconductor is ψ_s . The electron and hole concentrations are related by the Boltmann relation to the potential and the temperature i.e.

$$n_p = n_{p0} e^{\beta \psi} \tag{4.22}$$

$$p_p = p_{p0} e^{-\beta \psi} \tag{4.23}$$

Where

$$\beta = q / kT \tag{4.24}$$

When ψ is positive the bands are pushed downwards, see Figure 60 and Figure 61 as the energies of the electrons are reduced. The potential within the semiconductor is governed by Poisson's equation with charge density

$$\rho = q(N_D^+ - N_A^- + p_p - n_p) \tag{4.25}$$

i.e. where the first two terms on the right hand side are the densities of the ionizied donors and acceptors. At large distances from the interface the charge is neutral and the potential has been defined to be zero. Thus

$$N_D^+ - N_A^- = p_{p0} - n_{p0} \tag{4.26}$$

Poisson's equation may therefore be written as

$$\frac{\partial^2 \psi}{\partial x^2} = -\frac{q}{\varepsilon} \bigg[p_{p0} \left(e^{-\beta \psi} - 1 \right) + n_{p0} \left(e^{\beta \psi} - 1 \right) \bigg]$$
(4.27)

The electric field is related to the potential by

$$\boldsymbol{\mathcal{E}} = -\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{x}} \tag{4.28}$$

and thus Eq(4.27) may be integrated to give

$$\boldsymbol{\mathcal{E}}^{2} = \left(\frac{2kT}{q}\right)^{2} \left(\frac{qp_{p0}\boldsymbol{\beta}}{2\boldsymbol{\varepsilon}}\right) \left[\left(e^{-\boldsymbol{\beta}\boldsymbol{\psi}} + \boldsymbol{\beta}\boldsymbol{\psi} - 1\right) + \frac{n_{p0}}{p_{p0}} \left(e^{\boldsymbol{\beta}\boldsymbol{\psi}} - \boldsymbol{\beta}\boldsymbol{\psi} - 1\right) \right] \quad (4.29)$$

Defining the extrinsic Debye length for holes as

$$L_D = \sqrt{\frac{\varepsilon_s}{q p_{p0} \beta}} \tag{4.30}$$

and the function

$$F\left(\beta\psi,\frac{n_{p0}}{p_{p0}}\right) = \left[\left(e^{-\beta\psi} + \beta\psi - 1\right) + \frac{n_{p0}}{p_{p0}}\left(e^{\beta\psi} - \beta\psi - 1\right)\right]^{1/2}$$
(4.31)

Then the electric field becomes

$$\mathcal{E}_{surface} = \pm \frac{\sqrt{2kT}}{qL_D} F \tag{4.32}$$

Knowing the electric field at the surface the charge density at the surface may be inferred

$$Q_{surface} = -\varepsilon \mathcal{E}_{surface} \tag{4.33}$$

A schematic of the space-charge density at the surface of the semiconductor is shown in Figure 62 as a function of the surface potential.

Note that the differential capacitance defined by

$$C_D \equiv \frac{\partial Q_{surface}}{\partial \psi_s} = \frac{\varepsilon}{\sqrt{2L_D}} \frac{\left[1 - e^{-\beta \psi_s} + \left(n_{p0} / p_{p0}\right)\left(e^{\beta \psi_s} - 1\right)\right]}{F}$$
(4.34)

and that from this under the flatband condition $\psi_s = 0$

$$C_D^{flatband} = \frac{\varepsilon}{L_D} \tag{4.35}$$

In the depletion region the Si acts as an additional capacitor in series with the oxide so the capacitance of the drops in that region.

4.3.3 SiO₂-Si Interface

In reality the silicon dioxide will not be a perfect insulator. There may be mobile ionic charges (electrons or holes) trapped within the oxide, for example potassium and sodium ions. There may also be surface charges from surface states at the Si oxide boundary, Figure 63.

The surface states can act as recombination and generation centres for both holes and electrons in the material, and thus act (as in the case of bulk Si) to reduce the mobility of the carriers in the region of the interface and lead to leakage currents along the surface. The density of surface states will in general depend very strongly on the fabrication process, including the crystal orientation. In general the number of surface traps in Si is [100]<[110]<[111].



Figure 63: SEM View of the SiO2 Si interface. The interface may be made almost perfact.



Figure 64: Charges in thermally oxidised Si

In the region of the interface there are also fixed oxide charges introduced, in the main, by Si introduced into the oxide during oxidation (see next chapter). The dependence on the fixed charge on the crystal orientation again follows the rule [100]<[110]<[111].

Inside the oxide mobile charges (e.g., K) exist due to contamination. These will greatly effect the properties of the MOS junction. Thus every effort is usually taken to avoid this kind of contamination.

4.3.4 Radiation Effects

If electron hole pairs are generated within the oxide, for example by ionizing radiation some of these will remain trapped. These surface traps may be removed by high temperature annealing ($T\sim600K$) but unfortunately this is not normally an option with a constructed detector.

The effect of these charges is in general to change the electric field configuration at the interface, essentially introducing a surface sheet charge.

4.4 MOSFET

The MOSFET is the single most important devices for very-large-scale integrated devices.

4.4.1 Principle of Operation

The device is, like the JFET, a four terminal device. In the example shown (See Figure 65and Figure 66) there is a source, drain, gate and substrate bias. The central section, or the gate, is a MOS capacitor as discussed above. If the bulk is p-type, as drawn below, then the source and drain will be ohmic contacts to n implants. These act as two back to back p-n junctions. When no voltage is applied to the gate the only current that can flow from the source to the drain is thus a reverse bias current.



Figure 65: 3D schematic of a MOSFET (after Sze)



Figure 66: SEM of a MOSFET. Cross sectional view.

When a large, positive, voltage is put onto the gate an inverted n-channel is created the spans between the source and the drain. Under these conditions charge may be transported and large current may flow. The resistance of the channel is controlled by adjusting the gate voltage. We are then in the position that a drain current I_D will flow between source and drain. For low values of a (positive) drain voltage the electron current will simply increase linearly with applied voltage. Just as in the case of the JFET the n channel in the p bulk forms a reversed biased diode in its own right(Assuming the bulk is at zero volts) with a small depletion layer. As the voltage is increased further on the drain the depletion zone will tend to pinch-off the n-channel. At this point the current then saturates as with the JFET, see Figure 67. Further increasing the depletion region, or the drain voltage, will then not increase the current



Figure 67: Schematic showing the current in a MOSFET at a) linear region, b) at pinchoff and c) in saturation

4.4.2 Types of MOSFET

There are many types of MOSFET. Most practical devices insulate the MOSFET to other structures by the use of a very thick oxide layer called the Field Oxide. (FOX), see Figure 68.



Figure 68: Figure showing the field oxide surrounding a MOSFET structure.

JUNCTION DEVICES

TYPE	CROSS SECTION	OUTPUT CHARACTERISTICS	TRANSFER CHARACTERISTICS		
n-CHANNEL ENHANCEMENT (NORMALLY OFF)	<i>m</i> b b b b b b b b b b	$0 \xrightarrow{I_D \qquad V_G = 4V} V_D$	- O V _{Tn} +		
n-CHANNEL DEPLETION (NORMALLY ON)	n- CHANNEL	$0 \xrightarrow{I_D \qquad V_G=1V} V_D$	V_{Tn} - 0 + V _G		
p-CHANNEL ENHANCEMENT (NORMALLY OFF)		$-V_{D}$	$\begin{array}{c c} & v_{G} \\ \hline & v_{Tp} \circ & + \\ \hline & & \\$		
p-CHANNEL DEPLETION (NORMALLY ON)	p-CHANNEL	$-V_D $ $1 $ $0 $ $V_G = 1V $ $I_D $	- 0 + 1		

Figure 69

Figure 69: Different types of MOSFETs that may be constructed

The choice of bulk type and operating conditions also define whether or not the device is normally on, normally off and at what voltages this happens. See

References

- 1 S.M.Sze, Semiconductor Devices, John Wiley and Sons
- 2 Taur and Ning, Fundamentals of Modern VLSI Devices
- 3 Principles of Semiconductor Devices, B.v. Zeghbroeck, http://ecewww.colorado.edu/~bart/book/book/contents.htm

Problems

1 Describe qualitatively the functioning of a p-bulk MOSFET device. Draw a diagram that shows the conduction channel, and pinch-off. Explain why the drain current saturates at pinch-off.



WAFER PROCESSING

In this chapter a brief introduction to the basic Si fabrication techniques behind most semiconductor devices. Attention will be drawn to processes that are off particular importance to the creation of particle physics detectors. Detector design often begins with the choice of wafer to be used in construction.

5.1 Wafer Production

Imost without exception the semiconductor devices that are in use in particle physics today have started out as wafers cut from a crystal. We begin this chapter with a brief description of the different techniques of producing the wafers. The most important for tracking detectors is currently the traditional float-zone production method. But we start with the first to be discovered, the Czochralski method.

5.1.1 Czochralski Method

The first step in producing Si crystal is to start with a source of silicon, for example common sand. This can be reduced through chemical means to relatively impure metallurgical grade Si (98%). This solid form of Si is reacted with HCl to make trichlorosilane, which is a liquid a room temperatures. By first purifying the liquid through a distillation process and then reducing the trichlorosilane using hydrogen, pure polycrystalline Si may be produced.

The polycrystalline Si is the raw material for the large crystal Si we need.



Figure 70: Crucible of pure poly crystalline silicon



Figure 71: Schematic of the Czochralski method of producing a Si crystal

The Czochralski technique use apparatus which is sketched in Figure 71. The polycrystalline silicon is first melted in the furnace using RF induction. A small seed (Si) crystal is placed on the surface of the molten Si in an inert atmosphere (e.g. Argon). To achieve uniformity both the melt and the seed are counterrotated. Molten Si attaches itself in crystal layers to the seed which is then slowly withdrawn as it rotates. The withdrawing of the crystal is performed via a pulling mechanism that guarantees an even rate of movement at a few mm a minute. The crystalline Si solidifies as it is drawn away from the melt. The final product is a pure crystal. An example is shown below.



Figure 72: A single Si crystal grown using the Czochralski method

INTRODUCTION TO SEMICONDUCTORS

One advantage of the Czochralski technique is that it may be used to produce very large crystals. Large refers here not simply to length but also to the diameter of the crystal, which may now exceed over 300mm. The crystal is ground to produce a perfect cylinder and etched to remover any imperfections left by the grinding.

Characteristic	CZ	FZ
Oxygen Content	$>10^{18} \text{cm}^3$	$<10^{16} \text{cm}^{3}$
Carbon Content	$>10^{17} \text{cm}^3$	$<10^{16} \text{cm}^{3}$
Minority Carrier Lifetime	5-100(ms)	1000-20000ms

Table 4: Sample of purity differences between CZ and FZ Si

5.1.2 Float Zone Method

The float zone (FZ) process is used to produce crystals that are purer than the (cheaper) Czochralski technique. It entirely avoids contact with a potentially contaminating crucible.

The FZ process consists of the following steps.

1. A polysilicon rod is mounted vertically inside a chamber, which may be under vacuum or filled with an inert gas.

2. A needle-eye coil that can run through the rod is activated to provide RF power to the rod, melting a 2-cm long zone in

the rod. This molten zone can be maintained in stable liquid form by the coil.







Figure 74: Photograph of the FZ process in action

3. The coil is the moved through the rod, and the molten zone moves along with it.

4. The movement of the molten zone through the entire length of the rod purifies the rod and forms the near-perfect single crystal.

It is then ground using diamond wheels to make it a perfect cylinder that has the right diameter. It then undergoes an etching process to remove the mechanical imperfections left by the grinding process.

5.1.3 Wafer Cutting

After the crystal is grown, ground and chemically polished normally "flats" are ground into the crystal to aid mechanical location to position the crystal. Smaller secondary flats are also ground. Then position of the flats indicates the crystal orientation of the crystal.

The crystal is then diced by a diamond saw. The dicing delivers both a surface orientation (e.g. [100]) and a thickness that may be anywhere from about 50microns upwards.

Post cutting the wafers are polished, first by lapping and then chemically to produce smooth flat wafers. If very thin wafers are required this too may achieved by vigorous grinding and chemical means.



Figure 75: 300mm cut wafer from a large Si crystal. (MEMC corp).

5.1.4 Epitaxial Growth

Last but not least we note that extremely pure, but thin, sample of crystalline Si may be grown using a method called epitaxy. In epitaxy a substrate Si wafer acts as a seed. However rather than a high temperature process, such as the CZ or FZ crystal growth methods an epitaxial layer may be grown at substantially lower temperature than melting point.

The most common form of epitaxy is vapour phase epitaxy where Si, produced by chemical reactions at high temperature is precipitated onto the Si substrate. The major advantage of epitaxy is that engineers can control the doping profile of the Si much more easily than by ion implantation (see later). In addition the physical and chemical properties of the epitaxial layer(s) may be made very different from the bulk. This is of relevant to CMOS devices that we will briefly discuss at the end of this course.

5.2Wafer Properties

In the section above we have mentioned briefly some of the relevant parameters that effect our choice of wafer. Here we simple detail the key parameters and their relation to device characteristics.

5.2.1 Resistivity

The wafer resistivity is a function of the purity of the crystal and the number of defects. As both CZ and FZ wafers are nowadays almost completely defect free the major contributor to the resistivity is the carrier density which is controlled during the production process. The purer the crystal the higher the resistivity.

Although for VLSI purposes resistivities of $50\Omega cm$ are adequate for particle physics strip detectors we often require very high resistivities (> $1000\Omega cm$). This requirement comes not from the depletion depth, nor even from an exotic analysis of the high frequency behaviour of the devices but from a simple signal to noise requirement on the detectors.

As we shall see in many particle physics devices the fundamental key to their operation is the creation of a depleted zone within which electron-holes pairs can be created. These are then detected as the signal for the passage of ionizing radiation. In order to maximise the signal the leakage current, on the reverse biased diode, must be kept to a minimum. This is best achieved by having a high resistivity, or low number of carriers within the material.

5.2.2 Thickness

The thickness of wafers usually used in particle physics applications varies from about 50 microns to about 500 microns. We shall see that the advantage of thinner detectors is that they contribute little to the degradation of experimental performance through multiple coulomb scattering (or showering) within the Si. The major disadvantage, as we will discuss, will be that thinner detectors yield smaller signals are substantially harder to handle.

Typically devices will be fabricated on readily available commercially available thicknesses of wafer (rather than custom ground wafers) at about 200 or 300 micron thickness.

5.2.3 Crystal orientation

At one point in the 1990's much research was conducted on the performance of the different crystal orientations available to manufacture sensors with. It is usual now, for strip diodes, to use [100] wafers. As discussed above one of the advantages of [100] is has less surface and junction impurities.

5.3 Ion Implantation and Diffusion

Both diffusion and ion implantation serve to change the carrier concentrations inside the Silicon. Diffusion is a cheaper and more simplistic method used almost exclusively before the 1970, but can only be performed from the surface of the wafers. Dopants may also diffuse unevenly, and interact with each other altering the diffusion rate.

Ion implantation is more expensive and complex. It does not require high temperatures and also allows for greater control of dopant concentration and profile. It is an anisotropic process and therefore does not spread the dopant implant as much as diffusion. This aids in the manufacture of self-aligned structures which greatly improve the performance of MOS transistors or indeed in precise particle physics detectors.



Figure 76: Schematic of diffusion and ion implantation processes

Diffusion is normally achieved by inserting semiconductors wafers into a high temperature environment, a furnace, an passing a gas that includes the desired impurity over the surface of the wafers. Temperatures for the diffusion process vary between 800 and 1200°C for Si.

To produce an p-type impurity the Si wafers are usually doped with boron, to produce a n-type impurity arsenic and phosphorus are used. Typically, as mentioned above, the source material is carried by an inert gas (e.g. nitrogen) and reduced on the surface. For example in the following reaction an oxide layer is produced on the surface

$$2As_2O_3 + 3Si \rightarrow 4As + 3SiO_2 \tag{5.1}$$

5.3.1 Diffusion Equation

The diffusion process follows a similar behaviour to that of electrons or holes in regions of changing electron/hole density. The rate of movement of impurities will depend on the impurity gradient $\partial C / \partial x$ and a diffusion coefficient *D*. The impurities/dopant atoms will move away from a region of high concentration towards on of lower concentration. The diffusion equation (Fick's equation)

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2}$$
(5.2)

5.3.2 Diffusion Profile

If the surface is kept at a constant density then the diffusion equation may be solved

$$C(x,t) = C_{s} \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right)$$
(5.3)

Thus the density at a known depth may be calculated for a specific diffusion time. The diffusion coefficient itself is known to depend on the temperature of the diffusion process, as we may expect

$$D = D_s \exp\left(\frac{-E_a}{kT}\right) \tag{5.4}$$

The energy E_a is related to the energy necessary to move dopant atoms within the crystal from one site and another. In Si this is about 1eV and depends heavily on the dopant, boron for example diffusion 10 times faster than arsenic.



Figure 77:Diffusion profile versus Time.

5.3.3 Intrinsic and Extrinsic Diffusion

Diffusion that takes place for dopant concentrations less than the intrinsic carrier density at the diffusion temperature (e.g. $n_i=5\times10^{18}$ cm⁻³.at 1000°C) is called intrinsic diffusion. For densities higher than this the diffusion coefficients become a lot more complicated and depend on the local doing density. This is called the extrinsic doping region.

5.3.4 Ion Implantation

Ion implantation is effected by injecting highly energetic atoms of the dopant directly into the surface of the material. Ion implantation energies may vary anywhere from 30eV all the way to 1MeV. Ion implantation doses may vary from about 10^{11} to 10^{16} ions/cm².

The ion implantation process begins by the ionization of atoms of the dopant (e.g. B^+). These are extracted from a plasma, bent through an analysing magnet so they are almost monochromatic and then accelerated and focussed onto the Si wafers see Figure 78:Schematic of ion implanter. Real ion implanters are large complex pieces of equipment, e.g. Figure.



Figure 78:Schematic of ion implanter.



Figure 79 Varian Semiconductor, VIISion 80 Source/Terminal

5.3.5 Implantation Range

As we might expect the range of the ion within the semiconductor will be strongly dependent on the energy of the incoming ion, and the cross-section for interaction with the atoms within the crystal. As the ions collide with the atoms they will lose energy eventually coming to rest at average distance or range.



Figure 80: Schematic of the ion range R and the two dimension distribution of the ions

There is not only a longitudinal distribution of range, due to the stochastic nature of the collisions, but a lateral profile. The lateral profile is due to collisions with the crystal in which the ions acquire a substantial lateral deflection. Typical ranges for ions in Si as a function of ion energy are shown in Figure 81.



Figure 81: Range of ions in Si.

5.3.6 Channelling

Although it may be thought that the process of ion implantation is straightforward it is complicated by an effect called channelling. Due to the crystal nature of Si there are orientations of the wafer in which the ions will have much longer range that the average range above. If the ions are precisely aligned along certain crystal orientations they will tend to see channels through the crystal that present relatively low cross section for interaction. See for example Figure 82. The directions for channelling are well known in Si, see Figure 83



Figure 82: Diamond structure along the [110] axis



Figure 83: Channels in Si as function of orientation towards the wafer.

5.3.7 Disorder

When the ions enter the Si they lose energy via both nuclear and electronic interactions with the crystal. Electronic collisions generate electron-hole pairs but the nuclear collision have enough energy to move Si atoms within the lattice. The relative ratio of electronic to nuclear collisions depends on the mass of the incident ion. High energy boron ions, for example, have almost exclusively electronic interactions. These light boron ions (see Figure 84) thus cause very little damage until the end of their range when nuclear interactions occur.



Figure 84: Difference between a) damage caused by light ions and b) damage caused by heavy ions in Si.

									٩	0	0	0	0							
						0	0	0	0	ò	ο	0	0							
						0	0	0	0	ò	١O	0	0							
						0	0	0	0	0	``	0	0							
						0	0	0	0	0	0	0	0							
						0	0	0	0	0	0	0	0							
						0	0	0	0	0	0	0	0							
						0	0	0	0	0	0	0	0							
Vacancy defect																				
0	0	0	0	о	о	о	0						0	0	0	0	0	0	0	0
ō	· _``	0	0	о	о	о	0						0	о	о	о	0	0	0	0
0	0	01	0	0	0	0	0						0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0						0	о	0	0	0	<i>`</i> ,0	0	0
0	0	0	0	0	0	0	0						0	0	о	о	0	ò.	`~	0
0	0	0	0	0	0	0	0						0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0						0	0	0	0	0	0	0	0
Interstitial defect								Frenkel Defect												

Figure 85: Defects in Si caused by ion collisions

Heavier ions the energy loss is due to nuclear interactions over the whole length of the range. The damage is so substantial that locally the crystal will have a high density of vacancies, interstitial and Frenkel defects. For certain types of ion implantation the damage can be so severe that the silicon becomes almost amorphous.

5.3.8 Annealing

The severe damage caused to the crystal by the ion implantation causes severe degradation to the carrier properties in the Si. To restore the properties of the Si the semiconductor the material is annealed. This process allows the crystal atoms to move back to stable crystal positions.



a) Damaged Si lattice during implant



Figure 86: Repair to the defects in Si by annealing process



Figure 87: Annealing temperature for 90% activation of dopant

5.4 Lithography and Etching

5.4.1 Optical Lithography

The process of lithography is fundamental to the creation of complex structures in Si. It enables us to place patterns of dopant, oxide, metal or resistors in a complex 3D structure on one (or both) sides of a wafer. The patterns are transferred to the wafer by lithography through the use of photo-resist and (in most cases) optical exposure. The photo-resist material has the property that its chemical bonds are changed by exposure to light, making it possible to selectively remove, or etch, the photoresist, hence wither protecting or exposing parts of the surface.

The process of exposing the photoresist, with a sharp high resolution pattern is both complex and critical to the successful production of high quality semiconductor devices. In any lithography method an intense light source is focussed, through a mask that defines the shapes, onto the thin film of photoresist.



Figure 88: Three methods of imaging onto a wafer



Figure 89: Modern commercial lithographic and etching machinery.

Three basic techniques may be used. (see Figure 88) contact printing in which the mask is direct contact with the photoresist, proximity printing in which there a few microns gap between the mask and the photoresist, and projection printing in which complex optics allow an image to be focussed onto the photo resist. In most modern applications wither proximity or projection printing are used as contact printing introduces a lot of errors into the final device.



Common UV wavelengths used in optical

Figure 90: Range of wavelengths used in common lithography.

There is a wide range of wavelengths used for photolithography. In general the shorter wavelengths yield much higher resolution (certainly down to 0.25microns) for normal UV light but are substantially more expensive than the longer wavelengths.

5.4.2 Mask Production

As well as the exposer the object that enables to print a particular pattern on a photoresist is a precise image of the feature that needs to be reproduced. Normally this is a very high precision object as any defects will be replicated onto the Si. The mask is normally produced on high grade quartz and is a chrome metal. The mask itself may be a master from which other copy masks are made. The copies are usually much cheaper than the mask and have a finite lifetime due to handling and, in the case of proximity and contact exposure, the process of printing itself. The copies may be easily reproduced from the master. It is normal in detector design to order maybe 4-5 copies for each (critical) master for a typical production run of 100 detectors.

As will be seen from the next section the pattern on the mask will depend on whether positive or negative photoresists are being used.

INTRODUCTION TO SEMICONDUCTORS



Figure 91: Schematic of simple exposer and aligner.

The mask must also contain on it, as well as the pattern we want to produce, alignment marks to allow coregistration of patterns in different steps of the processing. As tolerances of much less than a micron are required in some physics detectors (and much less than that in CMOS device production) the alignment marks are critical.



Figure 92: Positive and negative masks for photolithography.

The masks are a surprising expensive part of producing a detector design. For a typical n+n strip over 13 master masks must be produced as well as copies.

5.4.3 Negative Lithography

If we use a negative photo resist the pattern that is left after the detector is exposed and the photoresist etched away is the negative of the original mask pattern i.e. areas exposed to light remain.



Figure 93: Negative photoresist in the lithographic process.

5.4.4 Positive Lithography

If we use a positive photo resist the pattern that is left after the detector is exposed and the photoresist etched away is the positive of the original mask pattern i.e. areas not exposed to light remain.



5.4.5 Photo Resist

The chemistry of photoresist is complicated. Both negative are positive resists may be carcinogenic and must be treated carefully. They must be stored in ideal conditions and applied evenly across the whole wafer before exposure. This is typically done with a spinning device that dispenses photoresist onto the centre of a rapidly rotating wafer (in some cases up to 5000 rpm). The viscous photoresist flow evenly across the surface and any excess is spun off. On contact with the air and the surface of the Si the photoresist becomes very sticky. If the process is performed correctly a known depth of photo-resist may be applied across the whole wafer. Before exposure the resist is baked to harden it further.



Figure 94: Simple diagram showing the spinning of photoresist onto a wafer.

5.4.6 Etching

After exposure the photoresist is removed in places where it is not needed. This may be done either by putting the wafers in a solvent (Wet etching) or by the used of plasma etching technique (dry etching). Dry etching is regarded as a superior technique as it produces cleaner, sharper corners in most cases.



Figure 95: Wet and dry etching techniques. Wet etching is in a solvent. Dry etching is in a plasma.



Figure 96: Wet etching an oxide layer. Note the shape under the photoresist



Figure 98: ODE Wet etching on Silicon



Si / SF6 + Sidewall Protective Gas Independent of Crystallographic Axis

Figure 97: Dry etching of Si

Figure 96 shows a diagram of wet etching an oxide layer. Notice how the etchant undercuts the photoresist and the oxide takes on a shape. This must be accounted for in all mask and device design.

Figure 97 and Figure 98 are a schematic of wet and dry etching on Si. For wet etching it is possible to produce orientation dependent etching (ODE) on Si in which different crystal orientations are etched selectively. This enables one to make, if necessary, different shapes in the Si.

5.4.7 High Resolution Lithography



Figure 99: Example of extreme UV lithography showing 50nm structures

After several years of feasibility studies ASML, Zeiss, Philips and TNO TPD started the development of an Extreme Ultra Violet (EUV) waferscanner capable of printing 50 nm dense lines with a corresponding overlay, see Figure 99. Extreme Ultraviolet Lithography operates at a wavelength of 13 nm. Compared to operating wavelengths of previous lithography technologies this short wavelength allows for high resolution and increased depth of focus. Since EUV light is strongly absorbed by all materials, operation in vacuum and use of mirror optics is necessary. Extreme UV lithography is viewed as being a relatively effective way of moving towards higher density of CMOS components.

5.5 Polysilicon

5.5.1 Deposition Techniques

Films of polycrystalline silicon, or polysilicon (also known as poly-Si or poly), are widely used as MOS transistor gate electrodes and for interconnection in MOS circuits. It is also used as resistor, as well as in ensuring ohmic contacts for shallow junctions. When used as gate electrode, a metal (such as tungsten) or metal silicide (such as tantalum silicide) may be deposited over it to enhance its conductivity.

Poly-Si is known to be compatible with high temperature processing and interfaces very well with thermal SiO2. As a gate electrode, it has also been proven to be more reliable than Al. It can also be deposited conformally over steep topography. Heavily-doped poly thin films can also be used in emitter structures in bipolar circuits. Lightly-doped poly films can also be used as resistors.



Figure 100:Polysilicon resistors in a silicon detector (BaBar). For particle physics experiments these resistors are typically of a value of about $1M\Omega$

Poly-Si is usually deposited by thermal decomposition or pyrolysis of silane at temperatures from 580-650 degrees C, with the deposition rate exponentially increasing with temperature. The deposition rate is also affected by the pressure of silane, which translates to silane concentration. Other important variables in polysilicon deposition are pressure and dopant concentration.

5.6 Dielectric Layers

Both standard MOS circuits and particle physics detectors depend critically on the use of thin insulating layers, usually SiO₂ although sometimes silicon nitride

is used as well. There are currently four commonly used techniques for putting down an insulating layer: thermal oxidation, chemical vapour deposition (CVD), low pressure CVD (LPCVD) and plasma deposition.

5.6.1 Thermal Oxidation

This process provides very high quality oxide. Typically either water or oxygen are introduced into a furnace at between 900 to 1200°C.



Figure 101: Schematic of thermal oxidation process

The oxide tends to grown slowly, rates of about 100Å/hour being typical. This techniques is used to put down thin layers of high quality oxide.

5.6.2 Chemical Vapour Deposition

Silicon dioxide may also be deposited using either CVD or LPCVD. Silicon dioxide may be created by reacting silane, dopant and oxygen e.g.

$$\mathrm{SiH}_4 + \mathrm{O}_2 \to \mathrm{SiO}_2 + 2\mathrm{H}_2 \tag{5.5}$$

This process takes place at about 500°C and is a low temperature deposition interaction. High temperature methods also exist. The advantage of the CVD methods is that relatively thick films may be put down, for example some detectors we will discuss used CVD layers of up to 4 microns thick.

5.7 Metallization

The final fabrication technique that we will briefly mention is the deposition of aluminium. This may be applied either a vapour deposition technique or by the physical sputtering of aluminium on the surface of a wafer.

5.7.1 Sputtering

In sputtering Argon ions are accelerated until they hot an aluminium cathode, or target. Through momentum transfer atoms on the surface of the cathode are ejected from the target and transported and deposited on the surface of the substrate. Quite thick (1 to 2 microns) layers of aluminium may be deposited in this way, although the quality is not as high as may be achieved in a CVD method. However sputtering does have the advantage of being much cheaper and faster.



Figure 102: Schematic of sputtering technique used to deposit AI on a wafer.

References

- 1 S.M.Sze, Semiconductor Devices, John Wiley and Sons
- 2 http://www.memc.com/

Problems

1 Discuss how many processing it takes to create a) a workable a p-n diode and b) a simple MOS capacitor.



SIMULATION

The role of simulation is discussed briefly. Its scope and technicalities are summarized. All discussion refers to the standard commercial package, ISE TCAD®, although freeware applications exist.

6.1 Simple models

More than the set of t

A good flexible, reliable and sophisticated package that is required. Simple models, that in essence are a straight forwards extensions of the sort of the analysis considered in the earlier chapters will fail to provide the accuracy embodied in programmes that contain sophisticated 3D meshing and numerical analysis techniques. To attain a realistic description of the behaviour devices it is also necessary to use libraries of known experimental cross-sections, material properties and phenomenological data. These are also not usually available to simple numerical models.

Moreover a detailed understanding of the behaviour of a particular device may function must contain detailed knowledge of the production process. Many features such as the mechanical dimensions, the implant profiles, and the properties and impurity level will depend on the exact production parameters. These must be simulated to provide precise data.

So, although simple models possess a pedagogical value, and may act as a guide to understanding device characteristics only a sophisticated 3D programme that contains good numerical techniques, production simulation and as complete knowledge of semiconductor behaviour as possible should be used when designing a real device.

6.2Commercial Package

There exist a number of commercial packages available to perform device simulations. We will report on one, used by the Liverpool Semiconductor Detector Centre (LSDC)³, for the design of strip detectors, CCDs and Active Pixel devices. The program is called ISE TCAD[®] and is now part of Synopsis[®].

6.2.1 ISE

The ISE TCAD programmes form part of a suite of design, simulation, optimization and graphical output that are used by professional designers, and educational institutes, to model realistic. For the purposes here we describe briefly, using the ISE's own words, taken from their web site, the function of three critical components. The MESH routines that provide the 3 dimensional meshes necessary to perform the finite element analyses. The DIOS suite that enables the user to fully simulate the production process and finally DESSIS that simulates the electrical, thermal, and optical characteristics of semiconductor devices.

6.2.2 MESH

ISE provides two state-of-the-art approaches for the automatic generation of meshes. The quadtree/octree-based method MESH creates meshes with an axisaligned structure that is fitted to the boundary. The mesh generator NOFFSET3D is fully unstructured and pays special attention to mesh elements near material interfaces.

Both methods create meshes that can be used for the discretization methods used in ISE process and device simulators. The meshes are not only adapted to the geometry, but also to the doping concentration in order to capture steep gradients.

Refinement is partly performed with anisotropic (stretched) elements, which means that meshes with fewer elements can be used to speed up the simulation without losing accuracy.

As the process simulators DIOS has an internal mesh generator for 2D applications, the external meshing tools are optional for DIOS. However, for device simulation, they are mandatory.



Figure 103: 3D mesh created by ISE's MESH routine

³ ISE-TCAD has been used by LSDC from 1998 until the time of writing these notes, 2004.
6.2.3 DIOS

DIOS is a multidimensional process simulator for semiconductor devices. It allows for the simulation of complete fabrication sequences including etching and deposition, ion implantation, diffusion, oxidation, and silicidation.

It is the most accurate process simulator available. In cooperation with Applied Materials, Inc., ISE engineers have calibrated DIOS to a large experimental data matrix that focuses on state-of-the-art deep submicron technology.

In addition to analytic implantation models, DIOS includes the Monte Carlo simulator Crystal-TRIM. Diffusion is simulated on the basis of state-of-the-art point defect models that are calibrated to a large number of experiments. Mechanical effects such as stress, flow, and thermal expansion are included.

DIOS has been applied to a wide variety of technologies such as VLSI CMOS, SiGe HBT, power devices, and advanced SOI processes, in many of the leading semiconductor companies



Figure 104:Device simulated using DIOS

6.2.4 DESSIS

DESSIS simulates the electrical, thermal, and optical characteristics of semiconductor devices. It is the leading device simulator and handles 1D, 2D, and 3D geometries, mixed-mode circuit simulation with compact models, and numeric devices. It contains a comprehensive set of physical models that can be applied to all relevant semiconductor devices and operation conditions.

DESSIS is used to evaluate and understand how a device works, optimize devices, and extract SPICE models and statistical data early in the development cycle.

Applications of DESSIS include VDSM silicon, where DESSIS has proven accuracy to well below 100 nm technology; silicon-on-insulator (SOI) devices, where DESSIS is known for its robust convergence and accuracy; double-gate and FinFET devices, where quantum transport is a reality; SiGe; thin-film transistors; heterojunction HEMTs and HBTs; and power and RF semiconductor devices.

6.3 Applications

Within particle physics ISE, or similar programmes are used, to perform various studies, some of which we will give examples of below.

6.3.1 Field Modelling

The first study to be perform is an analysis of the structure of the detector, in the example below an LHCb strip detector, in order to determine the electric field structure within the sensors. This will directly affect breakdown conditions, carrier mobilities etc. Superimposed on the 2D diagram below is the generated mesh that enables the finite element analysis.



Figure 105: ISE used to mesh and solve for the electric field within an LHCb Si sensors. The AI readout strips are in pink. The device I fully simulated using DIOS and details of manufacture from the fabrication company Micron SemiConductor Ltd. In this case the physicist is look for areas within the lower meshed area with high fields where breakdown may occur. Simulation from the LSDC.



Figure 106: Another field simulation on an LHCb sensor, in this case of the guard ring structure that separates the active region of the sensor from the non-active edge area. Simulation from Glasgow University. In this case DIOS was not used.



Figure 107 Potential, rather than field diagram for the LHCb sensor. The potential diagram is useful in studying surface effects.

6.3.2 Carrier Densities

The distribution of carriers, both holes and electrons, is vital to understand the creation of the signal in the sensors. These too may be fully simulated within ISE.



Figure 108: Density of electrons at thermal equilibrium within the LHCb sensor.



Figure 109: Density of holes at thermal equilibrium within the LHCb sensor

6.3.3 Passage of ionizing irradiation

The detection of ionizing radiation is the primary function of most silicon sensors within particle physics. ISE allows us to simulate the effect of ionization within our material and then study how the effects propagate.



Figure 110: Simulation of a heavy ion passing through a 300 micron thick n-type LHCb sensor. The mesh is not shown. The actual core of the charge density is no more than 2 microns thick. Note the scale is logarithmic.



Figure 111: A picture 4ns after the passage of the ionization through the sensor the holes are moving towards the junction side (at 300 microns Y). Notice the trail of charge leads up to two strips.



Figure 112 A picture 4ns after the passage of the ionization through the sensor the electrons are moving towards the strip pickup side (at 300 microns Y). Notice the chare charge moving towards two strips.

6.3.4 Process Optimization

Details of the processing are typically optimized, e.g. implant depths to produce the best possible signal to charge ratio in the sensor and to maximize the voltage at which the detector may operate. In the simplest caser this means simulation of the dopings, annealing times and electrode structures may be varied.

6.3.5 Charge Sharing

We shall see in the section on sensor design that features such as resolution or the charge sharing between strips are critical to the performance of the sensor. This is another feature modelled using ISE.

6.3.6 Charge Transfer

Finally DESSIS allows to simulate the signals that are seen in the readout electronics of the devices. The is especially useful in comparing data with simulation. It is necessary to have a detailed model of the front-end amplifier used for readout.

Radiation Damage Effects

One of the most important, and hard to model features, is effect of large scale, long term irradiation of the devices. These are studied using DIOS and DESSIS to investigate signal size and shape as a function of the lifetime of the sensors. To model the effects of ionizing radiation additional energy levels are added that may act as traps.



Figure 113: Signal observed in sensors as a function of proton irradiation. Scale on the x-axis is dose in p/cm² x 10¹⁴. y-axis is proportional to the signal observed in. Note the drop in total charge collected, which matches experimental data well.

6.4 Limitations

One of the biggest limitation seen in simulating devices is due to the numerically intensive nature of the modelling that the passage of a simple particle may take up to 24 hours to simulate. In addition true 3D modelling, rather than creating 2D simplifications, complicates and slows the simulation even further. This is not an

INTRODUCTION TO SEMICONDUCTORS

intrinsic limitation of the package used here but a limitation of the hardware used to run it. However it does indicate that substantial computing hardware resources should be made available to any team wishing to simulate or design a device.

References

1 http://www.ise.com

Part 2 Devices in Particle Physics



HEP EXPERIMENTS

An overview of accelerator and the technology used to generate the high energies involved in Particle Physics is given. This is followed by a short description of detector technology including the role of tracking and vertexing detectors that use semiconductor technologies. The notes here are a compilation of material available on the Web.

7.1 Accelerators

The purpose of accelerators is to provide a high energy beam of particles that may be used as a probe of fundamental particles. Since the kinetic energy particle energies are on the same order of or are above the rest energy of a proton, the acceleration of these particles is extremely complex process, typically involving many different accelerator components and technologies.

7.1.1 Accelerator Beam Energy

In any discussion of accelerators one typically discusses the energy of the accelerator, or more specifically the beam energy. The particles in the beam will be, for particle physics experiments, heavily relativistic.

For relativistic particles with a velocity (v) it is usual to define the Lorentz β as

$$\beta = v/c \tag{7.1}$$

Another quantity, the Lorentz γ , is defined in relation to β as

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} \tag{7.2}$$

The total energy, E, of a particle may be simply related to its rest mass by

$$E = \gamma m_0 c^2 \tag{7.3}$$

Modern large machines such as the electron positron collider (LEP) at CERN, which operated in the 1990's had beams in which the γ was over 80,000. The next generation Large Hadron Collider will collider protons with a γ of over 7000.

To achieve these energies the first step in the acceleration process often uses a LINAC. This is discussed in the section below.

7.2 Linac

A LINear ACcelerator or LINAC is a particle accelerator which accelerates charged particles - electrons, protons or heavy ions in a straight line.

Charged particles enter on the left and are accelerated towards the first drift tube by an electric field. Once inside the drift tube, they are shielded from the field and drift through at a constant velocity. When they arrive at the next gap, the field accelerates them again until they reach the next drift tube. This continues with the particles picking up more and more energy in each gap, until they exit the accelerator on the right.



Figure 114: Schematic of a Linac

The drift tubes are necessary because an alternating field is used and without them, the field would alternately accelerate and decelerate the particles. The drift tubes shield the particles for the length of time that the field would be decelerating Below is a photo of a small Linac (Figure 115).



Figure 115: Photograph of a small Linac.

The linac is ~ 2.5 meters long--not a great distance in which to get even an electron from zero to almost 300,000 kilometres per second. Below is a simplified drawing that shows how this linear accelerator works showing the electron source.



Figure 116: More detailed schematic of Linac showing the electron gun and buncher

7.2.1 The Electron Gun

The electron gun located at the left in the drawing, is where electron acceleration begins. The electrons start out attached to the molecules in a plate of barium aluminate or another thermionic materials such as thorium. This is the cathode of the electron gun. A cathode is a surface that has a negative electrical charge. In linac electron guns this charge is usually created by heating the cathode. Barium aluminate being "thermionic" means that it's electrons tend to break free of their atoms when heated.

The gate is like a switch. It consists of a copper screen, or "grid," and is an anode. An anode is a surface with a positive electrical charge. Every 500 millionth of a second the gate is given a strong positive charge that causes electrons to fly toward it from the cathode in tremendous numbers. As these electrons reach the gate, they become attracted even more strongly by the main anode, and pass through the gate.

Because the gate is pulsing at a rate of 500 million times per second (500 MHz), the electrons arrive at the anode in loose bunches, a 500 millionth of a second apart. The anode is a torus (a doughnut) shaped to create an electromagnetic field that guides most of the electrons through the hole into the next part of the accelerator, called the buncher.

The key part of an electron gun is an anode with a hole in it. Behind the anode was another plate with a hole in it, the holes being aligned. This second plate is usually called the suppressor, and the key thing to making this work is that the suppressor is either grounded or at a voltage opposite in polarity to the anode. So now the electrons coming off of the anode are accelerated and those that make it through the hole in the anode immediately pass through the hole in the suppressor as well. This suppressor (which can be a plate with a hole in it or a grid) shields the electrons that passed through the hole from the anode and a rough beam is formed. Now, generally a fine (collimated) electron beam is required so a second anode/suppressor section is used. As before, not all the electrons will hit this second hole but those that do tend to be well collimated. This may be repeated any number of times and the end result will be a high velocity, finely collimated beam of electrons which is generally steered The only limitations on this technique are physical size and magnetically. voltage breakdown between the various plates.

7.2.2 Electron sources

To obtain good characteristics, the electrons, in general, need to be emitted from a well defined surface in a controlled manner. The actual design of an electron gun is mainly a function of the use of the required beam and in general is amenable to computer simulation. Only the basics of electron emission will be dealt with here; the formation of the beam is dealt with in specialist texts.

Thermionic emission is the escape of electrons from a heated surface. Electrons are effectively evaporated from the material. To escape from the metal, electrons must have a component of velocity at right angles to the surface and their corresponding kinetic energy must be at least equal to the work done in passing through the surface i.e. the workfunction.. If the heated surface forms a cathode, then at a given temperature T ($^{\circ}$ K) the maximum current density emitted is given by the Richardson/Dushman equation:

$$J = AT^2 e^{-11600\phi/T} \tag{7.4}$$

where ϕ is the work function (eV) and A is a constant with a theoretical value of 120 A/cm2.K. In reality this value is not attained for real materials. Table 3 illustrates the basic characteristics of some thermionic emitter materials that are commonly used. It can be seen that the most important parameter for thermionic emission is that the work function as should be as low as possible to use a cathode at an acceptable temperature. The mixed oxide cathode is commonly found in small radio type valves. Cs/W/O, although not good for thermal emitters, is usually found in photo-tubes whilst the heavy metal cathodes are used in high power electron tube devices.

Material	А	φ	Temp (° K)	J (A/cm2)
Tungsten	60	4.54	2500	0.3
Thoriated W	3	2.63	1900	1.16
Mixed oxides	0.01	1.	1200	1.
Cesium	162	1.81		
Tantalum	60	3.38	2500	2.38
Cs/O/W	0.003	0.72	1000	0.35

Table 5: Important characteristics of some thermionic emitter materials

In a diode structure, electrons leaving the cathode surface lower the electric field at the surface. A stable condition exists when the field is zero as any further reduction would repel electrons back to the cathode. This stable regime is known as 'space-charge-limited emission' and is governed by the Child Langmuir equation:-

$$J = PV^{3/2}$$
(7.5)

where P, a constant which is a function of the geometry of the system, is known as the perveance. However, if the voltage becomes sufficiently high, the Richardson limit for current is reached when the emission becomes temperature limited. Figure below shows the characteristics of an ideal diode



Figure 117: Thermionic emission regimes

Thermionic emitters are used in electron tubes and in specialist electron guns, as for example in klystrons, welding, industrial materials processing and in accelerators for lepton production. Figure 16 shows a computer simulation of an electron gun used for hadron beam cooling.



Figure 118:Computer simulation of an electron gun

7.2.3 High field emission

The application of a high voltage between a fine point cathode and a contra surface can, by a tunnelling effect, give sufficient energy to an electron so that it escapes from the surface. This phenomena is known as high-field Fowler/Nordheim emission. It should not be forgotten that the electric field around a point is greatly enhanced relative to the apparent average electric field between the electrodes.

With fields of the order of 10^9 V/m, current densities can attain 10^{12} A/m2 but the actual current is quite small due to the small surface of the emitter. More reasonable currents can be obtained by multiplying the emitter sites. Needles or razor blades can be used as emitter arrays and arrays etched in silicon have shown some success in electron tubes. The major disadvantage of this type of source is that an excessive current density can destroy the points either by erosion or self heating.

7.2.4 Photo emission

Photons illuminating a metal surface may also liberate electrons. If the photon has an energy at least equal to the work function, then electrons will be emitted, i.e.:

$$\lambda < \frac{hc}{e\phi} \tag{7.6}$$

where λ is the wavelength of the incident light, c the velocity of light and h Plank's constant. For shorter wavelengths the electrons are emitted with an initial velocity given by

$$\frac{1}{2}mv^2 = hv - e\phi \tag{7.7}$$

but in general these velocities are low. To obtain reasonable emission with normal wavelengths, a low work function material is needed, for this reason the Cs/O/W material mentioned earlier is often used in photo tubes. Intense electron beams require intense light sources, and lasers have been used to obtain very short high intensity electron beam pulse trains intended for the generation of microwave power in future linear colliders.

7.2.5 The Buncher

The purpose of the buncher is to accelerate the pulsing electrons as they come out of the electron gun and pack them into bunches. To do this the buncher receives powerful microwave radiation from the klystron. The microwaves accelerate the electrons in somewhat the same way that ocean waves accelerate surfers on surfboards.



Figure 119: Field produced by buncher

Figure 119 represents the microwave radiation in the buncher The electrons receive more energy from the wave--more acceleration--depending on how near they are to the crest of the wave, so the electrons riding higher on the wave catch up with the slower ones riding lower. The right-hand wave shows the same group of electrons a split second later. On the front of the wave, the two faster electrons have almost caught up with the slower electron. They won't pass it though, because they are now lower on the wave and therefore receive less acceleration.

DEVICES IN PARTICLE PHYSICS

The higher electron on the back of the wave gets just enough acceleration to match the speed of the wave, and is in the same position as it was on the lefthand wave. This represents the last electron in the bunch. The lower electron on the back of the wave gets too little energy to keep up with the bunch and ends up even lower on the right-hand wave. Eventually it will fall back to the electron bunch forming one wave behind.

7.2.6 Acceleration

The linac itself is just an extension of the buncher. It receives additional RF power to continue accelerating the electrons and compacting them into tighter bunches. Electrons enter the linac from the buncher at a velocity of 0.6c. By the time the electrons leave the linac, they are travelling very close to the speed of light.

Linacs can be very long and may be used to accelerate light particles, which would lose energy through bremsstrahlung in a circular orbit.

The major structure of the particle accelerator is the copper tube. The copper tube has a strong vacuum inside through which the particles travel. The tubes are made of copper because copper conducts electricity and magnetism very well. At the SLAC linac, the copper tube is made of more than 80,000 copper cylinders brazed together for more than 2 miles (3.2 km).



Figure 120: Cavities in the copper tube

The copper tube at SLAC is arranged to form a series of cells called cavities. The spacing of the cavities is matched to the wavelength of the microwaves. The spacing allows the electric and magnetic fields to repeat their pattern every three cavities. Electrons or positrons in the beam come through the cavities in small bunches. The arrival of each bunch is timed so that it gets a push from the electric field across the cavities.



1/20,000,000,000 second later (notice how far the bunches have moved)

Figure 121: Bunches within cavities

The klystrons that make the microwaves, much like a kitchen microwave oven except that the klystrons' microwaves are about 1 million times more powerful. Klystrons produce microwaves by way of an electron gun. The electrons travel through the klystron in cavities, where their speed is regulated. As the electrons change speed in the klystron, they give off radiation in the form of microwaves. The microwaves are conducted through copper waveguides to the copper tube of the accelerator. Waveguides carry waves efficiently without losing intensity. The klystron and waveguides are kept under high vacuum to ease the flow of the waves.



Figure 122: Series of Klystrons

7.3 Cyclotrons

Circular accelerators do essentially the same jobs as linacs. However, instead of using a long linear track, they propel the particles around a circular track many times. At each pass, the magnetic field is strengthened so that the particle beam

DEVICES IN PARTICLE PHYSICS

accelerates with each consecutive pass. When the particles are at their highest or desired energy, a target is placed in the path of the beam, in or near the detectors. Circular accelerators were the first type of accelerator invented in 1929. In fact, the first cyclotron (shown below) was only 4 inches (10 cm) in diameter.



Figure 123 Photo from Lawrence Berkeley National Laboratory The first particle accelerator (cyclotron) developed by Ernest O. Lawrence in 1929

Lawrence's cyclotron used two Dshaped magnets (called Dees) separated by a small gap. The produced circular magnets a magnetic field. An oscillating voltage created an electric field across the gap to accelerate the particles (ions) each time around. As the particles moved faster, the radius of their circular path became bigger until they hit the target on the outermost circle. Lawrence's cyclotron was effective, but could not reach the energies that modern circular accelerators do.



Figure 124:Schematic of Cyclotron

Modern circular accelerators place klystrons and electromagnets around a circular copper tube to speed up particles. Many circular accelerators also have a short linac to accelerate the particles initially before entering the ring.



Figure 125 A large cyclotron magnet at the Lawrence Hall of Science. The black portions are the iron pole pieces and continue across below ground level. The electromagnet coils were within the white cylinders. The vacuum chamber would be placed in the horizontal gap between the poles of the magnet



Figure 126 A pair of Dee electrodes at the Lawrence Hall of Science. These would be contained within the vacuum chamber. The serpentine pipes are for cooling liquid.

7.3.1 Mathematics of the cyclotron

The centripetal force is provided by the transverse magnetic field B, and the force on a particle travelling in a magnetic field (which causes it to curve) is equal to Bqv. So,

$$\frac{mv^2}{r} = Bqv \tag{7.8}$$

(Where m is the mass of the particle, q is its charge, v is its velocity and r is the radius of its path.)

Therefore,

$$\frac{v}{r} = \frac{Bq}{m} \tag{7.9}$$

v/r is equal to angular speed, ω , so

$$\omega = \frac{Bq}{m} \tag{7.10}$$

Therefore the frequency is

$$f = \frac{Bq}{2\pi m} \tag{7.11}$$

This shows that for a particle of constant mass the frequency does not depend on the radius of the particle's orbit. As the beam spirals out its frequency does not decrease and it must continue to accelerate, as it is travelling more distance in the same time. As particles approach the speed of light they acquire additional mass, requiring modifications to the frequency or the magnetic field during the acceleration. This is accomplished in the synchrocyclotron

7.3.2 Limitations of the cyclotron

While a significant technical achievement at the time, the configuration of the device limits its cost effectiveness at higher power. These limitations were addressed with the invention of the synchrotron.

7.4 Synchrotrons



Figure 127:The CERN PS. Still operating after 40 years

A synchrotron (sometimes called a synchro-cyclotron) is а circular accelerator which has an electromagnetic resonant cavity (or perhaps a few placed at regular intervals around the ring) to accelerate the particles.

Thus, as the particles increase in energy the strength of the magnetic field that is used to steer them must be changed with each turn to keep the particles moving in the same ring. The change in magnetic field must be carefully synchronized to the change in energy or the beam will be lost. Hence the name "synchrotron". The range of energies over which particles can be accelerated in a single ring is determined by the range of field strength available with high precision from a particular set of magnets. To reach high energies, physicists sometimes use a sequence of different size synchrotrons, each one feeding the next bigger one.

7.4.1 Synchrotron Radiation

Synchrotron radiation is the name given to the electromagnetic radiation emitted by the charged particles circulating in a synchrotron. It occurs because the charged particles are accelerated (deflected) by the magnetic field from the dipole magnets to make the beam travel around the ring. Any accelerated charged particle produces some electromagnetic radiation.

The wavelength and intensity of the synchrotron radiation depends on the energy and type of the emitting particle. If all you are interested in is storing a high energy beam, then synchrotron radiation is a problem. The energy lost from the beam by this radiation effect must be restored by introducing accelerating cavities at one or more places in the ring, to give the particles a kick in energy every time they pass. The amount and energy of the radiation depends on the speed of the radiating particles and the magnetic field strength. As the particle approaches the speed of light, the effect increases rapidly. The energy loss for a

given electron energy is proportional to γ^3 .

For an 1.5 GeV electron in the SPEAR (at SLAC) storage ring, gamma is approximately 3000. For a 50 GeV electron in the SLC arcs, gamma is approximately 100,000. Thus, because a proton is so much more massive than an electron, a proton with 1 TeV = 1,000 GeV energy has a gamma factor of only 1,000. (1 TeV is the energy produced by the synchrotron at Fermilab. Thus synchrotron radiation is much greater for electrons than for equal energy protons. This is the reason why much higher energy synchrotrons can be built for protons than for electrons.

7.4.2 Focussing and Optics

As beams of particles are accelerated round a ring they must be focussed, otherwise due to dispersion, gravity and other effects they would be rapidly lost. Focussing of charged particles is done with quadrupole magnets. These have two North poles and two South poles and the magnetic field in the centre is zero.

Alternating focus/defocus/focus/defocus ... in a series of quadrupole magnets is called alternating gradient or strong focusing. This method of



Figure 128: Field in a quadrupole

keeping the beam particles tightly constrained inside a vacuum chamber is used in all synchrotrons, and also in the many beam lines.



Figure 129: Series of focussing defocusing quadrupoles

N

7.4.3 Storage Rings

One problem with a simple synchrotron is that of we wish to have collisions between particles at a fixed energy it is desirable to "store" the particles we have accelerated to maximize their lifetime in the ring (at constant energy) before they suffer collisions. This is the role of a storage ring. In essence it has all the properties of a synchrotron except that rather than accelerating particles it simply injects energy into the circulating beam to top up the energy lost bv bremsstrahlung. The total number of particles in the storage ring will be fed by a synchrotron.

N CESR The ring is nearly 1/2 mile around and is buried 40 ft deep. The electrons and positrons go a million times around the ring every 2.5 seconds! In



Figure 130: Schematic of Cornell Electron Storage Rings

Figure 131 we show picture of the tunnel for both CESR and the Synchrotron (looking away from CLEO). On the left is CESR, the Cornell Electron Storage Ring. The large red boxes are the CESR dipole magnets that bend the path of electrons and positrons to keep them going in a circle around the ring. The blue boxes are the CESR quadrupole magnets that focus the beams so that particles do not stray too far from the desired path. There may also be sextapole magnets behind the quadrupoles for further focusing, but they are hard to see in this picture. If you look carefully, you can see that dipole-quadrupole (and sextapole) magnet structure is repeated as far as can be seen. On the right is the Synchrotron accelerator where the electrons and positrons are accelerated to ~5 GeV before transfer to the storage ring. Most of the magnets on the Synchrotron are dipoles.

Since focusing isn't so important for that accelerator, there are only a few quadrupoles.



Figure 131: Photograph of the CESR ring

7.5 Accelerator Facilities

Below is a list of a few of the major accelerator facilities in the world for high energy physics.

- CESR (Cornell, USA), e⁺e⁻ at 10GeV (symmetric)
- DESY (Hamburg, Germany), ep asymmetric (27 and 920) GeV
- PEP2 (SLAC, Stanford USA) e^+e^- at 10GeV (asymmetric)
- Tevatron (FNAL, Chicago, Illinois) p-pbar at 2TeV
- LHC (CERN, Geneva, Switzerland) pp at 7TeV/beam

Other facilities at Brookhaven, Beijing, KEK (Japan) and Russia.

7.5.1 Large Hadron Collider

For the next big step in HEP a new accelerator is needed to make it possible to reach the whole predicted possible Higgs mass range (180-1TeV), to search for supersymmetry. and to probe energies beyond those governed by the Standard Model. Since CERN already played a central role in HEP, not only in Europe but worldwide, and since it would be possible to re-use part of the existing infrastructure, the CERN council decided to build that accelerator at CERN: the LHC. The LHC is going to be the biggest accelerator ever built with high centre of mass energy and high luminosity. It has been designed as a proton-proton collider with a centre of mass energy in that mode of 14TeV. Since the LHC is going to be a much bigger accelerator than any one built before several problems have had to be solved during the design process. From the physics point of view it was desirable to increase the beam luminosity. This is because the interaction

DEVICES IN PARTICLE PHYSICS

cross section for particle collisions (some sort of estimate for the likelihood of a specific interaction occurring) goes down as $1/E^2$ as the beam energy goes up. The only way to make sure enough interactions are generated within a reasonable timescale is to increase the number of collisions that can occur; pump up the luminosity. In the LHC that's done by sending more bunches of particles into the pipe and by putting more particles into each bunch. From the engineering side the biggest challenge was fitting the new accelerator in the existing LEP tunnels as much as possible. Another big achievement was the design and development of the superconducting magnets and the cooling and cryostats involved with them. Not only is it a lot cheaper than starting from scratch, another big advantage of putting the LHC in the old LEP tunnels is that it can be fed by the same particle sources and pre-boosters as LEP was. The whole process involved in getting high-energy protons into the LHC beampipe looks a bit like a longdistance train ride with many transfers. It starts with protons being generated in the duoplasmatron. Subsequently they are accelerated to the 'proton linear accelerator' and after that they will be boosted to about 1.4GeV in the 'Proton Synchrotron Booster' (PSB). The 'Proton Synchrotron' (PS) will then accelerate them to an energy of 26GeV and just before they finally enter the LHC they will reach an energy of about 0.45TeV in the 'Super Proton Synchrotron' (SPS). The LHC will give them the final boost till they reach the designed beam energy of 7TeV. This process is also depicted in



Figure 132: Schematic representation of the acceleration proces of protons in the LHC At the moment there are four detectors under construction for use in the LHC:

- ATLAS,
- the 'Compact Muon Solenoid' (CMS),
- the LHCb,
- Alice.

We will discuss these detectors in a little more detail when we cover the application of semiconductor devices in their design.

Both ATLAS and CMS are general purpose detectors whose main goal is to find experimental evidence for the Higgs boson and to investigate physics beyond the Standard Model.

The LHCb experiment is being designed to run at the 'B-factory end' (lower energy) of the LHC. Its aim is to study CP-violation using B-mesons, both for reproducing results from other B-factories (KEK, BABAR) around the world and for measuring new parameters involved in CP-violation. To do that it will run at lower luminosities than the other detectors by slightly defocussing the beams locally.



Figure 133: An arial view of the CERN site showing the outline of the LHC. Note the airport in the foreground

7.6 Detectors

We have discussed how particles may be accelerated and controlled and collided. Modern detectors consist of many different pieces of equipment which test for different aspects of an event. These many components are arranged in such a way that physicists can obtain the most data about the particles spawned by an event.

When protons collide, some events are "interesting" and may tell us about exciting new particles or forces, whereas many others are "ordinary" collisions (often called "background"). The ratio of their relative rates is about 1 interesting event for 10 million background events. One of our key needs is to separate the interesting events from the ordinary ones.



Figure 134 Two figures showing the simulation of a Higgs event at ATLAS. Data from the UA1 experiment (CERN) showing one of the first W candidates discovered.

The differentiation between these is based on the observed products of each collision...their identities, energies, directions of motion etc. For example, it may be possible to demonstrate that some observed configurations of outgoing collision products arise from the decay of a new particle. Such observations would then represent the discovery of this new particle. The instrument that provides the detailed information on the collision products is called the detector.

The purpose of a detector is to provide information about the collision products for each proton-proton collision. That information must be obtained and analyzed very rapidly. Only 10 to 100 of the billion collisions that occur each second must be flagged as potentially interesting and recorded for further study, while all the others are rejected.



Figure 135Photo of the CDF detector at Fermilab being opened

The detector must have the capability of processing the information fast enough to permit flagging about 10-100 potentially interesting events per second out of the billion collisions that occur and recording the measured information for just those events. The rejected collisions are known as "background". This flagging process is called "triggering", and is based on a rapid analysis of part of the measured information to reject most background events, followed by two additional steps of filtering, each requiring more sophisticated and timeconsuming analysis than the previous step, to produce the overall rejection factor of 10 million to 100 million

Most Triggers area a combination of dedicated hardware and large cluster processing farms containing many thousands of CPU's.

7.6.1 Detector Function

The following list the basic requirements of a modern general purpose detector

- Measure the directions, momenta, and signs of charged particles.
- Measure the energy carried by electrons and photons in each direction from the collision.
- Measure the energy carried by hadrons (protons, pions, neutrons, etc.) in each direction.
- Identify which charged particles from the collision, if any, are electrons.
- Identify which charged particles from the collision, if any, are muons.
- Identify whether some of the charged particles originate at points a few millimetres from the collision point rather than at the collision point itself (signalling a particle's decay a few millimetres from the collision point).
- Infer (through momentum conservation) the presence of undetectable neutral particles such as neutrinos.
- Have the capability of processing the above information fast enough to permit triggering about 10-100 potentially interesting events per second out of the billion collisions per second that occur, and recording the measured information.
- The detector must also be capable of long and reliable operation in a very hostile radiation environment.

7.6.2 Detector Structure



Figure 136: Configuration of fixed target and colliding beam experiments

Physicists are curious about the events that occur during and after a particle's collision. For this reason, they place detectors in the regions which will be showered with particles following an event. Detectors are built in different ways according to the type of collision they analyze.

7.6.3 Fixed Target

With a fixed-target experiment the particles produced generally fly in the forward direction, so detectors are somewhat cone shaped and are placed "downstream" from the collision. However, modern high energy detectors like Atlas use colliding beams and are somewhat cylindrically shaped.

7.6.4 Colliding Beams

During a colliding-beam experiment, the particles radiate in all directions, so the detector is spherical or, more commonly, cylindrical.

7.7 Components of a Detector

The reason that detectors are divided into many components is that each component tests for a special set of particle properties. These components are stacked so that all particles will go through the different layers sequentially. A particle will not be evident until it either interacts with the detector in a measurable fashion, or decays into detectable particles.

The interaction of various particles with the different components of a detector:



Figure 137: Schematic of detector structure Neutrinos are not shown on this chart because they rarely interact with matter, and can only be detected by missing matter and energy

A few important things to note:

- *Charged particles*, like electrons and protons, are detected both in the tracking chamber and the electromagnetic calorimeter.
- *Neutral particles*, like neutrons and photons, are not detectable in the tracking chamber; they are only evident when they interact with the detector. Photons are detected by the electromagnetic calorimeter, while neutrons are evidenced by the energy they deposit in the hadron calorimeter.

Each particle type has its own "signature" in the detector. For example, if a physicist detects a particle only in the electromagnetic calorimeter, then he is fairly certain that he observed a photon.

7.7.1 Tracking Detector

The Tracking Detector measures the directions, momenta, and signs of charge of electrically-charged particles produced in each proton-proton collision. It consists of up to three different systems of sensors normally all immersed in a magnetic field. In a modern detector these will typically be a pixel layer(s), strip

detectors and a large area/volume tracker. At least the first two are almost always semiconductor devices. We consider general forms of the resolution obtained from these types of devices in the last section in this chapter.

Pixel Detectors

The sensors closest to the collision point are the pixel detectors. These devices consist of thin layers of silicon subdivided into rectangular regions ("pixels") of typically dimensions 50 by 50 microns.

Each time a charged particle traverses such a layer, a signal is produced that identifies which pixel has been traversed, and thereby gives a precise measure of particle position. Indeed this position is precise enough to determine whether the particle originated at the collision point, or a few millimetres from it as a decay product of another particle.



Figure 138: The solid state (pixels and strips) tracker in the ATLAS detector

Strip Detectors

To provide additional position measurements a little further from the collision point (to best determine a full trajectory), additional layers of silicon subdivided into narrow strips. Typically from about 40-100 microns wide by several centimetres long, are used When a charged particle goes through the strip detector, signals identify which strip in each set has been traversed. The intersection of those two struck strips provides a 3-dimensional position measurement, very precise around the cylinder or disk and much less so in the other direction of the sensors. You may wonder why one simply does not put in more pixel detectors instead of the strip detectors: For larger detector areas, the pixel devices are simply too expensive.

Large Volume Tracker

Larger tracking capability is sometimes provided by gas chambers, either Time Project Chambers, Drift Chambers or Straw Chambers. These very low mass, relatively low cost detectors may still have very high resolution, better than 30 microns/point in many cases, but are generally quite slow compared with the readout times of Si strips. The large volume trackers are responsible for the pattern recognition and the momentum resolution of particles within the experiment.

Generally a tracker differs from a vertex finding device by having more measuring points over a larger volume (at least 5 or 6 with resolutions of order 20-50 microns, but of less precision than the vertex detector (typically 2 or 3 measuring points with resolutions of 10 microns or better).

Large volume trackers must also be constructed to have much less mass /per measuring point than to avoid creating particles "showers" which would destroy both its own track finding capability but also "neutral" energy reconstruction in the calorimeters.

Magnet Field

Large electro-magnets, often superconducting, are use to produce magnetic field, nearly 100,000 times stronger than the Earth's. This field is parallel to the beam axis if a solenoid magnet is used, or perpendicular if a dipole is used. The magnet deflects each charged particle coming from the collision point.

A solenoid magnet may be used as follows. After a particle emerges perpendicular to the beam, it continues perpendicular and travels in a circle whose radius is proportional to its momentum. The paths of very slow particles can "curl up" within the detector, whereas those of very fast particles have very large radii and leave the detector (unless absorbed or deflected). The direction of rotation around the magnetic field (clockwise or counter-clockwise) indicates the sign of charge of the particle. If the particle is not perpendicular to the beam, the field changes the trajectory to a circular helix with axis parallel to the beam line. The radius of the helix is proportional to the momentum of the particle.

Dipole magnets are often used in an experiment when the forward going particles need to be resolved.

Electromagnetic Calorimeter

The Electromagnetic Calorimeter absorbs the energies of all electrons and photons traversing it (this constitutes the "electromagnetic energy"), and produces signals proportional to those energies. It is finely subdivided so that it can measure the directional dependence of the electromagnetic energy.

The typical e-m calorimeter consists of thin lead plates (about 1.5 mm thick) separated by sensing devices. When high energy photons or electrons traverse the lead (or any high-atomic-number material), they produce an electron shower. What happens is that their initial energies are transformed into the rest masses of numerous low energy (but still fast-moving) electrons and positrons. The number of such electrons/positrons is proportional to the incident energy, and their presence is detected by a sensing system between the lead plates. Electromagnetic calorimeters of this kind may have an energy resolution of

$$\sigma_{em} \sim \frac{10\%}{\sqrt{E}} \tag{7.12}$$

Higher resolution e-m calorimeters work by absorbing photon/electrons into a crystal, such as CsI, and measuring precisely the light that is re-emitted.

Hadronic Calorimeter

The hadronic calorimeter is placed after the electromagnetic calorimeter. It absorbs and measures the energies of hadrons, including protons and neutrons, pions and kaons (electrons and photons have been stopped before reaching it). Hadronic calorimeters usually consist of steel absorbers separated by tiles of scintillating plastic. Interactions of high energy hadrons in the plates transform the incident energy into a "hadronic shower" of many low energy protons and

neutrons, and other hadrons. This shower, when traversing the scintillating tiles, causes them to emit light in an amount proportional to the incident energy.

The newly created low-energy hadrons and the knocked out protons and neutrons that get out of the plates are detected by the sensing devices, and the total signal is proportional to the incident energy. Since hadrons can initiate their shower in the electromagnetic calorimeter, the signals from both calorimeters must be combined to get the full hadronic energy. Electrons and photons are already absorbed in the electromagnetic calorimeter, so they do not reach the hadronic calorimeter.



Figure 139: Part of the ATLAS hadronic tile calorimeter being constructed.

Muon System

Muons are the only charged particle that can travel through all of the calorimeter material and reach the outer layer. The Muon System determines the signs and momenta of muons with better precision than the inner tracking system does. It is able to measure momenta even at the highest luminosities.

Muons interact with matter almost entirely through their electric charge. However, because they are about 200 times more massive than electrons, they are much less affected by the electric forces of the atomic nuclei that they encounter. They therefore do not produce the same kind of electromagnetic shower.

They lose energy almost solely by the formation of electron-ion pairs along their path, and for a substance like steel or copper, this amounts to an energy loss of about 1 MeV per millimetre of path. Thus muons with energy above, say, 5 GeV will penetrate about 5 meters of steel, whereas hadrons of almost any energy are completely absorbed in about 1.5 meters of steel. Thus energetic particles seen outside the hadron calorimeter are guaranteed to be muons.

7.8 Vertex Resolution

We have discussed the overall layout of a modern particle physics experiment. An extremely important application of semiconductor devices is the vertexing and tracking detectors; the high precision core of most experiment. Below we provide simple arguments that dictate the resolution that may be obtained.

7.8.1 Impact Parameter

Consider a simple two plane measuring device (see Figure 140)



Figure 140: Schematic of a simple two plane vertex detector. A particle is assumed to emanate from (0,0) at a small angle to the perpendicular of the two planes positioned at $x_1=d_1$ and $x_2=d_2$. The point intersection with the two planes is recorded as y_1 and y_2 , and x_1 and x_2 are known. A straight line may be reconstructed through the two space points. In general the line will "miss" the true generation point (0,0) by a distance, b known as the impact parameter.

In two dimensions the passage of an ionizing particle through the two planes is recorded and an impact parameter, the point of closest approach to the nominal generation point, *b* calculated.

If we write the equation of the line connecting the two points as y = mx + c (7.13)

the values of m and c may be shown to be

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$c = \frac{y_1 x_2 - x_1 y_2}{x_2 - x_1}$$
(7.14)

The point of closest approach is given by the intersection of this line with its perpendicular going through the origin

$$y = -\frac{1}{m}x\tag{7.15}$$

This point is

$$x_b = -\frac{cm}{m^2 + 1}, y_b = \frac{c}{m^2 + 1}$$
(7.16)

Which yields

$$b = \frac{c}{\sqrt{m^2 + 1}} = \frac{y_1 x_2 - x_1 y_2}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$
(7.17)

The aim of a vertex detector is to make a precise and accurate measurement b.

If we assume we have close to normal impact in general one of the dimensions (the spatial separation of the planes) will be much greater than the others e.g. $x_2 - x_1 \gg y_2 - y_1$. In this case

$$b \sim \frac{y_1 x_2 - x_1 y_2}{x_2 - x_1} \tag{7.18}$$

References Problems



PHOTODETECTORS

The properties of simple photodectors are described. Photodectors allow the conversion of photons in the optical or near optical regions into electrical signals. Applications of photodetectors as detectors in particle physics are discussed.

8.1 Photons

Photodetectors perform the function of converting optical photons into electrical signals. The process within the semiconductor during this conversion is essentially the carrier generation by incident light followed by the sensing of these carriers. Below we discuss first how this conversion takes place and then the techniques used to detect the electrical signal.

8.1.1 Absorption

When light falls onto a semiconductor it may be absorbed by several different processes depending on the wavelength of the light and the bandgap of the material.

We note here that the wavelength of the light is related to the energy of the photon by

$$\lambda = \frac{c}{\nu} = \frac{1.24}{E_{\text{photon}}} \,\mu\text{m} \tag{8.1}$$

If the energy of the photon matches the band gap in the semiconductor (Figure 141(a)) then an electron hole pair may be created at the edge of the valence and conduction bands, this is an intrinsic absorption. If the energy of the photon exceeds that of the bandgap Figure 141(b)) a photon may again be absorbed via an intrinsic transition but to an excited state. This excited state will in general deexcite releasing energy into the crystal. If the energy of the photon is less than that of the bandgap the photon will not be able to make a direct transition and absorption may only process either by an extrinsic transition where additional



phonon energy is gained from the lattice or to a mid-bandgap impurity state Figure 141(c)).

Figure 141: Absorption of a photon with (a) $hv=E_g$, (b) $hv>E_g$, (c) $hv<E_g$

The absorption as a function of depth in the semiconductor will be typified by an absorption coefficient

$$\Phi(x) = \Phi_0 e^{-\alpha x} \tag{8.2}$$

where the equation above gives the flux as a function of the depth (x) into the seminconductor. The absorption coefficient for Si is shown in Figure 142. Note that when the energy of the photon is less than the bandgap of Si (1.1eV) the absorption coefficient drops dramatically. Thus Si starts to become transparent to photons at wavelength of about 1 μ m.



Figure 142: Absorption coefficient of photons in Si.(1&2 Sze 1981 and Jellison and Modine [1982]).
8.2 Photoconductor

The simplest photoconductor is composed of piece of semiconductor with two ohmic contacts attached, Figure 143



Figure 143: Schematic of a photoconductor, showing a semiconductor sandwiched between two ohmic contacts

The light falling on the semiconductor will generate additional carriers in the material which, if a field is placed across the bulk of the photoconductor, will give rise to a current that may be measured.

8.2.1 Operation

If a photoconductor is illuminated with a monochromatic light source of total power P on the area A where (8.5)

$$A = w \times L$$

(See Figure 143) then the number of photons arriving per second on the photoconductor is

$$n_{gen} = P / hv \tag{8.3}$$

where the energy of each photon is hv.

Assuming that the thickness is much thicker than the absorption length i.e.

$$d \gg 1/\alpha \tag{8.4}$$

then the number of electrons (holes) generated per second per unit volume will be equal to

$$G = \eta \frac{P}{hv} \cdot \frac{1}{Lwd}$$
(8.5)

where η is the quantum efficiency for converting photons to electrons. In the steady state the number generated per second must equal the number that are recombining per second i.e.

$$G = n/\tau \tag{8.6}$$

If we apply a voltage, V, across the photoconductor this will induce an electric in the bulk given by

$$\mathcal{E} = -\frac{V}{L} \tag{8.7}$$

The current in the photoconductor will be

$$I_p = \sigma \mathcal{E} w d = q \mu_n n \mathcal{E} w d = (q n v_d) w d$$
(8.8)

Where the photocurrent I_p is expressed in terms of the drift velocity of the electrons which have a carrier density, n. Substituting n from Eq. (8.5) and Eq. (8.6) yields

$$I_p = q \left(\eta \frac{P}{hv} \right) \left(\frac{\mu_n \tau \mathcal{E}}{L} \right)$$
(8.9)

Thus defining the "primary" photocurrent as

$$I_{ph} = q \left(\eta \frac{P}{hv} \right) \tag{8.10}$$

the ratio (gain) of the current to the primary photocurrent is

$$gain = \frac{I_p}{I_{ph}} = \frac{\mu_n \tau \mathcal{E}}{L} = \frac{\tau}{t_r}$$

Where the carrier transit time is

$$t_r = \frac{L}{v_d} \tag{8.11}$$

8.2.2 Limitations

Photoconductors are limited by the dark current in the active region. To get extended wavelength sensitivity impurities may be added so the detector is more extrinsic in its behaviour. However these impurities add to the dark current in the sensor hence limiting the sensitivity of the sensor.

Another limitation is the recombination in the sensitive region. One solution to this is the use of Blocked Impurity Band detectors.

8.3 Photodiode

Diodes used as Photodetectors (photodiodes) offer another solution to improving the behaviour of photoconductors. Photodiodes offer:

- High impedance through the depletion region
- High impedance at elevated temperatures, compared with photoconductors
- Little recombination noise
- Great gain that allows photon counting.

8.3.1 Principle of operation

One of the simplest kind of photodiodes is the p-i-n photodiode in which an intrinsic piece of semiconductor is sandwiched between two heavily (oppositely) doped regions. At the metallurgical junctions charges migrate from the heavily

doped regions to fill any vacancies or accept any free electrons. A depleted region is then set up within the Si. The two charge sheets (on the n+ and p+) sides produce a field which, even without an external field supplied, will tend to separate charges produced in the depleted region. The separated charges will be swept to either terminal and be detected as a current provided that they are not recombined.



Figure 144: Schematic of a p-i-n photodiode.

8.3.2 Construction

A typical photodiode may be constructed as shown above, *Figure 144*. A thin ptype region is implanted into an n-type bulk. The n+ region provides the ohmic contact as well as additional carriers. A depletion region develops as shown above.

8.4 Avalanche Photodiodes

If a high enough reverse bias is applied to the photodiode avalanche multiplication may take place. The multiplication will result in substantial gain in the diode. A larger reverse-bias voltage results in a larger gain. It should be noted that increasing the reverse-bias voltage also results in increased noise levels. The main advantage of the APD over the p-i-n diode is the gain.

Large excess noise resulting from the avalanche multiplication process places a limit on the useful gain of the APD. The noise arises because every photogenerated carrier does not undergo the same multiplication.

8.5 Applications

Photosensitive devices are used in many HEP applications. They may be used to detect light various sources e.g.

- Calorimeters (e.g. CMS experiment at CERN)
- Crystals (CLEO at Cornell, Babar SLAC)
- Radiation Sensors
- Optical Coupling for decoupling electrical signals

References

http://www.ioffe.rssi.ru/SVA/NSM/Semicond/Si/optic.html http://sales.hamamatsu.com/assets/applications/SSD/CERN_award_mayer.pdf **Problems**



STRIP DETECTORS

An introduction to strip based sensors using reverse based diodes is given. Simple overview of the charge collection mechanisms in these sensors is provided, together with a brief description of the fabrication techniques for strip detectors. The small energy required to create electron-hole pairs compared with the energy lost per unit length by an ionizing particle make Si an excellent material out of which to construct position sensitive detector.

9.1 Simple Sensors

ne of the simplest detectors to build is a reverse biased diode. In Figure 145 a schematic of a reverse biased diode is shown. If the detector is depleted then effectively the whole of the bulk is devoid of free carriers. The only ones that will be generated in thermal equilibrium are, to first order, those from thermal excitation (that form part of the dark current) if the diode is not exposed to any ionizing radiation including light.

9.1.1 Charge Collection

If the diode is kept unilluminated the passage of ionizing radiation through the bulk will cause electron-hole pairs to be created. Holes will drift towards the negative potential on the p^+ implant and electrons towards the positive potential on the ohmic contact.

The field within the depletion volume of a fully depleted diode was given in chapter 3 and, in the broad n-type depletion region varies linearly

$$\mathcal{E}(x) = \frac{qN_D(x-d)}{\varepsilon}$$
(9.1)

With the depletion voltage being

$$V_{dep} = \frac{qN_D d^2}{2\varepsilon} \tag{9.2}$$

Under the condition the diode remains fully depleted, but not over depleted, we may write





Figure 145: Schematic of a reverse biased n-type diode. This diode has a heavy pimplant on one side and n+ implant to provide ohmic contact on the other. Electrons move towards the p-implant. Holes will drift towards the n⁺ side.

The field within the depletion volume of a fully depleted diode was given in chapter 3 and, in the broad n-type depletion region varies linearly

$$\mathcal{E}(x) = \frac{qN_D(x-d)}{\varepsilon}$$
(9.4)

With the depletion voltage being

$$V_{dep} = \frac{qN_D d^2}{2\varepsilon} \tag{9.5}$$

Under the condition the diode remains fully depleted, but not over depleted, we may write

$$\mathcal{E}(x) = \frac{2(x-d)}{d^2} V_{dep} \tag{9.6}$$

If we apply an overvoltage this becomes

$$\boldsymbol{\mathcal{E}}(x) = \frac{2(x-d)}{d^2} V_{dep} - \frac{V - V_{dep}}{d}$$
(9.7)

As the electron moves towards the n^+ ohmic contact its drift velocity remains connected to the electric field by

$$v_e(x) = -\mu_n \mathcal{E}(x) \tag{9.8}$$

Equations (9.7) and (9.8) give a relatively simple differential equation for the electron velocity as a function of position which can be solved (see for example Lutz) starting a position x_0

$$v_{e}(t) = \mu_{n} \left[\frac{2V_{dep}}{d^{2}} x_{0} - \frac{V + V_{dep}}{d} \right] e^{-2\mu_{n} \frac{V_{dep}}{d^{2}} t}$$
(9.9)

There be a similar differential equation for the hole velocity as a function of time.

$$v_e(t) = -\mu_p \left[\frac{2V_{dep}}{d^2} x_0 - \frac{V + V_{dep}}{d} \right] e^{2\mu_p \frac{V_{dep}}{d^2} t}$$
(9.10)

The current induced by a moving charge with velocity in the circuit is given (see the next section) by

$$i = \frac{q}{d}\frac{dx}{dt} = \frac{qv}{d}$$
(9.11)

Thus a simple expression can be formed that gives the induced current at any time i.e.

$$i(t) = i_e(t) + i_h(t)$$
 (9.12)

This current may be plotted for a sample diode with 20% over depletion, Figure 146. As the hole has a lower mobility it continue to induce a current after the electron has arrived at the plate. The transition time for a charge produced in the middle of in fully depleted diode for both holes and electrons is of the order of

$$t \sim O\left(\frac{d^2}{4\mu V_{dep}}\right) \tag{9.13}$$

i.e. for a 300 micron thick diode at the depletion voltage of 70Volts and a carrier mobility of $1700 \text{cm}^2/\text{Vs}$ the approximately 2ns.



Figure 146: Signal for separation of an electron-hole in a diode. The e-h pair is created in the centre of the diode. The electron current is the dashed line, and hole current the dot-dashed line. Note the hole current is still arriving after the electron current is finished (after Lutz)

9.1.2 Ramo's Theorem

Above we assumed a form for the induced current on one of the electrodes. In fact a general theorem exists (called Ramo's or the Schokley-Ramo Theorem) that in its general form states that the current induced in electrode j due to a moving charge at point P is

$$i_j = q\vec{v}.\nabla\phi_P \tag{9.14}$$

The derivative is that of the electrostatic potential that occurs when we put a unit voltage on electrode j and keep all other electrodes at zero potential and is known as the weighting field.

For the case of the parallel plate capacitor (or diode) we can easily obtain the epression used above, (9.11). In general when we have more complex weighting fields, that it becomes critical to know the electrostatic potential in the detector. Note also than in many cases, Figure 147, that qualitatively the biggest current is induced close to the strips in the high field region. Also if a particle is moving at right angles to the weighting field that little current is induced.



Figure 147: Weighting filed for a strip detector. Note the field is high close to the strip coordinate (at 300 microns), and thus the large currents are induced close to the high field region i.e. strip.

9.1.3 Minimum Ionizing Particles

As an ionizing particle passes through matter it loses energy through elastic collisions with electrons. For particles with momenta between 10MeV/c and 100GeV/c the rate of ionization loss is given approximately by the Bethe-Bloch Equation

$$\frac{dE}{dx} = \frac{DZ^2 n_e}{\beta^2} \left[\ln\left(\frac{2mc^2\gamma^2}{I}\right) - \beta^2 - \frac{\delta(\gamma)}{2} \right]$$
(9.15)

Where x is the distance travelled through the medium, m is the electron mass and Z is the atomic number and

$$D = \frac{4\pi\alpha^2\hbar^2}{m} = 5.1 \times 10^{-25} \,\mathrm{MeV cm^2}$$
(9.16)

For a Si the dE/dx (for protons) is minimum at about 0.4keV which would predict the creation of 110electron holes pairs/micron. The actual figure is measured to be about 80 e-h pairs per micron. The difference between the two being accounted for by energy lost by the incident particle to the lattice via phonon interactions.



Figure 148: dE/dx for protons in Si

For a 300 micron thick sensor we expect of order 24,000e/h pairs to be created on the passage of a minimum ionizing particle through the sensor. Given a collection time of approximately 2ns the peak current generated due the deposition of this charge is approximately given by

$$I \sim \frac{24,000 \times 1.6 \times 10^{-19}}{2 \times 10^{-9}} \sim O(1 \mu \text{A})$$
(9.17)

The dark currents (saturation currents) must be kept at least 2 orders of magnitude below this in order to have a useable detector.

9.1.4 Landau Distribution

The most probable value of the energy deposited per unit length is given by the Bethe-Bloch equation. However each passage through material is the sum of a series of stochastic events and the actual energy deposited is distributed around this mean. The form of this distribution, for thin layers, was first calculated by Landau and given by

$$\phi(\lambda) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} e^{s\log(s) + \lambda s} ds$$
(9.18)

The most probable value is c, and λ is a dimensionless number proportional to the energy loss. The Landau distribution possesses a long "tail", see Figure 149 which means the mean energy loss is much greater than the most probable value.



Figure 149: Landau distribution for 500MeV/c pions in thin silicon.

9.1.5 Diode Bulk

The diode considered above was n-bulk diode The bulk may of course be either n-type or p-type and may be either partially or fully depleted. Note that for the nbulk detector the depletion zone grows from the p+ implant towards the n+ implant with increasing (negative) voltage applied to the p+ connection. The pbulk depletion zone grows from the n-implant side with increasing (positive) voltage applied to the p+ connection.



Figure 150: Diagram showing both p and n-bulk diodes in reverse bias both partially and fully depleted.

9.2DC Coupled Strip Devices

The diode detector is, in itself, not a particularly useful high resolution device. The assembly of many individual, tiny diodes, to form a large area tracking device would be both a laborious and expensive task. Each diode would also possess dead areas around the edge of the silicon leading to excess material, the need to overall and excessively complicated mounting assemblies. In order to achieve high resolution we try to segment the readout of the diode detector.

9.2.1 Principle of Operation

In its simplest form the diode could simply have the p-implants placed as strips. with Al lines added on top of the p^+ to conduct any current to the amplifiers with pitch, *d*. Figure 151. Note the Al is directly in contact with implant, hence the nomenclature DC; any dark current in the sensor will be an input to the amplifier.



Figure 151: Diode split up with strips on p+ implant side

As long as the "cell" dimension is much larger than any diffusion length then any charge generated (or injected) into the bulk within a centre of a cell (e.g. cell 2 in diagram above) will only induce a large average current on strip 2. Note that although all the holes in cell 2 will end up on strip 2 that does not imply that the weighting field of strip 2 is zero outside of that cell. Charge moving directly up the real field lines (not the weighting field lines) will produce net zero current on adjacent strips. Figure 152.



Figure 152: weighting field in simple strip detector. Charge far away from a strip tends to produce small bi-polar signals.

9.2.2 Geometrical Considerations

The precise choice of the geometry of the segmented diode will depend on applications. There is no need for the strips to be strips at all, they may be pads or even concentric circles. The basic considerations are

- does the pattern match the needs of the experiment?
- is this pattern consistent with the properties of silicon e.g. compatible with diffusions distances>
- Will the segmentation yield a useful signal to noise and (closely linked) an adequate spatial resolution?

These issues will be dealt with in separate sections towards the end of this chapter.

9.2.3 Readout Schematic

A typical readout schematic for the DC coupled sensors is shown below in Figure 153. The p+ implants are kept at a negative voltage with respect to the n+ side of the sensor by applying a positive bias to the n+ side. The amplifier inputs, if wired "normally" will be close to ground potential allowing the sensor to be biased.



Figure 153: Schematic for the readout of a simple p^+n strip detector. The implant lines on the p+ side are kept at a negative voltage by the input to the electronics that will normally operate close to ground voltage

9.3 AC Coupled Strip Devices

9.3.1 Principle of Operation

The AC coupled detector is designed to over come a major obstacle with the use of DC devices, that of the dark (reverse) current in the diode. Not only may this overtax the amplifiers driving them into saturation, or still seriously reduce their dynamic range, but also the noise on the current will also tend to mask the small high frequency signals from particles that we wish to detect. An obvious way of overcoming this problem is to use capacitative coupling to the readout electronics

The circuit for achieving the AC coupling is shown below.



Figure 154: Schematic showing how AC coupling of the implant to the amplifier input may be achieved.

Critical to the functioning of the AC coupled detector is that the decoupling capacitance C_d must be large compared to the interstrip capacitance C_{is} , the interstrip capacitance, and C_{sb} the capacitance to the backplane. A charge q

signal arriving on the p+ implant will induce a charge q_c on the input of the amplifier

$$q_{C} = q(\frac{C_{d}}{C_{d} + C_{sb} + \sum_{i} C_{is}^{j}})$$
(9.19)

where we have summed over all the interstrip capacitance. In order to not to "lose" charge to the adjacent strips it is necessary to make C_d about 100 times larger than the other capacitances.

Note that it is also necessary to bias the implants. If the implant is not kept at the correct potential relative to the bulk then depletion will be lost. The capacitor alone does not permit us to hold a DC voltage as the reverse bias current would simply discharge the capacitor. This it is necessary to apply a separate bias resistor, R, to allow the strips to remain biased.

If the value of the bias resistor is too low, of order of a few ohms, (or we tried to simply provide a ground "short") to the strip then any current will be "lost" to ground rather than driving the input amplifier. Thus we need to provide a high resistance between the p+ implant and ground.



Figure 155: Resistors and decoupling capacitors. The values of the bias resistors must remain high to avoid shorting all the strips

This has the crucial effect of also isolating strips from each other. If the resistance to a common line is low then in effect all the strips on a detector would be shorted leading to a spreading of the charge across the entire detector and a loss of signal.

If a typical bias voltage is of order 100V then in order to maintain the equilibrium in the system, (and resolution as we will see later) it is desirable that this bias voltage does not change. In particular if the current induced from the moving charges is of order of a μ A then this will induce a voltage change of over a volt in a M Ω resistor. Thus we try and keep the effective bias resistance to be of order 1M Ω .

9.3.2 Fabrication

The fabrication of AC coupled detectors is more complicated than that of DC detectors. If a large area array with many strips is required then the only viable low cost option (for readout strips of pitch less than about 1mm pitch) is to integrate the resistor mechanism and AC coupling into the sensor.

The AC coupling This is achieved by separating the implants from the aluminium readout lines with a thin (~ 0.1μ m thick) oxide layer. This high quality oxide is usually thermally grown and capacitatively couples the implant to the readout line.



Figure 156: Schematic of AC coupled strip detector showing the thin thermally grown oxide layer below the AI lines.

9.3.3 Geometrical Considerations

As soon as we move away from the simplest strip diode structures the parameters it becomes difficult to define simple algorithms for the geometrical design of the sensors. Detailed simulation, as shown in our chapter on simulation is always used to check the field and potentials within a particular design in order to avoid high field which may lead to break downs, whilst optimising the weighting fields to give the biggest signal possible and the best resolution.

As with the DC couple sensors the implant and metallization on the strip side should be designed to match the application, although the majority of devices used have a simple repeating strip geometry.

Although it is advantageous to have wide strips, as this minimizes the resistance of the aluminium strips and increases the capacitance to the implant, wide strips will also tend to have a larger interstrip capacitance.

In addition it is necessary to plan the layout of the sensor to allow for the biasing mechanism to be integrated onto the wafer, whilst minimizing the amount of dead area.

9.3.4 Technical Limitations

Two of the most important technical problems for this kind of sensor have already been mentioned. The requirement to design and implement the biasing mechanism in a reliable fashion and the difficulty in implementing a high quality decoupling oxide.

Should the oxide develop pin-holes (i.e. a direct contact) between the implant and the input of the amplifier it is usual that the particular strip and its immediate neighbours are no loner useable or sensitive for physics. Either because the amplifiers channels have saturated or become very noisy, or in a more extreme case where many pinholes have developed on different input channels to the same ASIC, that the input amplifiers are driven into an unstable mode. Thus great care is taken in producing oxide that contains few defects.

Typically for modern strip detectors, depending on manufacturer and process, one observes about 1 pinhole for every 1-10cm² of MOS structure.

9.4 Strip Bias Techniques

One of the biggest problems in fabricating these sensors is the biasing. There are three common techniques in use today. The first is to deposit a doped polycrystalline silicon structure that acts as resistor to connect the implants to a nominal bias voltage. This technique is favoured in high radiation environments. The other two, reachtrhough and FOXFET, use a more subtle techniques to achieve the same effect but are more sensitive to ionizing radiation.

9.4.1 Polysilicon Resistors

The deposition of polysilicon has already been described. As we would expect the electrical characteristics of a poly-Si thin film depends on its doping. As in singlecrystal silicon, heavier doping results in lower resistivity. Poly-Si is more resistive than single-crystal silicon for any given level of doping mainly because the grain boundaries in poly-Si hamper carrier mobility.



By controlling the doping a thin sheet of polysilicon with known properties may be produced. By laying out the appropriate

Figure 157: Polysilicon resistors on a Si strip detectors

geometry, usually a long snake like structure, large resistances of order of $1M\Omega$ may be produced. Space must be reserved on the wafer for the resistors; as many resistors as there are strips must be imaged.

The resistivity of polysilicon is usually measured in Ohms/square which is a unity of Surface Resistivity. For electric current flowing across a surface, the ratio of DC voltage drop per unit length to the surface current per unit width. In effect, the surface resistivity is the resistance between two opposite sides of a square and its independent of the size of a square or its dimensional units. Surface resistivity is expressed in ohms/square. For the poly used in strip detectors resistivities of > 1000 Ω /square are used. Thus with poly lines of order 10 microns wide a 1M Ω resistor will require a length of order 1cm. Hence the long snake like forms of the resistors.

9.4.2 Reachthrough Bias

A different method of biasing is called punch-through or reach-through biasing. The precise details of the structures used depends on the bulk and the strip implant. For examples for p^+ in n bulk as the detector is biased the depletion region grows. When the depleted region reaches from the bias rail to the strip we have achieved our goal of effectively putting a high series resistance between the bias rail and the strips.



Figure 158:Reachthrough bias for p-strips in n-bulk. The diagram shows a cut along the length of the strips. Originally the strips sit in a conductive undepleted Si. Note how the depleted region grows and "reaches" across from the bias rail towards the strips as the negative voltage on the bias rail is increased. (After Lutz)

For n-strips in n-bulk the situation is slightly more complicated. On this side the bias resistance is controlled via an electron accumulation layer under the oxide when the sensor is fully depleted. Before this point the region is connected and the strips are shorted to the bias rail and each other.



Figure 159: Electron layer resistor in n+n sensor when fully depleted.(After Lutz)

9.4.3 FOXFET bias

A development of the reachthough technique is to add a p-implant to control the electron accumulation layer on the n+ side. To control the effect of the implant an MOSFET structure (as it is in the field oxide it is often called a FOXFET) is added. This enables us to control the resistivity of the channel. However the FOXFET bias system is very sensitive to radiation effects.



Figure 160: The bias of n+ strips in n- bulk controlled by a p+ implant (a). This may be modified to include a MOS structure (b) known as FOXFET biasing.

9.5 Spatial Resolution

We have ignored up to this point one of the most important features of silicon detectors. That is that they are able to provide high resolution information about the position that particles passed through the sensor.

If the information from a detector is purely binary, i.e. if a particle passes through a "cell" then that cell, and only that cell responds, the resolution may be simply related to the pitch d.

The familiar formula for the resolution of a "binary" system (either a cell hit or not) is given by considering the average deviation from a true position is given by

$$\left\langle \Delta x^{2} \right\rangle = \frac{\int_{-d/2}^{d/2} (x)^{2} dx}{\int_{-d/2}^{d/2} dx} = \frac{x^{3}}{3d} \Big|_{-d/2}^{d/2} = \frac{d^{2}}{12}$$
(9.20)

I.e. if we have a pitch d with binary resolution then we expect to obtain a resolution of approx

$$\sigma = \frac{d}{\sqrt{12}} \tag{9.21}$$

i.e. a 50 micron pitch detector should yield a resolution of at least $15\mu m$ or so.

9.5.1 Charge Sharing

The resolution of the detector may improved if the signal is induced on adjacent strips so the signal depends on the position of the particle between the two strips.



Figure 161: Schematic of particle passing between two strips at xp.A particle leaves a charge QL on the left strip and QR on the right strip.



Figure 162: η distribution for perfect resolution(solid) and binary resolution (dotted)

If the charges induced on the strips are Q_L and Q_R respectively and depend on the linear distance from the strip then if we define

$$\eta = \frac{Q_R}{Q_L + Q_R} \tag{9.22}$$

then η maps directly onto the position the particle went through the detector i.e.

$$\frac{x}{d} = \eta \tag{9.23}$$

The difference between the "perfect" η distribution and the binary resolution is shown in Figure 162. In reality we often achieve something in between, showing we expect the resolution to be worse than ideal but better than binary!



Figure 163: Eta distribution from a real detector. (no intermediate strips). LHCb.

A small amount of reflection reveals that if noise N is present in our measurements of the charge that this degrades the resolution. This may be quatified.

If we possess a system with noise N and total signal size S the resolution, σ , will be approximately related to the eta distribution by,

$$\sigma \sim \frac{N}{S} \times \left(\frac{d\eta}{dx}\right)^{-1} \tag{9.24}$$

Which gives an easy (but not necessarily accurate) way of estimating the resolution at a particular point for the sensor. Note that using this approximation we must integrate over the entire cell width to get the average resolution.

If we have ideal resolution (i.e. the eta distribution is the straight line drawn above) then

$$\sigma \sim d / (S / N) \tag{9.25}$$

For a narrow strip detector (e.g. 25 microns pitch) with ideal η distribution (for thick detector with long collection time) and S:N of better than 25:1 we would expect a resolution of approximately 1 micron. This can be achieved.

9.5.2 Charge Collection

We note here that very fast time collection gives rise to poor S:N and degrades the resolution. The noise arises both from the degraded performance of the readout electronics as we decrease the time available for sampling and, if we make the time short enough, that we may collect as little as 50% of the charge. This is discussed in a later chapter.

9.5.3 Diffusion

The mechanism that for sharing charge between different cells has different sources. One of the most important is "diffusion".

As the charge clouds move, holes towards the cathodes and electrons towards the anodes, they tend to grow. In a region where there is no gradient in the electric field that focuses the moving charges, the clouds will grow according to

$$\sigma = \sqrt{2Dt_d} \tag{9.26}$$

It should be remembered the diffusion depends on the temperature and the mobility of the carriers. The diffusion is due to the scattering of the carriers with the crystal. Depending on the experimental configuration, the levels of radiation damage, and where in the detector the cloud is generated, the transverse size of the charge cloud may be as big as 10 or 20 microns or as small as 1 micron. Long drift times clearly increase the transverse size of the charge cloud. Thus the diffusion leads to signals that may be picked up between two cells. Likewise at very high operating voltages (high mobility but very short collection time) we expect the sharing to be reduced.



Figure 164: Schematic of diffusion leading to charge being picked up on two strips

If the total width if the diffusion cloud is σ_d then the fraction of the a cell that will have binary resolution is

$$f_{bin} = \frac{d - \sigma_d}{d} \tag{9.27}$$

This fraction will thus have a resolution

$$\sigma = \frac{\left(d - \sigma_d\right)}{\sqrt{12}} \tag{9.28}$$

The remaining fraction

$$f_{linear} = \frac{\sigma_d}{d} \tag{9.29}$$

will have resolution approximately given by

$$\sigma = \frac{\sigma_d}{(S:N)} \tag{9.30}$$

The net resolution will thus be



$$\langle \sigma \rangle = \frac{(d - \sigma_d)}{d} \frac{(d - \sigma_d)}{\sqrt{12}} + \frac{\sigma_d}{d} \frac{\sigma_d}{(S:N)}$$
 (9.31)

Figure 165: Approximate improvement of resolution with size of diffusion cloud for 40micron pitch detectors for three S:N ratios. In the limit the diffusion becomes zero the detector achieves pure binary resolution.

Note that at the LHC we expect the diffusion cloud to have a transverse size of 1-2 microns hence the improvement to the resolution is almost negligible from this source.

9.5.4 Intermediate Strips

A technique to improve the charge sharing is to insert intermediate strips between the readout strips. The effect of the intermediate strip(s) is to locally collect charge in a quasi-binary fashion, but by using the fact that it is part of a capacitatively coupled network, the charge is divided between the nearest strips.



Figure 166: Schematic of sensor with intermediate strips

There is a small degradation of the signal in when using the intermediate strips as a larger proportion of the charge is shared with the back plane. This technique is limited by the ability to fabricate intermediate strips that are isolated from each other.

9.5.5 Angled Tracks

Another feature that improves charge sharing is the angling of tracks. Non perpendicular *tracks* may easily traverse more than one cell, hence improving the charge sharing and giving more information about the average position of the track within the Si.

For example in Figure 167 a particle is incident at angle θ to the normal at a distance x_p from the centre of a strip. The pitch of the sensors is *d*. If the track would normally deposit charge Q at normal incidence it will deposit

$$Q_{tot} = Q / \cos \theta \tag{9.32}$$

Assuming that the depth of the silicon is w, and then assuming that the only charge sharing mechanism in action is the angle of the track then the fraction of the cell that will only contain binary information is f_{bin}

$$f_{bin} = \frac{d - w \tan \theta}{d} \tag{9.33}$$

This fraction will thus have a resolution

$$\sigma = \frac{(d - w \tan \theta)}{\sqrt{12}} \tag{9.34}$$

The remaining fraction in which the track traverses a cell boundary and hence splits its charge is f_{linear} .



 $f_{linear} = \frac{w \tan \theta}{d} \tag{9.35}$

Figure 167: Angled, non-perpendicular track, within Si sensor

In this linear region the eta distribution changes between 0 and 1 over a distance $w \tan \theta$. This we expect the resolution in this area to be given by

$$\sigma_{linear} = \frac{w \tan \theta \cos \theta}{(S:N)_{perp}} = \frac{w \sin \theta}{(S:N)_{perp}}$$
(9.36)

Where the S:N is larger compared to the perpendicular value due to the value of Q_{tot} .

Thus the average value of the resolution for angled tracks is

$$\langle \sigma \rangle = \left(1 - \frac{w}{d} \tan \theta\right) \frac{\left(d - w \tan \theta\right)}{\sqrt{12}} + \frac{w \sin \theta}{(S:N)} \cdot \frac{w \tan \theta}{d}$$
 (9.37)

under the condition that

$$w\tan\theta < d \tag{9.38}$$



Figure 168:Schematic of resolution showing enhanced performance with angle as a function of s:n.

9.6 Edge Effects

An important feature of the construction of Si sensors is the control of the fields at the edge of the active Si. We may wish to apply bias voltages of up to 1000V between the two sides of the Si. This is only possible if the two sides are effectively isolated.

Broken and cut silicon will in general have defects (surface and bulk) at the surface and close to the cut which will conduct current. Enough defects make it impossible to bias the detector. Thus we need a technique to smoothly reduce the voltage from the high voltage regions across the edge of the sensor.

Such structures are usually called guard rings and may be seen around most particle physics strip sensors. They usually made of a series of metallic rings and implants surrounding the active regions. They may be biased or floating.

The guard rings are often left floating rather than biased to a specific voltage.



Figure 169: Guard rings surrounding an active region of a sensor.





Figure 170: Guard ring structure and simulated field potential in LHCb sensors. The aim of the guard rings is to keep the field below the breakdown potential. This structure has resisted up to 1000V.

A problem with some guard ring structures the large area occupied on the Si.The LHCb structure shown above occupies over 5mm on the wafer. Ideally one would like to reduce the space occupied by the inactive guard rings whilst controlling the field at the edge of the sensor.

9.7 Bulk and Implants

The choice of either n-type or p-type bulk is subtle Both p-type and n-type sensors have been shown to work and recent studies have achieved extremely high radiation tolerance for p-bulk sensors. We will discuss this topic in detail in a later chapter on radiation performance of Si sensors.

9.7.1 Strip Implants

The biggest choice is whether to read out from the p-side or the n-side. In n-bulk detectors (prior to irradiation) if it is wished to operate the sensor underdepleted

we must use p^+ implants as this is the side the sensor depletes from. We note the high field region closest to the strips will be attracting the slower moving holes.

If the high field region is to collect the higher velocity electrons then the strips should be placed on the n^+ side, i.e have n^+ implants. In this case the detector must be operated fully or over depleted.

9.7.2 Strip Isolation

The p-strips in n-bulk are negatively biased with respect to the bulk to produce the reverse bias. We assume that we wish to build AC coupled p-strips. From our discussion of MOS capacitors we expect there to be a local hole accumulation under the strips. This hole accumulation will eventually absorb any mobile charges in the SiO_2 leaving the trapped, immobile holes as the only carriers in the intervening region. Thus the strips are well isolated.

For the n-strip side the case is more complicated. There is electron accumulation in the n^+ implant and mobile electron carriers inside the intervening oxide. If nothing else were done the n-strips would not be isolated from each other and thus inoperable.

There are several techniques to solve this on the n-strip side. The first is to implant (with an ion-implanter) insulating p-strips (called p-stops) in between the n-strips. These collect the mobile electrons.



Figure 171: P-stops on n-strip readout.

Another option is to apply an overall p-implantation (before other processing) to what will become the n+ side. Thus we will be left with a lightly p-doped layer in between the n+ strips under the oxide, which again will act to absorb the electrons.

Another option, call graded-p-spray, is to apply both a p-spray and an diffused p-structure (p-stop like) into the n+ side.

In each case the p-implants are designed to both isolate the strips and minimize the high fields within the n-strip side to avoid micro-discharges.

Isolation may also be achieved, but at the cost of complicating the detector structure, by adding a MOS capacitor structure (field plate) in between each n-strip and biasing it so as to repel the electrons. This would close the channel between the n strips.

9.8 Double Sided Detectors

The astute observer will have noted there is not reason, since we can build a double sided detector with strips on both sides.



Figure 172: Schematic of double side detector.

If the strips on opposite sides of the sensor are placed at right angles to each other a particle a space point (rather than a single coordinate) will be produced for the hit. In principle it is possible to correlate the charge detected on either side of the sensor to reduce the number of "ghost" reconstructions. See next chapter.

Double sided detectors are very useful but complex and expensive to make and reaedout. It is often easier to use two single sided sensors in close proximity to achieve the same effect.

A bigger problem is that unless the electronics is to be operated "floating" at least one side of the sensor must hold-off a large voltage across the AC coupling oxide. With the high voltages used for operation (>100V) this will place a severe strain on the oxide, and make the development of pin-holes during operation more serious. The alternative, "floating" the electronics so the ground of one set of amplifiers is at the bias voltage, makes for complicated construction and for very great care in the design of the front-end electronic supports.

9.9 Readout

Readout of the signals takes place via the signal being collected on the implant, (which is high resistive and hence localized). The pulses are AC coupled to the aluminium strips which connect via bondpads to the front-end electronics.

9.9.1 Bondpads



Figure 173: Typical bond from an AI wire to a bond-pad on the sensor that is an area large enough to allow the "welding" of AI wire to the AI strip. Typically these are at the edge of the detector.

The bondpad structures are large enough to allow bonding (using a bonding machine and are typically at the end of strips close to the edge of the detector. They may or may not be over the active area of the strip.



Figure 174: Schematic of bondwire being welded by ultrasonics to the Al bond pad.



Figure 175: Photograph of bonded strips.

9.9.2 Readout Layers



To effect the readout of the signals it is sometimes desirable to use, as one does in PC boards, another readout layer which enables the readout to happen at an angle to the strip layout.

It is usual in strip detectors to have a maximum of two metal layers for readout. Detectors of this sort are called

double metal detectors. The Figure 176: THe LHCb sensors that use a conducting layers are separated by a double metal layer to route our signals thick (CVD) 4-5micron oxide. The thickness is reduce the readout/strip capacitance which will degrade the signal.

Connections between the two layers is done by etched holes in the oxide which are filled, at the time the second metal is deposited with aluminium. These holes are called vias.

9.9.3 Electronic noise

There are several source for electronic noise in strip detectors. We summarize here, for completeness, the most important:

- Capacitive load (C_d). Often the major source, the dependence is a function of amplifier design. Feedback mechanism of most amplifiers makes the amplifier internal noise dependent on input capacitive load. ENC \propto Cd
- Sensor leakage current (shot noise). ENC $\propto \sqrt{I}$
- Parallel resistance of bias resistor (thermal noise). ENC $\propto \sqrt{(kT/R)}$
- Total noise generally expressed in the form (absorbing the last two sources into the constant term a): $ENC = a + b \cdot Cd$

The noise has been typified in term of ENC or electron noise equivalent. This provides us with an easy way of estimating the signal to noise. If a detector provides 20,000 electrons signal and the noise is the equivalent 1000 electroncs then the S:N will be 20:1.

The front end amplifiers used for the LHC (For example) have noise performances that are of the order of 500e- and 50e- per pf. Thus for a signal of 20,000 electrons and a typical load capacitance of 10 pf we expect to obtain S:N of about 20:1.

9.10 Sensor Testing

Both the sensor and electronics for particle physics experiments are expensive. Complex sensors may cost about 100Euro/cm^2 and their readout electronics about 1Euro/channel. With total silicon areas varying from about 0.1 to 100m2 and the number channel counts ranging from 0.1m to 10m the testing of sensors before they are assembled is a priority.

The acceptance of sensors usually depends on the sensor passing a number of criteria to ensure that the diode and strips/pixels are correctly fabricated.

• The reverse bias current for the entire silicon sensor is measured to ensure that the bulk leakage currents is small and typically of the order 1-10 μ A. At high voltages a larger than expected leakage current may well be a precursor to avalanche breakdown or a surface discharge.

- The capacitance as a function of bias voltage is measured to make sure the bulk is within specified tolerances. Typically depletion voltages of about 50V are expected for 300micron thick sensors.
- If the detector is AC coupled then each strip is measured to ensure there are no "pin-holes" in the insulating layer.
- In many cases the strips are checked to ensure that they are isolated from each other (especially important for n+n sensors).

References

Ramo, Simon, 1939, Currents Induced by Electron Motion, Proceeding of the IRE, vol. 27, pp. 584-5.

Semiconductor Radiation Detectors, G.Lutz, Springer

A. Honma, Silicon Detectors, NATO Advanced Study Institute, 2002.

http://www.semiconfareast.com/polysilicon.htm

J.Kemmer and G.Lutz, NIM A236 (1993), 209-213

P.Allport et al, Nuclear Instruments and Methods in Physics Research Section A, Volume 310, Issue 1-2, p. 155-159

L. Landau. On the energy loss of fast particles by ionization, J. Phys (Moscow), VIII:201,1944.

Problems



PIXEL DETECTORS

The adaptation of the strip technology to build pixel detectors is described. The sensors possess no inbuilt electronic logic nor amplifiers. The bump-bond method of connecting the pixels to readout electronics is described.

10.1 Pads, Strips and Pixels

S ilicon strip detectors, as described in the last chapter, come in a large range of dimensions and are used for a number of diverse applications. Very high (micron level) resolution strip detectors may be used for vertexing while wider pitch strip detectors are used for tracking. Whether single or double sided, strip sensors suffer when used in an environment when multiple hits/strip are likely. The fraction of time a strip is hit is known as the occupancy.

If the occupancy of a strip detector becomes larger than about 1% thus usually complicates the analysis and use of the sensor. By the time the occupancy reaches 10% the performance of the detector as a whole starts to degrade noticeably. The degradation is usually due, not to problems reconstructing charge clusters, but to pattern recognition.

In a double sided (or two single sided) detectors with n real clusters will in general give rise to a number of ghost clusters, see Figure 177. If a track only hits a single strip then the number of ghosts given by is

$$n_{ghost} = n^2 - n \tag{10.1}$$

If the occupancy is f, and the number of strips m, the ratio of ghost clusters to real clusters R, is

$$R = \frac{n_{ghost}}{n} = \frac{fm(fm-1)}{fm} = fm-1$$
 (10.2)

In a sensor (or pair of sensors) with 1000 strips in each direction and an occupancy of 1%, R~10 i.e. there are 10 times more ghosts than real clusters.



Figure 177: Schematic of double sided strip detector with 3 real hits

The only way to reduce the ratio R is to use a large number of small sensors (make m small). In the limit that m=1 i.e. each sensor is a single small pad of the dimensions necessary to produce the resolution needed the problem has reduced to that of the single diode detector.

As discussed before a myriad of small independent detectors is not practical. Therefore we try and integrate all these separate sensors onto a single wafer, to make a pixel sensor.

10.2 Pixel Detectors

The pixel detector described above is simply multiple diode sensor whose fabrication has been integrated onto a single wafer. The pixels themselves may be almost any shape. The ATLAS pixels, for example are sensitive diodes $50\times400\mu$ m. Pixels of these dimensions give rise to enormous readout densities e.g. ~ 50×10^6 m². In most cases electronics much be attached separately to the diodes. Thus the final detector is often called a hybrid pixel.

10.2.1 Principle of Operation

The pixels may be constructed in identical fashion to strip detectors. For example they may be built on n-bulk with the pixel structures being n^{-} implants isolated with a p-spray.

The advantage of the p-spray isolation technique is that it does not require a complex inter-pixel isolation structure and can be fabricated without complex imaging.

As with the strips all pixels must be biased. In Figure 179, which shows an example from the ATLAS experiment, the pixels are biased using the punch-through technique.



Figure 178: Schematic of a CMS pixel structure using puncghrough biaising and p-spray isolation



Figure 179: An ATLAS pixel bias scheme. The pixels use a punch-through bias and isolated using floating p-stops..

10.3 Bonding

Bonding using a wire bonder is practical only if the pixel sensor contains a small number of pixels (<100). For any higher number a new technique must be used.

The channel count of a pixel sensor depends on the area. A pixel sensor has a channel count that usually depends only on one linear dimension, and the pitch usually matches that of the readout electronics, i.e. a pitch of approximately 40 microns. Thus it is straight forwards to connect the electronics to the strips by lining the edges of the electronics up with the bond-pads on the strips and using bonding machine. For the pixel detector a few hundred pixels may be aligned along the same line of real estate that would read out a single pixel. As each pixel has no addressable electronics a direct connection to the front-end must be made.

The only practical way to do this is to put the electronics on top of the pixels and finding a method of interconnecting the two.

10.3.1 Bump Bonding

In bump bonding the high density interconnects are achieved by the creation of small bumps (down to 15micron radius) on a surface that may be bonded, through sutiable application of temperature and pressure, to matching pads.

Bonds must first be grown on one surface see Figure 180 and Figure 181 before the second layer is attached e.g. Figure 182



Figure 180: Indium bumps grown on the CMS readout electronics before bump-bonding to the pixels



Figure 181: Lead-tin bumps at 50micron pitch after reflowing (FNAL)


100µm 250X

Figure 182: Cross-section of bump-bonded materials

10.4 Readout

The readout chips for the pixel sensors are inevitably an expensive and almost always customized item. See for example Figure 183. Given the scale of the chips and the bumps the tolerances for placing the chips on the detectors has to be of the order of 1-2 microns.



Figure 183: Schematic of CMS pixels bonded to electronics

The design of the chip must contain not only the structure to make the bump bonds but also the connection to the outside world. See Figure 184.

With pixel detectors the overall construction of the sensors is extraordinarily complex and there is substantial material in the beam. The high resolution and tolerance to high occupancy often more than compensate for this which is why many experiments have chosen to deploy pixels as their inner vertexing layers.

It should be noted that for the design of pixel detector shown here there is onereadout channel for each active element.



Figure 184: CMS pixel chip. The cells are not visible on this scale. The front end chip contains the structures to allow bump bonding to the pixels but also readout to the detector acquisition.

References

Problems



CHARGE COUPLED DEVICES

A simple description of CCD devices is given. The main features relating to charge transfer efficiency are discussed. Problems relating to the use of CCD devices in High Energy Physics applications are listed..

11.1 CCD

This is exists in many easily recognizable forms, for example in cameras (all digital cameras on the market), digital video recorders, and webcams.

11.1.1 Principle of Operation

The simplest form of a CCD can be thought of as a series of MOS capacitors. A simple one dimensional view is shown in Figure 185. Different voltages may be applied to each one of the capacitors in the CCD. If a higher voltage is applied to a capacitor than to its neighbours, charge can be localized. In essence an electric field is shaped that can retain charge in a "bucket".

The shape of the electric field can be changed in the device by the manipulation of the voltages on the electrodes. For example if a voltage is applied to a capacitor adjacent to the "bucket" that exceeds that voltage on the bucked and charged trapped will tend to be transferred from the existing bucket to the new "minimum", see Figure 186. This shows how charge can be moved under the capacitors.

A judicious application of voltages in series would then have the effect of systematically moving charge on the device. Figure 187 shows a schematic of series of capacitors connected to three "control" lines which are labelled $\varphi_1, \varphi_2, \varphi_3$. If in the initial state $\varphi_1=5V, \varphi_2=10V, \varphi_3=5V$ then charge will be trapped under the capacitors connected to φ_2 . By applying $\varphi_1=5V, \varphi_2=10V, \varphi_3=15V$ charge is then moved to the right under the capacitors connected to φ_3 . The voltage φ_3 can then be reduced so that $\varphi_1=5V, \varphi_2=5V, \varphi_3=10V$. The charges are now in the original configuration except displaced to the right by one unit cell.



Figure 185: Schematic of CCD. This one dimensional view shows three MOS capacitors with different voltages applied. The central capacitor has a high voltage applied to its neighbours. This forms a kind of "bucket" than can localize charge..



Figure 186:Schematic of CCD showing charge transfer towards to right.



Figure 187: A1D schematic of a series of MOS capcitors in a CCD. Each of the capcitors is connected in turn to one of three lines that control their voltages. These are labelled $\varphi_1, \varphi_2, \varphi_3$.

Note that it is important for there to be three or more (rather than two) control lines as this allows charge to be uniquely transferred in one direction.



185



- 11.1.2 Geometrical Considerations
- 11.1.3 Fabrication
- 11.1.4 Testing Techniques
- 11.1.5 Efficiency and Resolution
- 11.1.6 Technical Limitations
- 11.2 High Speed Readout
- 11.2.1 Clocking



11.2.2 Column Parallel Readout



References

Problems



APPLICATIONS

Chapter subtitle

12.1 Tracking and Vertexing sd

12.2 SLAC

Sjdaksjdhaksd









	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Total
min radius	3.2 cm	4.0 cm	5.4 cm	12.4 cm	14.0 cm	
modules	6	6	6	16	18	52
Si wafers	6x4=24	6x4=24	6x6=36	16x7=112	18x8=144	340
readout pitch (φ)	50 µm	55 µm	55 µm	100 µm	100 µm	
readout pitch (Z)	100 µm	100 µm	100 µm	210 µm	210 µm	
readout chips	144	192	240	256	288	1120
channels	18432	24576	30720	32768	36864	~140k

12.2.2 SLD – CCDs







12.3 Fermilab

Jsdhkajsdhaks



12.3.1 CDF – Strip Detectors





2 double-sided 1.2° layers, r=20,28 cm •Central layer (|e| < 1): •link tracks from COT to SVXII •Forward layers (1 < |e| < 2): •Standalone silicon tracking

Large-scale module production •Less stringent alignment •Hybrids at ladder ends



Goal: provide closest possible point measurement before particles scatter





Beampipe-mounted layer (1.5 cm radius)
25 μm pitch with alternate (50 μm) readout
Single-sided radiation-hard sensors (CMS)
Kapton cables to readout electronics at ends of active region at larger radii
Incredibly tight assembly clearance: 300 um between L00 and SVXII!
Active region covers SVXII bulkheads



HEP EXPERIMENTS



12.3.2 D0 – Strip Detectors







- 144 wedges
- 2.6 cm < r < 10 cm
- 50 μm (p-side)
 62.5 μm (n-side) pitch
- Double sided wedges with ±15⁰ (30⁰ effective stereo)
- Variable strip length



BTeV - Pixels

- 12.4 CERN
- 12.4.1 ALICE Pixels



HEP EXPERIMENTS





• Baseline design:

∠n+ pi×el in n-bulk material: ∠Moderated p-spray isolation.

- ∠Bias grid to allow testing before module assembly.
- ∠Oxygenated silicon to improve radiation resistance and increase allowable time to room temperature (for repair/upgrades).









Middle forward module



- silicon microstrip p⁺ on n detectors
- = 768 AC-coupled readout strips
- pitch size
 - 80 µm (barrel)
 - 50 90 µm (endcap)

<u>5 different module types</u>

- 1 barrel type
- 4 endcap (forward) types
 - outer
 - middle
 - short middle
 - inner

Difference between types

- geometry (mainly)
- different radial position on disk



- 12.4.2 ATLAS Pixels
- 12.4.3 ATLAS Strips















HEP EXPERIMENTS









References

Problem

Part 3 Advanced Techniques



RADIATION TOLERANCE

Radiation Tolerance

- **13.1 Effects of Radiation**
- 13.1.1 Low energy photons
- 13.1.2 High energy photons
- 13.1.3 Electrons
- 13.1.4 Ionizing Radiation
- 13.1.5 Neutrons/NEIL









215



References

Problems


3D PROCESSING

- 14.1 Edge Processing
- 14.1.1 Principle of Operation



Fabrication Steps

- etch border trenches
- diffuse in dopant
- grow protective oxide cover
- fill trench with poly
- vertical, directed etch (to dotted lines)
- n and p electrodes can be reversed



- 14.1.2 Geometrical Considerations
- 14.1.3 Fabrication
- 14.1.4 Technical Limitations
- 14.2 3D detectors

14.2.1 Principle of Operation



Ineffcient Area of Electrodes

DEVICES IN PARTICLE PHYSICS

- 14.2.2 Geometrical Considerations
- 14.2.3 Fabrication
- 14.2.4 Testing Techniques
- 14.2.5 Efficiency and Resolution
- 14.2.6 Technical Limitations



14.3 Laser Cutting

References

Problems



DRIFT DETECTORS

Chapter subtitle

- **15.1 Solid State Drift**
- 15.1.1 Principle of Operation





Detail of junctions and oxide layers



- 15.1.2 Geometrical Considerations
- 15.1.3 Fabrication
- 15.1.4 Testing Techniques
- 15.1.5 Efficiency and Resolution
- 15.1.6 Technical Limitations
- **15.2** Applications

References

Problems



CMOS

Chapter subtitle

16.1 Technology

16.2 Active Pixel Sensors







16.2.1 Principle of Operation



16.2.3 Fabrication



16.2.4 Testing Techniques



- 16.2.5 Efficiency and Resolution
- 16.2.6 Technical Limitations

References

Problems





B

bremsstrahlung, 125

C

cavities, 125 cyclotron, 127

D

Dee, 127 density, 10

н

hadronic calorimeter, 139

intrinsic semiconductors, 10

S

synchrocyclotron, 129 synchro-cyclotron, 129 synchrotron, 129 Synchrotron Radiation, 130