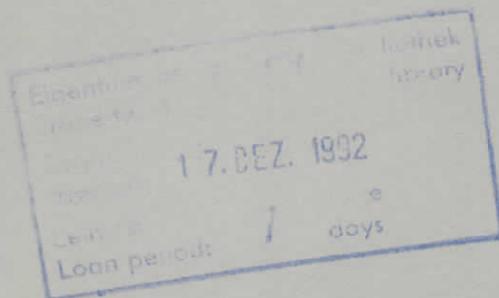


Internal Report  
DESY F35D-92-05  
December 1992

# Characteristics and Tolerance to Radiation of the Silicon Detectors Used in the Hadron-Electron Separator of ZEUS

by

M. Momayezi



**DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.**

**DESY reserves all rights for commercial use of information included in this report, especially in case of filing application for or grant of patents.**

**“Die Verantwortung für den Inhalt dieses  
Internen Berichtes liegt ausschließlich beim Verfasser”**

Characteristics and Tolerance to Radiation  
of the Silicon Detectors used in the  
Hadron-Electron Separator of ZEUS

Dissertation  
zur Erlangung des Doktorgrades  
im Fachbereich Physik  
an der Universität Hamburg

vorgelegt von  
Michael Momayezi ✓  
Hamburg  
1992

Abstract

At DESY in Hamburg an electron-proton collider (HERA) is presently taking up operation. ZEUS, one of HERA's two experiments, will employ a hadron-electron separator (HES) whose task it will be to find electrons, mainly originating from heavy-quark decays, inside jets of hadrons. The HES consists of a layer of  $3 \times 3 \text{ cm}^2$  silicon detectors built into the ZEUS calorimeter. They sample the particle cascades that grow when impacting high-energy particles dissipate their energy in the calorimeter. Being located at a depth of four radiation lengths the HES resides at a position where electromagnetic showers have already reached their peak intensity while hadronic ones are still in a premature phase. Reading the energy deposition in the HES planes allows to distinguish between the two shower types.

For particle momenta between  $2 \text{ GeV}/c$  and  $10 \text{ GeV}/c$ , a discriminating cut in the energy deposition in the HES detectors that correctly identifies 90% of the genuine electrons, will mistake 5% of the hadrons for an electron.

A much bigger range of particle momenta is successfully covered if the HES consists of two layers of silicon detectors, at depths of four and seven radiation lengths. In that case, hadron misidentification probabilities are about 1% for particle momenta between  $2 \text{ GeV}/c$  and  $75 \text{ GeV}/c$  while the odds for correctly identifying the electrons are 90%.

The final HES will comprise 50,000 silicon detectors. These must conform to very demanding standards concerning operating voltage, dark currents, electronic noise, and tolerance to penetrating radiation. To have a quality control and to provide guidance for the manufacturer during the development phase, a semiautomatic but flexible test station has been built. The physics and technology of the silicon detectors and the efforts on quality control are described.

The radiation damage caused by electromagnetic showers and neutrons is treated quantitatively by measuring macroscopically observable effects like dark-current increase and the formation of negative space charge in the depletion zone of the diodes.

These effects are attributed to the action of crystal defects. To study their characteristics a new method is introduced. It is based on recording in real time the electric current caused by charge carrier emission from the defects and has been named current-transient recording, or CTR for short.

It can supplement the most often used DLTS in various ways. First, it retains all information on how the emission proceeds with time, so no assumptions are needed in the subsequent analysis. Next, it knows no high-concentration limit, which makes it well suited for studies involving weakly doped material. And finally, CTR allows to measure the defect occupancy as a function of the Fermi level position when this crosses the defect level. From these data one can extract the ionization entropy to high precision. Moreover, if the defect concentration exceeds the doping, the defect type, acceptor or donor, can be derived unambiguously.

The first acceptor level of the divacancy could be studied in great detail. The reaction enthalpy and entropy of the process  $VV^- \rightarrow VV^0 + e^-$  have been measured very precisely:  $\Delta H = (0.4216 \pm 0.0007) \text{ eV}$  and  $\Delta S = -(1.05 \pm 0.05) \text{ k}$ , respectively. Further, it was discovered that the electron emission proceeds non-exponentially with time.

The macroscopic radiation damage is shown not to be due to the presence of divacancies, but to that of another kind of defect with an acceptor level just below the bandgap centre. The dopant concentration was found to be the same before and after the irradiation.

## Zusammenfassung

Am DESY in Hamburg wird derzeit der Elektronen- und Protonenbeschleuniger HERA in Betrieb genommen. ZEUS, eines der beiden Experimente an HERA, wird mit einem Hadronen-Elektronen-Separator (HES) ausgestattet. Dessen Aufgabe wird es sein, Elektronen zu identifizieren, die sich innerhalb von hadronischen Teilchenjets befinden und hauptsächlich vom Zerfall schwerer Quarks herrühren. Der HES besteht aus Lagen von  $3 \times 3 \text{ cm}^2$  großen Siliziumdetektoren, die in das ZEUS-Kalorimeter integriert werden. Mit ihnen mißt man die Teilchenkaskaden an, die entstehen wenn hochenergetische Teilchen ihre Energie im Kalorimeter verlieren. Der HES befindet sich in einer Tiefe von vier Strahlungslängen und somit an einer Stelle, an der elektromagnetische Schauer ihre volle Intensität erreicht haben, während die hadronischen Schauer noch unterentwickelt sind. Bestimmt man die Energiedeposition im HES, so kann man die beiden Schauerarten voneinander unterscheiden.

Im Impulsbereich zwischen  $2 \text{ GeV}/c$  und  $10 \text{ GeV}/c$  werden durch einen Auswahlchnitt in der deponierten Energie, der 90% der wahren Elektronen als solche erkennen lässt, etwa 5% der Hadronen als Elektronen identifiziert.

Ein wesentlich größerer Impulsbereich kann erfolgreich abgedeckt werden, wenn der HES aus zwei Lagen von Siliziumzählern, in Tiefen von vier und sieben Strahlungslängen, besteht. In diesem Fall beträgt die Wahrscheinlichkeit ein Hadron falsch zu identifizieren ca. 1% über einen Impulsbereich von  $2 \text{ GeV}/c$  bis  $75 \text{ GeV}/c$ , wenn die Chancen ein tatsächliches Elektron korrekt zu erkennen auf 90% festgesetzt werden.

In der letzten Ausbaustufe soll der HES etwa 50.000 Siliziumdetektoren umfassen. Diese müssen äußerst strengen Anforderungen in Hinblick auf Betriebsspannung, Dunkelstrom, elektronisches Rauschen und Widerstandsfähigkeit gegen durchdringende Strahlung genügen. Zur Qualitätskontrolle und um den Herstellern während der Entwicklungsarbeiten Hilfestellung geben zu können, wurde ein halbautomatischer und dennoch vielseitiger Teststand aufgebaut. Physik und Technologie der Siliziumdetektoren und der Umfang der Qualitätskontrollen werden beschrieben.

Strahlenschäden, durch Neutronen und elektromagnetische Schauer verursacht, werden quantitativ durch Messung der makroskopisch beobachtbaren Auswirkungen, wie Dunkelstromerhöhung und Bildung negativer Raumladung in der Verarmungszone der Dioden, behandelt.

Diese Effekte werden als Auswirkung von Kristallfehlern angesehen. Eine neue Methode zur Bestimmung ihrer Eigenschaften wird vorgestellt. Sie stützt sich auf das Aufzeichnen der Stromtransiente, die durch Ladungsträgeremission von den Defekten entsteht und wurde in der englischen Entsprechung current-transient recording genannt, abgekürzt CTR.

Diese neue Methode kann den durch die am häufigsten verwendete Methode, DLTS, gesetzten Rahmen um einige Punkte erweitern. Zum einen bleibt sämtliche Information über den zeitlichen Ablauf der Ladungsträgeremission erhalten, so daß diesbezüglich keinerlei Annahmen in der nachfolgenden Analyse gemacht werden müssen. Weiters gibt es keine Obergrenze in der zulässigen Defektkonzentration, so daß CTR für Untersuchungen an hochohmigem Material bestens geeignet ist. Und schließlich erlaubt CTR die Messung der Besetzungswahrscheinlichkeit der Defekte als Funktion des Fermi-niveaus während dieses das Defektniveau passiert. Aus diesen Daten kann die Ionisierungsentropie mit großer Genauigkeit gewonnen werden. Wenn zudem die Defektkonzentration höher als die Dotierungsdichte ist, kann der Typus des Defekts, Donor oder Akzeptor, unzweideutig ermittelt werden.

Das erste Akzeptorniveau der Divakanz konnte sehr genau untersucht werden. Die Enthalpie und die Entropie der Reaktion  $VV^- \rightarrow VV^0 + e^-$  konnten sehr präzise bestimmt werden zu

$\Delta H = (0.4216 \pm 0.0007) \text{ eV}$  und  $\Delta S = (1.05 \pm 0.05) \text{ k}$ . Es zeigte sich ausserdem, daß der Emissionsprozess keinem exponentiellen Zeitgesetz folgt.

Es wird nachgewiesen, daß der makroskopisch sichtbare Strahlenschaden nicht auf die Anwesenheit von Divakanzen zurückgeführt werden kann, sondern statt dessen einem anderen Defekt, der ein Akzeptorniveau knapp unterhalb der Bandmitte besitzen muß, zugeordnet werden muß. Die Dotierungskonzentration wurde durch die Bestrahlung nicht verändert.

# Contents

<b>1</b>	<b>Physics at HERA</b>	<b>1</b>
1.1	Introduction	1
1.2	Deep-inelastic physics	1
1.3	Event classification and kinematic range	2
1.4	QCD at HERA	3
1.5	Heavy-quark physics	4
1.6	ZEUS—A detector for HERA	5
<b>2</b>	<b>A hadron-electron separator for ZEUS</b>	<b>9</b>
2.1	Introduction	9
2.2	Hadron-electron separation using a calorimeter	10
2.3	Theoretical aspects	11
2.3.1	Shower development	11
2.3.2	Limitations	15
2.4	Proof of principle	16
2.4.1	The setup	16
2.4.2	Performing the experiment	17
2.4.3	One-plane analysis and results	18
2.4.4	Two-plane analysis and results	19
2.4.5	Conclusion	21
2.5	Prototyping	22
2.5.1	The HES prototype	22
2.5.2	Test beam setup	25
2.5.3	Analysis and results	26
2.5.4	Scans across diodes	27
2.6	Summary	31
<b>3</b>	<b>Technical description of the HES detectors</b>	<b>32</b>
3.1	Fabrication	32
3.2	Fields and potentials	35
3.2.1	Equilibrium case	35
3.2.2	Backward bias	37
3.2.3	Detector capacitance	38
3.2.4	MOS-capacitance	39
3.2.5	Reverse currents	41
3.3	Detector performance	42
3.3.1	Energy deposition	42
3.3.2	Charge collection	43
3.3.3	Incomplete charge collection	43
3.4	Noise	45
3.4.1	Introduction	45
3.4.2	Shot noise in pn diodes	45
3.4.3	Signal processing	46
<b>4</b>	<b>Detector testing</b>	<b>48</b>
4.1	Introduction	48
4.2	Items under survey	48
4.3	The test facility	49
4.3.1	General layout	49
4.3.2	Software	50
4.3.3	The features	51
4.4	Test results	55
4.4.1	Ansaldo	55
4.4.2	Hamamatsu	55
4.4.3	Siemens	56
4.5	Conclusion	56
<b>5</b>	<b>Phenomenology of radiation damage</b>	<b>60</b>
5.1	Sources of radiation	60
5.2	Radiation effects in the detectors	61
5.3	Experiments	62
5.4	Results	63
5.4.1	Uranium activity	63
5.4.2	Results on dark current	63
5.4.3	Results on doping changes	67
5.4.4	Results on oxide charging	68
5.5	Operation at ZEUS	71
<b>6</b>	<b>Point defects and their characterization</b>	<b>72</b>
6.1	Theory of point defects	73
6.1.1	Crystallography	73
6.1.2	Coulombic defects	74
6.1.3	Thermodynamics involving deep defects	75
6.1.4	Non-equilibrium but stationary processes	83
6.1.5	Non-stationary processes	86
6.2	Current-transient recording	88
6.2.1	Sample preparation	88
6.2.2	The CTR equipment	89
6.2.3	Raw data analysis	93
6.3	Experimental findings	96
6.3.1	Raw data analysis	96
6.3.2	Chemical identity of the deep defect	108
6.3.3	The midband defect	112
6.3.4	Comparison with DLTS	116
<b>7</b>	<b>Summary</b>	<b>122</b>
7.1	Resumé	122
7.2	Quantitative results	124
<b>A</b>	<b>Symbols and their Meanings</b>	<b>126</b>

# 1 Physics at HERA

## 1.1 Introduction

The first notion on a possible granular substructure of the nucleon came from electron-proton scattering experiments performed at the fixed-target facility at SLAC in 1968, [Pa68]. There it was found that an unexpectedly large fraction of electrons were scattered under wide angles from a hydrogen target. This was attributed to the existence of smaller particles inside the proton. Until today, deep-inelastic-scattering experiments with electrons, muons, and neutrinos have delivered a detailed picture of how the constituents of the proton (the 'partons') share spin and momentum. The description is condensed into structure functions  $F(x, Q^2)$ . It extends to momentum transfers of  $Q^2 = 200 \text{ GeV}^2$ , and the range covered in  $x$ , the parton's fraction of the proton momentum, is 0.02 to 0.8.

At DESY in Hamburg (Germany), an electron-proton collider, HERA, is presently taking up operation. This unique machine produces head-on collisions of 30 GeV electrons with 820 GeV protons. It will extend the accessible kinematic range by two orders of magnitude in both,  $Q^2$  and  $x$ . Besides its potential to vastly improve our knowledge about the proton structure, it will also allow to test QCD and to explore heavy-quark physics.

## 1.2 Deep-inelastic physics

In elastic electron-proton scattering the proton recoils as a whole. The invariant mass of the hadronic system remains exactly the proton mass. In an inelastic reaction the proton is excited, and the hadronic invariant mass after scattering is increased. A reaction is dubbed deep-inelastic if the energy and momentum transfer to the proton are big enough to leave the hadronic system with an invariant mass much bigger than the proton mass.

The proton itself is an extended object, and the elastic-scattering cross-section very rapidly falls when the momentum transfers ( $Q^2$ ) exceed the square of the proton mass. At higher values, inelastic processes dominate.

The study of deep-inelastic electron-nucleon scattering (DIS) soon led to what is now called the naïve parton model. It is assumed that a rapidly moving nucleon can be regarded as a collinear jet of point-like partons which share the momentum. The basic process is thought to be elastic scattering of the incoming electron on one of the partons. The total deep-inelastic scattering cross-section is obtained by incoherently summing the contributions of all partons present in the nucleon.

In the meantime this model has evolved into the quark-parton model. The point-like partons have been identified with the quarks; their interactions with the incoming electron are described by the Standard Model. The structure functions of the nucleon are linear combinations of quark and anti-quark densities. These are chiefly a function of the fractional momentum carried by the struck quark. To a lesser extent, they also depend on the momentum transfer. Information about the quark content of the proton can, in this model, be extracted from measured doubly-differential cross-sections  $d^2\sigma/dx dQ^2$  of deep-inelastic electron-proton scattering.

In the Standard Model reactions are classified as being neutral-current (NC) or charged-current (CC) events. In the NC case, in which the scattered electron preserves its identity, the exchanged particle is the electromagnetic  $\gamma$  or the neutral, weak  $Z^0$  boson. CC interactions are mediated by the two charged bosons  $W^+$ ,  $W^-$ , and the electron is turned into a neutrino, which escapes from the detector.

Because of the large mass of the weak bosons ( $M_Z = 91 \text{ GeV}$ ,  $M_W = 80 \text{ GeV}$ ), the weak interaction becomes fully important only at  $Q^2 \sim M_Z^2, M_W^2 \sim 10000 \text{ GeV}^2$ . At much smaller momentum transfers, ep scattering is dominated by  $\gamma$  exchange. Consult Ingelmann, [In 87], for details on cross-sections and structure functions.

A measurement of the differential cross-sections yields the structure functions, which are, by construction, linear combinations of the quark densities. The electromagnetic structure function,  $F_2^{em} = \sum_j e_j^2(xq_j + x\bar{q}_j)$ , is determined from the NC cross-sections for  $Q^2 < 3000 \text{ GeV}^2$ . Up to this  $Q^2$  the weak contributions are negligible within the statistical errors of a  $200 \text{ pb}^{-1}$  (1 year) data sample. The charged-current cross-section embodies information on  $u + c$  and  $\bar{d} + \bar{s}$  at a  $Q^2$ -scale of  $\sim 10000 \text{ GeV}^2$ .

Single-quark distributions can be obtained by unfolding procedures from combined  $e^\pm p$  NC and CC cross-section measurements. To do this with reasonably small statistical errors, data samples of  $\sim 200 \text{ pb}^{-1}$  for electron and positron runs are needed, corresponding in sum to a few years of data taking. But uncertainties in the absolute calibration of the detectors can make it difficult to merge data sets recorded years apart. One handle to arrive at a mutual normalization is the demand that the low- $Q^2$  NC cross-sections obtained with positrons have to reproduce the  $e^-p$  cross-sections; the electromagnetic structure functions have to be identical in both cases.

## 1.3 Event classification and kinematic range

Obviously, determining the structure functions will be possible only in those kinematic regions ( $x, Q^2$ ) where NC and CC events can be recognized with little losses and small chances for misinterpretations. The step after classification is to sort the NC and CC events into bins of ( $x, Q^2$ ). Due to finite resolutions some events will not go in the right bin, and, conversely, any bin will erroneously receive events that should have gone to its neighbours. The useful range in  $x$  and  $Q^2$  is restricted to a domain where the net corrections do not exceed 20% of the bin content. Miscalibrations of the calorimeter and a misalignment of the detector introduce systematic errors, which are particularly important.

Event classification makes use of two dissimilarities of NC and CC events. The transverse momentum ( $p_t$ ) of the current jet and of the scattered electron balance well in an NC event, but there is a large imbalance in a CC event due to the escaping neutrino. However, limited detector coverage close to the beam and finite calorimeter resolution also cause missing  $p_t$ . The other difference is the energy deposited in total which is smaller in the CC case. Using appropriate cuts in  $p_t$  and  $E_{tot}$  allows to separate NC from CC events. There remains a small kinematic domain,  $6 \text{ GeV}/c < p_t < 20 \text{ GeV}/c$  &  $E_{tot} < 250 \text{ GeV}$ , in case of which an explicit search for the scattered electron is necessary. It can be identified with the calorimeter by comparing the energy deposition in the electromagnetic and hadronic sections. A Monte Carlo study showed that for  $Q^2 > 1000 \text{ GeV}^2$  losses and contaminations in the two samples should be smaller than 1%, [In 87].

The range in the kinematic variables ( $x, Q^2$ ) over which the cross-sections can be measured is not the same in the two event classes.

In NC events,  $x$  and  $Q^2$  will most naturally be reconstructed from the energy and the flight-direction of the outgoing electron. From the dependence of  $x$  and  $Q^2$  on these variables it follows that the relative resolution in  $Q^2$  is the same as the relative energy resolution, except for the case of nearly backward scattered electrons when the angular resolution dominates. Similarly, in that case the resolution in  $x$  is also reduced. Over most of the solid angle the precision in  $x$

is governed by the calorimeter performance. It is worse than the energy resolution by a factor of  $1/y = x_s/Q^2$ , which restricts measurements to  $y > 0.1$  because the corrections of the  $x$ -bin contents would otherwise exceed the 20% limit.

Systematic errors are introduced by a miscalibration of the calorimeter and a misalignment of the tracking-detector. Because of the inverse- $y$  dependence of the relative error in  $x$ , a miscalibration of  $\sim 1\%$  already sets the most stringent limits to the useful kinematics domain, if systematic errors in the cross-sections have to be kept below 10%. The drift chamber alignments are expected to be better than 1 mrad which keeps small the error from this source. In the kinematics domain where systematic errors are less than 10%, the net corrections to the  $(x, Q^2)$  bin contents due to finite detector resolutions are also below 10%.

In CC events, the outgoing lepton is an undetectable neutrino. The kinematics variables have to be deduced from the hadron flow. Besides a poorer resolution, there is additional systematic error introduced because of particles escaping through the beam pipe. The result of a simulation involving the ZEUS detector can be summarized as follows: At low  $x$ , ( $x \leq 0.03$ ), no observable systematic errors occur down to  $Q^2$  as low as  $100 \text{ GeV}^2$ . The resolution in  $x$  and  $Q^2$  is  $\sim 20\%$ . At high  $x$ , however, the systematic errors exceed 30% for both,  $x$  and  $Q^2$ , if  $Q^2 \leq 1000 \text{ GeV}^2$ . They become small only at  $Q^2 \geq 10000 \text{ GeV}^2$ . Again, as in the electron case, a calorimeter miscalibration of 2% restricts the useful range to  $x < 0.5$  while the expected misalignment poses no problem. The lower limit in  $Q^2$  is set by the trigger requirement of  $p_t > 10 \text{ GeV}/c$ , and the range in  $y$  is limited to  $y > 0.03$ .

Summing up, the accessible kinematic range in the determination of cross-sections is:

$$\begin{array}{ll} \text{NC case:} & 5 \cdot 10^{-5} < x < 0.7 \\ & 0.1 < y < 1 \\ \text{CC case:} & 0.01 < x < 0.7 \\ & 0.03 < y < 1 \end{array}$$

It is necessary to measure  $\sigma_{NC}$  and  $\sigma_{CC}$  simultaneously to extract single-quark densities, At  $\sqrt{s} = 314 \text{ GeV}$ , this is possible in the domain

$$\begin{array}{l} 0.01 < x < 0.7 \\ 0.1 < y < 1 \end{array}$$

At the lowest  $x$  this corresponds to  $Q^2$  between  $100 \text{ GeV}^2$  and  $1000 \text{ GeV}^2$  while at  $x = 0.5$  the  $Q^2$  range covered is  $5000 \text{ GeV}^2$  to  $20000 \text{ GeV}^2$ .

At the one existing proton collider, HERA, the structure of the proton will be investigated over an unprecedented kinematical domain, up to momentum transfers of  $20000 \text{ GeV}^2$  and down to  $x \sim 10^{-4}$ . This is an impressive extension to any possible fixed-target experiment.

## 1.4 QCD at HERA

At HERA, QCD can be tested within the context of the QCD-improved parton model. It will be possible to precisely measure the expected scaling violations of the structure functions as well as to determine the gluon content of the proton.

If the quarks were not interacting, but were confined in a common potential well, no further structure could be resolved at higher and higher  $Q^2$ . The quark-density distributions would not depend on  $Q^2$ ; a characteristic called scaling. However, to first-order QCD, the quark struck by the photon probe may just have emitted a gluon that now carries some fraction of the quark's momentum. For that reason, the high- $x$  region depletes with increasing  $Q^2$  while the low- $x$  region becomes more populated. From first-order QCD calculations a logarithmic

scale breaking has been predicted. The starting point are the Altarelli-Parisi equations (APE). These are two coupled, linear integro-differential equations, which approximate QCD to first order. Their kernels are the splitting functions. These give the probability of finding a parton inside a parton with a certain fraction of the parent parton momentum. Here, 'parton' stands for quark or gluon. The input needed for solving the APE are the quark and gluon distributions at a fixed  $Q_0^2$ . Their evolution in  $Q^2$  is then calculated from the APE.

The results from deep-inelastic scattering experiments at HERA will allow us to compare the structure functions at a scale of  $Q^2 \sim 10 \text{ GeV}^2$  obtained from fixed-target experiments with the structure functions at  $Q^2 \sim 5000 \text{ GeV}^2$ .

The complete set of the APE need the gluon structure function  $xg(x)$  as an input. Gluons carry neither weak nor electromagnetic charge; their momentum distribution cannot be obtained from DIS. At HERA, the handle is heavy-quark production by photon-gluon fusion. Such events are recognized by observing the decay of the produced heavy-quark mesons. In extracting  $x_g g(x_g)$  from the measured cross-section, there arise two problems.

The measured  $x$  is not the  $x_g$  of the gluon but the  $x$  of the struck quark-antiquark pair. A more exact relation is  $x_g = x(1 + M_{q\bar{q}}^2/Q^2)$ , in which the transverse momentum of the gluon is still neglected. It is difficult to measure the invariant mass of the two quark jets due finite detector resolution and possible gluon emission from one of the quarks. But the transverse mass with respect to the beam direction,  $M_t^2$ , is a good estimator for the invariant mass of the quark pair. In place of the true  $x_g$  the approximation  $x'_g = x(1 + M_t^2/Q^2)$  is used.

Secondly, the cross-section is not proportional to  $x_g g(x_g)$  but to the integral over the gluon density times the quark-gluon splitting function. The latter gives the probability that the gluon annihilates into  $q\bar{q}$  such that the pair carries a certain fraction of the gluon momentum. One has to employ an unfolding procedure which is far from trivial to arrive at the gluon distribution function. Woudenberg et al. state the accessible kinematic range will be  $10^{-3} < x < 0.1$  at  $Q^2 \approx 0$ , [Wo91].

Other methods for obtaining  $xg(x)$  involve the measurement of the longitudinal structure function, which is of order  $\alpha_s$ , or the observation of resonant production of the  $J/\Psi$  at low  $Q^2$ . Both methods are quite restricted in the kinematic range they cover, but they may serve as cross-checks for the above mentioned technique.

## 1.5 Heavy-quark physics

At HERA, heavy quark-antiquark pairs will be produced abundantly by photon-gluon fusion. For an integrated luminosity of  $100 \text{ pb}^{-1}$  about  $10^8$  charm events and  $10^6$  bottom events are expected. The primary interests in heavy-quark physics are mixing phenomena in the neutral heavy mesons and CP violations. Moreover, the production cross-sections at HERA should be big enough to search for rare, Cabibbo-suppressed decays in order to measure some CKM-matrix elements.

The neutral mesons will be produced in pairs, eg  $B^0 - \bar{B}^0$ . They are not eigenstates of the weak interaction by which they decay; hence, mixing has to occur, [Pe87, chapter 7.14]. It is recognized by observing the semileptonic decays of both produced mesons. If no mixing occurred the two leptons in the final state should always be of opposite sign. Otherwise dilepton final states with like-sign leptons will be found in a certain fraction of the events. The size of this fraction serves as a first rough measure of the amount of mixing. If the decay vertex and the momentum of at least one of the two mesons can be reconstructed it is possible to obtain the oscillation length and the oscillation parameter  $x = \Delta M/\Gamma$ . A sizeable fraction

of the heavy mesons will be produced with large Lorentz boosts of up to  $\gamma_B \sim 15$ ,  $\gamma_D \sim 40$ , leading to decay lengths of  $O(1-2\text{ cm})$  and  $O(0.5\text{ cm})$ , respectively, [A188]. Given a good vertex detector, efficient reconstruction, and unambiguous lepton identification, it should be feasible to measure mixing parameters and a number of individual bottom-hadron lifetimes.

CP violation is possible in the Standard Model with three families if the one irreducible phase of the CKM matrix is nonzero. So far, CP violations have only been observed in the decays of neutral kaons. It was observed that the weak eigenstates  $K_S$  and  $K_L$  are not true CP eigenstates, but that the  $K_L$  (ideally CP=-1) has a small,  $\epsilon = 2.28 \cdot 10^{-3}$ , admixture of the CP=+1 state. Experimentally, this shows up as a small branching ratio  $\Gamma(K_L \rightarrow 2\pi)/\Gamma(K_L \rightarrow \text{all}) = \epsilon$ . Since it is associated with  $K^0 - \bar{K}^0$  mixing  $\epsilon$  is called the  $\Delta S=2$  CP violation parameter.

A  $\Delta S=1$  CP violation ( $\epsilon'$ ) is also possible, but in the Standard Model it is expected to be very small. Measurements at CERN gave  $\epsilon'/\epsilon = (3.3 \pm 1.1) \cdot 10^{-3}$ , [NA3188].

In the case of mesons containing bottom quarks, the situation is reversed with respect to the kaons. The CP violation in the  $B^0 - \bar{B}^0$  mixing (ie  $\Delta B = 2$ ) is predicted to be of the order of  $O(10^{-4})$ . On the other hand, CP violation in the decay of bottom mesons (ie  $\Delta B = 1$ ) may be large. Two elements,  $V_{cb}$  and  $V_{cd}$ , of the CKM matrix have a large ( $O(\sin^3 \theta_C)$ ) contribution from the CP violating phase. Through them a measurable CP violation in  $B$  decays can occur.

The best prospect for finding  $\Delta B = 1$  CP violations is by searching for asymmetries in hadronic decays. Consider a non-leptonic final state  $f$  such that  $B^0$  can decay into  $f$  and  $\bar{B}^0$  can decay into its CP conjugate. If CP is violated the decay rates are unequal. One has to identify one of the  $B$  mesons as  $B$  or  $\bar{B}$  by means of its semileptonic decay and observe the hadronic decay of the other. Note, however, that the number of  $B$  decays needed to measure these asymmetries is large. The branching ratios multiplied with the expected detection efficiencies are of the order of  $O(10^{-5})$ , and the need for tagging costs another factor of ten. Demanding about 100  $B$ 's for a 3 $\sigma$ -effect leads to the requirement of producing  $O(10^6)$   $B$ 's. It will therefore be difficult but not impossible to detect CP violations in the  $B$  system at HERA. The aim is to determine whether the CKM matrix alone can explain the CP violation.

Finally, the top quark is an item for every new collider. It is the as yet undiscovered third u-type quark, which is predicted and needed by the Standard Model for consistency and renormalizability. Recent measurements at the Tevatron at FNAL have established that  $m_t > 89\text{ GeV}$ , at 90% confidence level, [CDF90]. Through radiative corrections, the top mass enters into important parameters of the Standard Model like  $\sin^2 \theta_W$ , the weak-boson mass ratio  $M_W/M_Z$  as well as the  $Z^0$ -width. A global fit using all electroweak data involving these parameters leads to estimates for the top mass of  $m_t = 132_{-37}^{+31}\text{ GeV}$ , [CDF90], and  $m_t = 139 \pm 38\text{ GeV}$ , [Dy91]. The discovery limit at HERA has been estimated to be  $\sim 70\text{ GeV}$ , [A188]. Thus, chances to produce and find the top quark at HERA are dim.

## 1.6 ZEUS—A detector for HERA

ZEUS is an omnipurpose detector not dedicated to perform only one kind of measurement. It is constructed to ensure good reconstruction of the reaction vertices, unambiguous particle identification, and excellent momentum and energy resolution.

Particles with very high momenta, especially the scattered electron, can emerge from the interactions because of the high center-of-mass energy at HERA. Of such particles only the direction of flight can be determined accurately by the tracking chambers. Their momentum resolution deteriorates in direct proportion to the particle momentum. The energy resolution of the calorimeter, on the other hand, improves with increasing particle energies. At an electron

energy of 30 GeV, both instruments perform with the same accuracy. At ZEUS, the calorimeter and the tracking devices will equally share the task of measuring energy and momentum.

The main calorimeter is a depleted-uranium/scintillator calorimeter. The sampling frequency is  $1X_0^1$ . The forward<sup>2</sup> calorimeter is  $7\lambda$  deep. The barrel and rear calorimeter have depths of  $5\lambda$  and  $4\lambda$ , respectively.

The scintillator plates that alternate with the uranium plates are read out by wavelength shifters on both sides. These are cut into strips, which guide the light to the photomultipliers. That way, the front area where the particles enter is structured into readout strips. In the electromagnetic calorimeter (EMC) these are  $5 \times 20\text{ cm}^2$  in size. The readout of the hadronic calorimeter (HAC) parts the area into squares of  $20 \times 20\text{ cm}^2$ . In the rear calorimeter the EMC strips are  $10 \times 20\text{ cm}^2$ , and the second HAC section is missing. Readout in the barrel calorimeter is similar, but with two HAC sections.

The calorimeter is compensating; its response to an electron is the same as to a hadron of equal energy. The electromagnetic and the hadronic calorimeter form a unit. There is no difference in material, sampling frequency, or mode of construction. They only differ in the readout segmentation. This feature offers some noteworthy advantages over non-compensating calorimetry.

- It offers the best possible energy resolution for hadrons. In a non-compensating calorimeter, the response to electrons is bigger than to hadrons by up to 50%. But during the development of a hadronic shower a varying fraction (roughly 30%) of the incident energy is converted into  $\pi^0$ 's. These decay into two  $\gamma$ 's, which causes electromagnetic energy deposition. The relative size of this electromagnetic component is subject to large fluctuations. As a result, the hadron energy resolution is degraded in a conventional calorimeter—not so at ZEUS. Here, an energy resolution of  $17\% / \sqrt{E/\text{GeV}}$  for electrons and an unprecedented  $35\% / \sqrt{E/\text{GeV}}$  for hadrons are achieved.
- The average fraction of hadronic energy that goes into  $\pi^0$ 's rises logarithmically with the energy. A non-compensating hadron calorimeter is therefore intrinsically nonlinear, unless complicated weighting techniques are employed. These involve searching for electromagnetic energy deposition in a finely segmented hadron calorimeter. The signal found in the electromagnetic clusters is added to the rest with a some weighting factor applied. This time consuming procedure is unnecessary at ZEUS, [H189].
- Since there is no need to find electromagnetic showers deep inside the ZEUS calorimeter, the readout segmentation has been made very coarse in depth. There are only three segments in the forward and the barrel region and just two of them in the rear part. The innermost readout segment has a depth of  $25X_0$ , which equals  $1\lambda$ , and serves as the electromagnetic calorimeter (EMC). It is followed by two identical hadronic-readout sections (HAC1 and HAC2) each being  $3\lambda$  in depth. In the barrel region the HAC sections span only  $2\lambda$ , and in the rear HAC2 is missing completely. Note that the naming EMC and HAC only refers to the readout segments. Otherwise there is no difference between the EMC and the HAC sections. The benefit is a great reduction in the number of electronic channels. The ZEUS calorimeter has only  $\approx 12,000$  channels.

<sup>1</sup> $X_0$ : radiation length,  $\lambda$ : nuclear inelastic-interaction length

<sup>2</sup>The proton direction of flight is referred to as forward.

The uranium calorimeter is not deep enough to completely contain all hadronic showers. The iron yoke enclosing the inner calorimeter is instrumented as a backing calorimeter to catch showers that have punched through the main calorimeter. Its resolution is  $100\%/\sqrt{E}$ .

The tracking system chiefly consists of two barrel-shaped drift chambers: a small vertex tracker (VXD) close to the beam pipe is surrounded by the big central track detector (CTD). Additional trackers are installed in the forward and backward direction. Their planes are set perpendicular to the beam closing the openings of the CTD. The combined data of the VXD and the CTD allow to reconstruct the reaction vertex with an accuracy of  $\sim 25\mu\text{m}$ . The momentum resolution is  $\sigma(p)/p = 0.003p/\text{GeV}/c$  plus a constant term of  $\sim 0.5\%$ , for particles moving perpendicularly to the beam axis.

The magnetic field (1.6T) for the tracking system is provided by a superconducting coil. It had to be thin,  $< 1X_0$ , to avoid a deterioration of the energy resolution for electrons and  $\gamma$ 's. Good field homogeneity is achieved by a  $z$ -variation of the number of coil windings per unit length. The iron yoke that returns the magnetic flux houses the main calorimeter, the coil, and the tracking chambers.

Charged-particle identification is a task shared by a number of subsystems.

- Low-momentum particles ( $p < 10\text{GeV}/c$ ) have their kind revealed by  $dE/dx$  measurements performed by the tracking chambers.
- In the forward direction, two transition radiation chambers interleaving the forward-tracker planes will be installed. They provide a  $\pi$ -e rejection power of  $\sim 100$  over the momentum range  $1\text{GeV}/c < p < 30\text{GeV}/c$ .
- A fine-grain sampling at the position of the electromagnetic shower maximum in the calorimeter using silicon detectors ( $3 \times 3\text{cm}^2$  in size) gives another independent hadron rejection of  $\sim 25$  over the momentum range  $2\text{GeV}/c < p < 100\text{GeV}/c$ . Together with the calorimeter this component covers the full solid angle, down to the beam pipe. In the forward calorimeter two fine-grain sampling layers will be installed pushing up the rejection power to about 100.
- Isolated particles can be identified if calorimeter and tracking data are combined. One requires a track pointing at that calorimeter cell for which a hit was reported. The energy and the momentum measured by the two devices must coincide. If, in addition, the major part of the energy is deposited in the EMC the particle will be called an electron. A hadron, on the other hand, should deposit only a small fraction of its energy in the EMC.

The muon system consists of layers of limited-streamer tubes. They surround the whole detector on the inner and on the outer side of the iron return yoke. The yoke is deliberately magnetized; observing the deflection of the muons passing it, provides a measure of their momentum. Multiple scattering in the calorimeter and in the yoke, however, limits the resolution to  $\approx 20\%$ .

Muons emerging at angles of less than  $200\text{mrad}$  do not cross the two central trackers, and their path of flight is nearly parallel to the solenoid field. A forward muon spectrometer then provides the momentum measurement. It employs two toroidal magnets interleaved with planes of drift chambers and limited-streamer tubes. It has been designed to catch muons down to angles of  $100\text{mrad}$  and to feature a momentum resolution of better than 25% for muons between  $30\text{GeV}/c$  and  $100\text{GeV}/c$ .

The luminosity is monitored by measuring the rate of the well-understood elastic-scattering bremsstrahlung process  $ep \rightarrow ep\gamma$ . The momentum transfer to the proton is negligible and the electron and the  $\gamma$  emerge at very small angles ( $\theta < 0.5\text{mrad}$ ). The electron has lost a sizeable fraction of its energy, and it is ejected out of the beam pipe by the next bending magnet it encounters. It is detected in a small calorimeter 35m away from the interaction point. The photon energy is measured by a second calorimeter located close to the proton beam pipe 108 m upstreams. Integrating the differential cross-section over photon energies between 8 GeV and 14 GeV one obtains 24.6mb. The acceptance of the luminosity detector is close to 100%, and at HERA's design luminosity one expects a rate of  $4 \cdot 10^6$  events/s. Therefore, this process is well suited for fast luminosity monitoring. The goal is to achieve an accuracy of  $\approx 1\%$ .

The bunch-crossing rate at HERA will be 11MHz to achieve high event rates. Since no trigger decisions can be made in the time between two bunch crossings, the data have to be stored (in analog or digital form) and the trigger has to work like a pipeline to avoid dead time.

## 2 A hadron–electron separator for ZEUS

### 2.1 Introduction

The particles produced in inelastic electron–proton scattering are the scattered lepton (either an electron or a neutrino), a jet of hadrons associated with the struck quark, and proton debris that largely escapes through the beam pipe. There is more than one particle jet, if a quark–antiquark pair is created or if hard-gluon bremsstrahlung occurs. In a neutral-current reaction the incoming electron preserves its identity and emerges well isolated from the produced hadrons. Beyond the Standard Model, there may be a number of ‘exotic physics’ reactions that could also create isolated electrons. Those of current interest are SUSY<sup>3</sup> events like the production of a ‘selectron’ and its subsequent decay into an electron and a photino or the production of heavy leptoquarks decaying into an electron and a quark. Finally, an excited electron would decay by emitting a high-energy gamma. In any of these cases, the calorimeter together with the tracking devices would be able to identify the electron.

This is different when heavy quarks are to be studied. Especially *B*-mesons can be tagged well by their semileptonic decay which can create an electron inside a hadron jet. At HERA the heavy quarks are predominantly produced by photon–gluon fusion alongside with the much more frequently occurring light quarks. The ratio of electrons, from *c* or *b* decay, to hadrons is of  $O(10^{-3})$  at best. In the forward direction, this value is attained over an electron momentum range of  $0 < p < 30 \text{ GeV}/c$ ; in the backward direction, the momenta range only up to  $10 \text{ GeV}/c$ . Outside the quoted ranges the electron-to-hadron ratio decreases rapidly.

In order to tell electrons from hadrons one makes use of the different behaviours of the two particle types in the various detector components. Appropriate cuts in the quantities measured by each component allow a distinction which, however, can never be perfect. Regardless how the cuts are chosen, there will always be some electrons that are not recognized as such. The probability for a genuine electron to pass the cut, shall be called the electron efficiency. The goal is, of course, to have a high electron efficiency, but there is a trade-off to be accepted: The higher the electron efficiency the less demanding is the cut, and the more hadrons will be mistaken for electrons. The odds for this are called the hadron misidentification probability; naturally it is a function of the electron efficiency. Its inverse is the hadron-rejection power.

If we return to the example of *B*-meson physics, with the electron-to-hadron ratio being of  $O(10^{-3})$ , we note that the hadron-rejection power needed is on the order of  $10^3$ . No detector component can achieve this on its own; the task must be shared.

The tracking chambers provide a hadron-rejection power of  $O(10^2)$  from  $dE/dx$  measurements for particles with momenta below  $10 \text{ GeV}/c$ . In the backward region, however, a cone of  $200 \text{ mrad}$  around the beam pipe is not covered at all. Transition radiation chambers (installed only in the forward direction) add another factor of  $O(10^2)$  up to  $p = 30 \text{ GeV}/c$ . So, especially in the rear direction, electron identification is insufficient without an additional component.

Clearly, including an independent hadron–electron separator that covers the full solid angle and is operative over the whole momentum range indicated would considerably improve the performance of the ZEUS detector.

<sup>3</sup>from super symmetric theories

### 2.2 Hadron–electron separation using a calorimeter

Isolated electrons can be identified using the momentum information from the tracking chambers together with the calorimeter response. First of all, the energy measured by the calorimeter has to match the momentum measured by the tracking devices, to within two standard deviations, and, secondly, there may be only little leakage of energy into the hadronic section. A cut that retains 90% of the electrons falsely accepts 1% of the hadrons as electrons. The method is suitable only if the electron is not part of a hadron jet. To see why, consider a hadron jet of total energy  $100 \text{ GeV}$  containing a  $10 \text{ GeV}$  electron. The electron will deposit its energy in the EMC and the hadron jet may easily add another  $10 \text{ GeV}$ , which are a mere 10% of its total energy. The electron cannot be singled out.

If a calorimeter has a fine granularity both in depth and width, it can achieve a very good  $\pi$ - $e$  separation by recognizing the rather different patterns of hadron and electron showers. An example is the CHARMII neutrino detector. There, the longitudinal profiles of intensity are the same for electromagnetic and hadronic showers, but the lateral spread is much bigger for the latter. By this characteristic the two can easily be distinguished. To do so, the longitudinal sampling and readout frequency of such a calorimeter must be  $\sim X_0$ ; the lateral partitioning should be smaller than the Molière radius  $R_M$ . Hadron misidentification probabilities of  $10^{-3}$  at 90% electron efficiency have been obtained, [CHARMII].

The ZEUS calorimeter has a sampling frequency of  $1 X_0$  throughout the whole depth, but the optical readout sums up the signals of many layers, and its transverse segmentation is only  $\sim 2.5 R_M \times 10 R_M$  ( $R_M = 2.0 \text{ cm}$ ). Therefore, hadron–electron separation by pattern recognition cannot be done with the ZEUS calorimeter.

In an active-target experiment, like the CHARM neutrino detector, the showers initiated by an interaction can start anywhere in the calorimeter, whereas in a collider experiment, all produced particles enter the calorimeter at the front, and the shower development begins there. When an electron enters the calorimeter it loses energy by radiating energetic photons, which in turn undergo pair conversion. This way, a cascade of electrons, positrons, and gammas develops. It fades away when the particle energies become too small. The cascade is most intense a few radiation lengths behind the entrance point. Hadrons, on the other hand, lose their energy by inelastic reactions with the nuclei of the absorber material. At the expense of the particle's energy, more hadrons are created, and a hadronic cascade grows. But its characteristic scale of length, the interaction length, is, in uranium, roughly 30 times bigger than the radiation length. Hence, electron-initiated showers build up in the very first part of the calorimeter while hadronic ones extend deep into it, see figure 1. Recording separately the energy deposition at a suitable depth inside the calorimeter should thus provide a good handle to distinguish hadrons from electrons. As the calorimeter itself does not offer this feature, an independent detector component is needed to accomplish the task—the Hadron–Electron Separator, or HES for short.

It has been decided within the ZEUS collaboration to realize the HES by inserting layers of particle counters into the electromagnetic calorimeter, at a position where the electron-initiated showers are most intense. The counters are silicon detectors with a cross-section area of  $\sim 3 \times 3 \text{ cm}^2$  and a thickness of  $400 \mu\text{m}$ . In the rear and the barrel calorimeter, one such layer is foreseen, and two layers are envisaged in the forward direction. All told, the full-sized HES will cover an area of  $50 \text{ m}^2$  and involve some 50,000 detectors. At the time of writing, a small section of the rear calorimeter has been equipped with a HES; its 2,000 detectors are delivering data. The full-sized HES will be installed only as part of the stage-two upgrade of ZEUS.

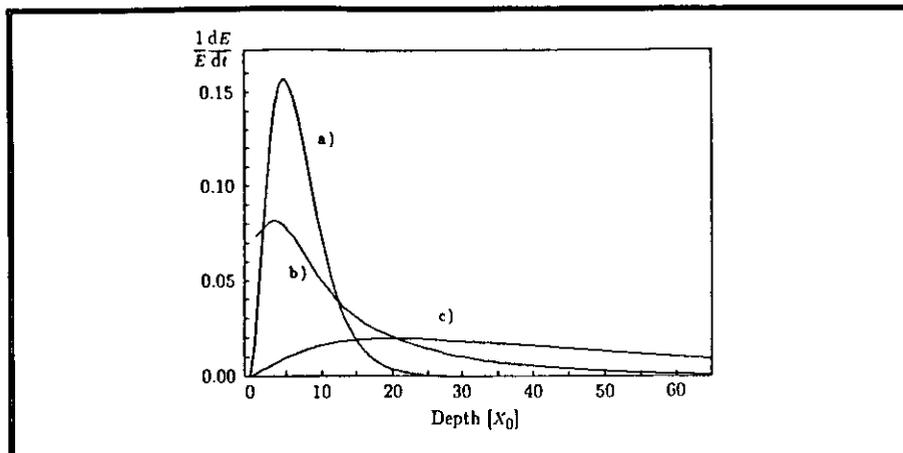


Figure 1: The operating principle of a calorimetric hadron–electron separator. Observe the average shower profiles in the ZEUS calorimeter at  $E = 5\text{ GeV}$ : a) electron shower, b) hadron shower starting at  $x = 0$ , c) average over hadronic showers (starting-point distribution folded in). Reading the energy deposition at eg  $4X_0$  allows to distinguish between hadrons and electrons.

## 2.3 Theoretical aspects

### 2.3.1 Shower development

The radiative energy loss per unit length of relativistic electrons grows linearly with the electron energy, while the ionization loss grows only logarithmically. Both are equal at a critical energy ( $E_c$ ) below which the energy is lost predominantly by ionization. Electrons with energies greater than  $1\text{ GeV}$  will radiate off all but a fraction of  $1/e$  of their energy across one radiation length ( $1X_0$ ) of material. On average a high-energy photon ( $E > 1\text{ GeV}$ ) undergoes pair-conversion within  $(9/7)X_0$ . This way, a cascade of electrons, positrons, and photons develops. As it grows, the particle energies decrease, and it fades away when the average particle energy falls below the critical energy. In a sampling calorimeter with high- $Z$  absorber layers and low- $Z$  readout layers, the signal is mainly due to ionization from the  $e^+$ ,  $e^-$ , whereas the photons contribute only to a small extent, [Wi87].

The mean longitudinal shower profiles are described by a universal function of  $X_0$  and the electron energy, if one scales the length as  $t=x/X_0$  and the energy as  $y = E/E_c$ . This function is well approximated by the density function associated with a gamma distribution. Its two parameters depend on  $X_0$  and weakly on the energy as has been derived from measurements and Monte Carlo calculations; see box on next page.

The profile of energy deposition by an electromagnetic shower, as shown in figure 1, represents an average only. Since the cascade development is governed by random processes, no two showers are alike. Their intensity at a given depth is subject to substantial fluctuations. Figure 3 gives an example of this. Shown are the intensity fluctuations of 1,000 showers at various depths (Monte Carlo calculations, [Pe91]). At  $3X_0$  the average intensity of the  $1\text{ GeV}$

The average energy deposition per unit length (the longitudinal profile) of an electromagnetic shower is described by the density function associated with the gamma distribution, [Jo70].

$$\mathcal{L}(x) = E \frac{b}{\Gamma(a+1)} (bx)^a e^{-bx}$$

$$\int_0^\infty \mathcal{L}(x) dx = E$$

$E$  denotes the electron energy. The parameters  $a$  and  $b$  depend on the absorber material and the electron energy. For the EMC of ZEUS, they are:

$$a = 1.15 + 0.54 \log(E/\text{GeV})$$

$$b = (0.395 + 0.022 \log(E/\text{GeV}))/X_0$$

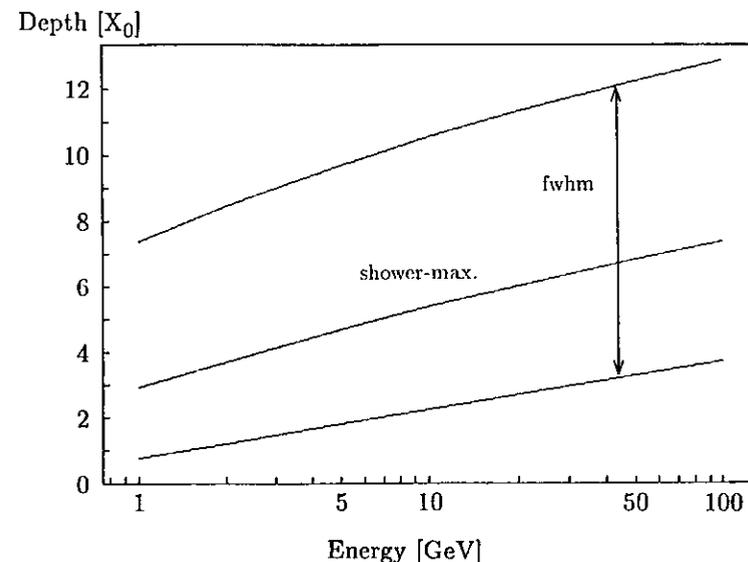


Figure 2: Positions of the shower maximum and of the points of half-maximum intensity v electron energy. The centre line traces the depth at which an electron shower attains, on average, its maximum intensity. The two depths corresponding to half that intensity are indicated by the lower and the top curve, respectively. Note that these are average values only which are subject to substantial fluctuations as the next two pictures will show.

showers attains its maximum. The r.m.s. fluctuations are 55%, and the distribution is nearly Gaussian. At the two depths where the average shower intensity is half the maximum value, the distributions look rather different. They are asymmetric and pile up at small values. A sizeable fraction of the events give pulse heights close to zero.

The goal is to recognize electrons impinging onto the calorimeter by measuring the intensity of the associated particle cascade at a certain depth. In view of the fluctuations, it is clear that the electron identification cannot be 100% efficient. A trade off has to be accepted, but the efficiencies should be high since the intent is to single out rare events. To be specific, the efficiency of recognizing genuine electrons shall be fixed to 90% throughout the remainder of this chapter.

Figure 4 shows the variations of energy deposition of 1GeV and 10GeV electromagnetic showers at a depth of  $4X_0$  which is close to the shower maximum for both energies. In order to correctly identify 90% of the electrons, we demand that the shower intensity exceed 6.1mips and 63mips, respectively.

If we rely on the HES information only, we attribute any event passing the test to an electron. Hadrons should of course not pass the test. But, as we shall see below, they sometimes initiate reactions that produce big enough energy deposits to have them mistaken for an electron.

The most prominent difference of hadronic cascades to electronic ones is their characteristic length, which scales with the nuclear inelastic-interaction length  $\lambda_i$ . It depends on the atomic weight  $A$  of the absorber, but is independent of the particle energy for  $E > 2\text{GeV}$ . In uranium it is about 30 times bigger than  $X_0$ .

When a high-energy hadron hits a nucleus, the most probable reaction is spallation. In a first step, the incoming hadron collides with a nucleon which in turn hits others creating an intranuclear cascade. Pions, but also heavier hadrons, are created at this stage. Together with the cascade nucleons, they escape from the hit nucleus, mainly along the flight direction of the primary hadron. Cascade protons and neutrons are produced in proportion to their abundance in the nucleus. In a second, much slower step, nucleons evaporate from the remaining highly excited nucleus which afterwards may even undergo fission. Because of the high Coulomb barrier present in heavy elements (14MeV in the case of uranium), it is mainly neutrons that evaporate. The cascade nucleons have an exponential energy distribution with an average of 60MeV of kinetic energy. The evaporation neutrons, like fission neutrons, have a Maxwellian distribution corresponding to a temperature of 1.9MeV.

Roughly half of the primary energy goes into hadron production. At 5GeV, on average five hadrons (chiefly pions) emerge from the inelastic reaction. Their number rises logarithmically with the energy. The charged pions will in turn cause further spallations thus creating a hadronic cascade. The neutral ones decay on the spot in two gammas, and an electromagnetic shower starts there. The rest of the primary energy is carried away by the cascade particles and, to a smaller extent, by the evaporation neutrons, the nucleus recoil, and gammas from nuclear de-excitation. Depending on the construction of the calorimeter, a fraction of this energy will turn up as visible energy. The binding energy of the nucleons that have been separated from the nucleus is, however, lost in any case.

The mean longitudinal energy deposition profile of a hadronic shower can be parametrized by gamma distributions in a manner similar to the electromagnetic case. The two components,  $\pi^0$  and purely hadronic, with their different depth profiles have to be folded with the probability density of the first interaction to occur at a certain depth.

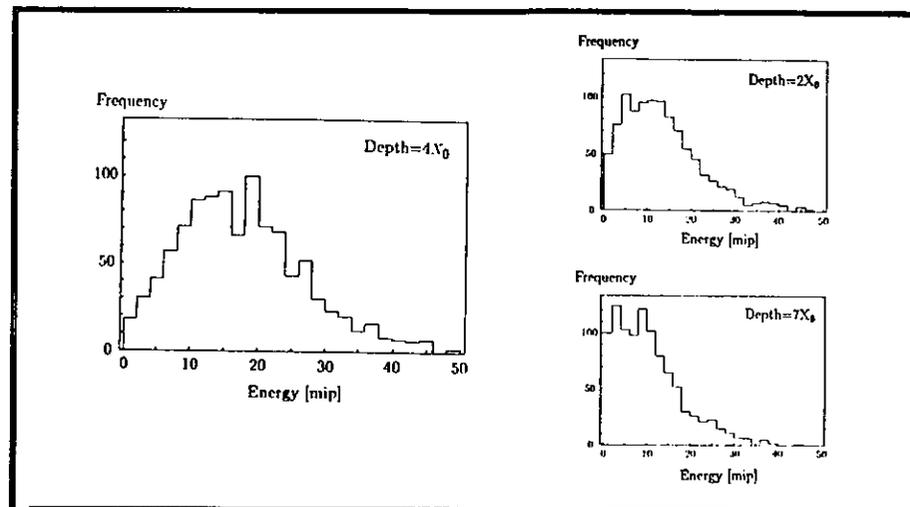


Figure 3: Energy deposition of a 1GeV shower at the shower maximum ( $4X_0$ ) and at those depths where the average deposition is half the maximum value. Note that at the half-height depths a sizeable fraction of the events give pulse heights close to zero. The energy unit 1mip denotes the energy deposited by one minimum-ionizing particle. Pictures by courtesy of del Peso [Pe 91].

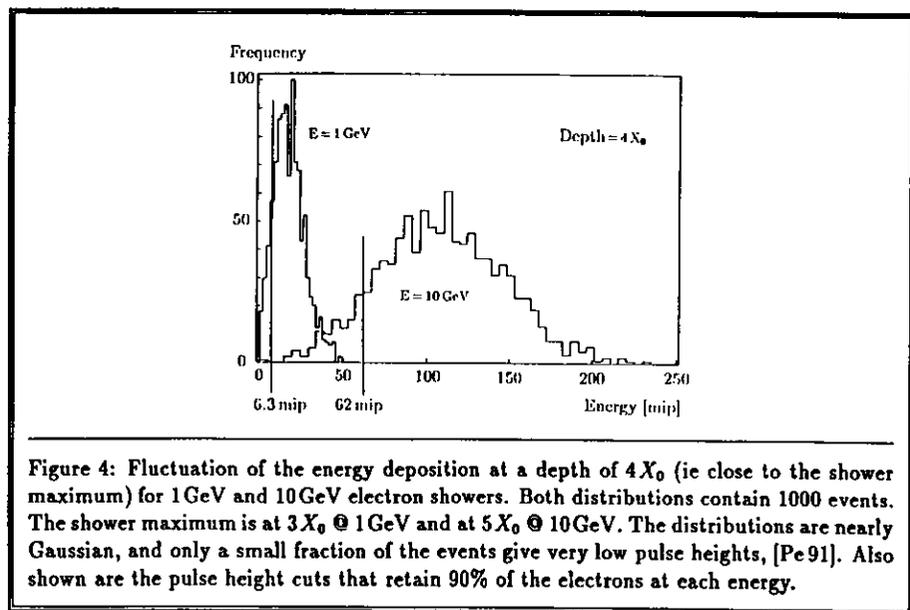
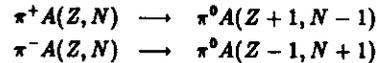


Figure 4: Fluctuation of the energy deposition at a depth of  $4X_0$  (ie close to the shower maximum) for 1GeV and 10GeV electron showers. Both distributions contain 1000 events. The shower maximum is at  $3X_0$  @ 1GeV and at  $5X_0$  @ 10GeV. The distributions are nearly Gaussian, and only a small fraction of the events give very low pulse heights, [Pe91]. Also shown are the pulse height cuts that retain 90% of the electrons at each energy.

### 2.3.2 Limitations

For once, the obtainable hadron-electron separation is limited by the fluctuations of electron-initiated showers. Around shower maximum, the energy deposited in a readout layer, scintillator or silicon, roughly follows a Gaussian distribution, the width of which is large. At a depth of  $4X_0$  the r.m.s. width is 55% at 1 GeV gradually decreasing to 35% at 10 GeV. The hadron-electron discriminating cuts that retain 90% of the electrons have to be placed quite low and are energy dependent. From figure 4 one reads that at 1 GeV the cut has to be as low as 6 mips. Clearly, the hadron-electron separating power is severely limited by a demanding cut like this because there are hadronic reactions which also cause big pulse heights in the silicon detectors.

The one that immediately springs to mind is a charge exchange of the incident pion with a nucleus:



The created  $\pi^0$  practically carries away all of the primary energy because the momentum transfer is small ( $< (0.2 \text{ GeV}/c)^2$ ). It immediately decays into two gammas and an electromagnetic shower instead of a hadronic one is created. Of course, the incident pion is then mistaken for an electron. The cross-section for this process has been measured by a group of physicists at Serpukhov for a number of different nuclei, [Ap 79] [Ap 87]. On  $^{238}\text{U}$  the total charge-exchange cross-section for  $\pi^-$  is  $\sim 0.8 \text{ mb}$  at 39.1 GeV. The probability for a charge exchange to happen within one radiation length is found to be only 0.0012%.

Charge-exchange cross-sections on heavy nuclei have otherwise only been measured below 0.25 GeV of kinetic energy. On lead  $\sigma(\pi^- \text{Pb} \rightarrow \pi^0 \text{X})$  is 0.25 b and  $\sigma(\pi^+ \text{Pb} \rightarrow \pi^0 \text{X})$  is 0.11 b at 0.16 GeV. Hence, at energies  $E > 1 \text{ GeV}$  quasi-elastic charge exchange can be neglected.

But  $\pi^0$ 's are also frequently produced during the inelastic reaction of a high-energy hadron with a uranium nucleus. At 5 GeV on average five pions emerge from the reaction. One or two of them are possibly  $\pi^0$ 's. They have an average energy of  $\sim 0.5 \text{ GeV}$ , and produce quite short electromagnetic showers. The maximum intensity (10 mips) occurs at around  $2X_0$  behind the  $\pi_0$  production point, but this is subject large fluctuations.

Not only  $\pi^0$ -initiated showers but also the cascade protons contribute to the unwanted hadron signal in the silicon detectors once a spallation has occurred. At 5 GeV six spallation protons are produced with an energy of 60 MeV (average values), [Wi 87]. The 3 mm thick ( $1X_0$ ) uranium plates used in the ZEUS calorimeter can stop protons only if their energy is less than 40 MeV. For this reason, the highly ionizing spallation protons mostly reach through the uranium plate in which they have been produced. The cross-section for a spallation reaction is quite independent of the hadron energy if  $E > 2 \text{ GeV}$ , [Ru 66]. The probability for a spallation to occur along the thickness of one uranium plate is 3.2%.

An example shall illustrate the HES response to a spallation reaction induced by a 5 GeV pion: Assume that three charged and two neutral pions as well as six protons of 60 MeV each have been produced. Let the event have occurred in the middle of the uranium plate next to the HES, ie at a depth of 1.6 mm inside the uranium plate. The protons lose some 15 MeV in the uranium. The ionization density in silicon of a 45 MeV proton is about six times that of a mip<sup>4</sup>. The six protons deposit as much energy as  $6 \times 6 = 36 \text{ mips}$ . Add to this the signal of the three charged pions and obtain a HES response equivalent to 39 mips. In order to correctly

<sup>4</sup>mip means minimum ionising particle.

identify 90% of the electrons, the discriminating cut must, at  $E = 5 \text{ GeV}$ , be placed at 23 mips. In our example the pion would have been misidentified as an electron.

If a spallation occurs in the uranium plate next but one to the HES, the signal from the protons is greatly reduced, but it is in part compensated by the  $\pi^0$ -initiated showers, which on average will contribute some 10 mips per shower.

For a realistic analysis, the probability distribution of the number of protons and pions emitted, of their energy, and of the point of production would have to be taken into account. It is clear, however, that inelastic nuclear reactions in the uranium absorber present the most severe limit for hadron-electron separation, at least if only one plane of detectors is used.

## 2.4 Proof of principle

Two major experiments have been performed at CERN during the development phase. The aim of the first experiment was to proof that silicon detectors can be used for hadron-electron separation in a sampling calorimeter. Clearly, the depth at which the HES detectors are located is the important parameter that determines its performance. Best results are to be expected if the diodes are located near the electromagnetic-shower maximum. The higher the electron energy the deeper in the calorimeter do the showers reach their maximum intensity. At 1 GeV and 10 GeV these depths are  $3X_0$ ,  $5X_0$ , respectively. We studied the HES performance at depths from  $2X_0$  to  $8X_0$  using electrons and pions with momenta of 3 GeV/c, 5 GeV/c, and 9 GeV/c. Further efforts were made to study the simultaneous use of two planes of detectors; see also [Dw 88].

### 2.4.1 The setup

In this experiment, the device to contain the HES was an electromagnetic lead-scintillator calorimeter with a total depth of  $23.4X_0$ . It consisted of thirteen identical layers (1 cm of lead, 0.3 cm of scintillator) separated from each other by 0.7 cm wide air gaps into which the HES diodes could be inserted. The scintillators were read out by wavelength shifters, placed left and right, which guided the light to the two photomultipliers. Downstreams, the setup was followed by several hadron calorimeters.

The diodes had a size of  $1 \times 1 \text{ cm}^2$  and a thickness of  $280 \mu\text{m}$ . They were arranged into rectangular patterns of  $4 \times 4$  diodes. The spacing between the diodes was 2 mm. Sixteen diodes were mounted between two epoxy boards. The unit was wrapped in copper foil to provide the necessary electrical shielding. Shielded flat-band cables connected the diodes to the power supply and the preamplifiers.

The preamplifiers and line drivers were installed on top of the calorimeter in order to keep as short as possible the cables between the detectors and their amplifiers. The preamplifiers (MSDII) had been designed by Jarron, [Ja 84]. They are current-to-voltage converters with a transimpedance of 10 k $\Omega$  when connected to an output load of 50  $\Omega$ . Their signal was fed into fast video line drivers with a gain of ten. The output impedance of the line drivers had been matched to the 100  $\Omega$  of the twisted-pair cables used to route the signals into the electronics hut. There, the differential signals were, after further amplification, adapted by ferrite transformers to the 50  $\Omega$  single-ended input of the charge-sensing LeCroy 2282 ADC's. The amplifier chain showed good linearity for signals smaller than 20 mips but saturated above 100 mips. The electronics noise was 1/5 (r.m.s.) of the signal produced by one minimum-ionizing particle.

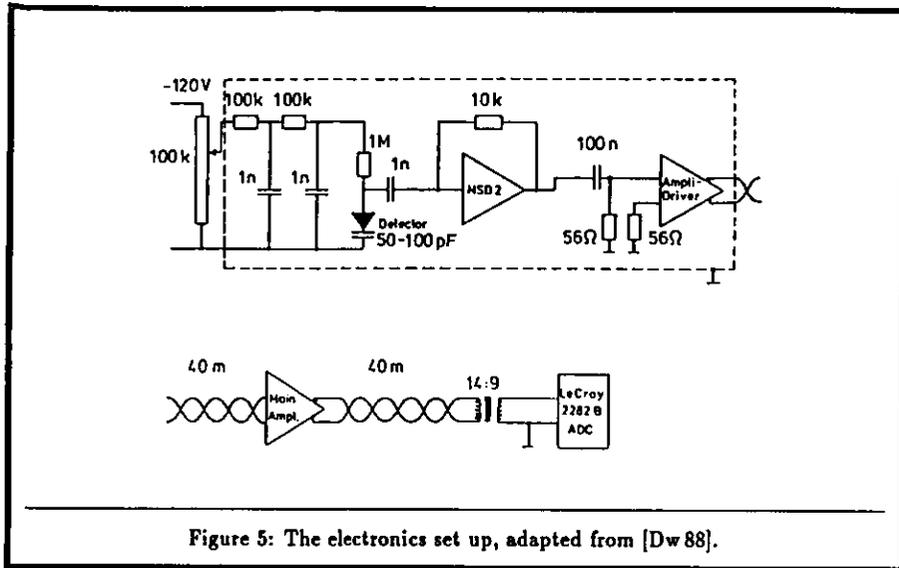


Figure 5: The electronics set up, adapted from [Dw 88].

### 2.4.2 Performing the experiment

Since the test-beam facilities at the CERN proton synchrotron (PS) provide unseparated beams consisting of electrons, pions, protons, and muons, a clean particle identification done by the experimenter is vital. Two Čerenkov counters filled with CO<sub>2</sub> did the task. At 3 and 5 GeV/c, their pressures were set well below the pion threshold, but just above it at 9 GeV/c (albeit the pion efficiency had been negligible). Muons do not induce showers but behave like minimum ionizing particles. Adding together the visible energy deposition in all calorimeters gives for muons a much smaller value than for either electrons or hadrons, and they can thus be excluded from the sample during the offline analysis.

We calibrated each channel by removing the diodes from the calorimeter and letting the beam (3 GeV/c pions) fall onto the centre of each diode. The most frequent signal height of the obtained distribution is in the following referred to as 1 mip. We calibrated the analog chain to up to 100 mip by injecting charge via 0.5 pF capacitors into the input of the preamplifiers.

Close to its maximum, the electromagnetic shower in lead has a diameter of  $\approx 4$  cm (95% containment). In order to catch most of it, it is sufficient to add together the signals of a  $3 \times 3$  subarray of diodes at the centre of which the beam was always aimed. The diode arrays were moved from slot to slot to cover a range of depths from  $2X_0$  to  $8X_0$ . The layers of the calorimeter had a thickness of  $1.8X_0$ , but adding a 0.5 cm ( $0.9X_0$ ) lead plate in front of the calorimeter allowed to half the step size to  $0.9X_0$ . About 10,000 hadron events and 3,000 electron events were recorded at each position and momentum. During a study devoted to using two planes of diodes simultaneously, we varied their mutual distance and their position in the calorimeter.

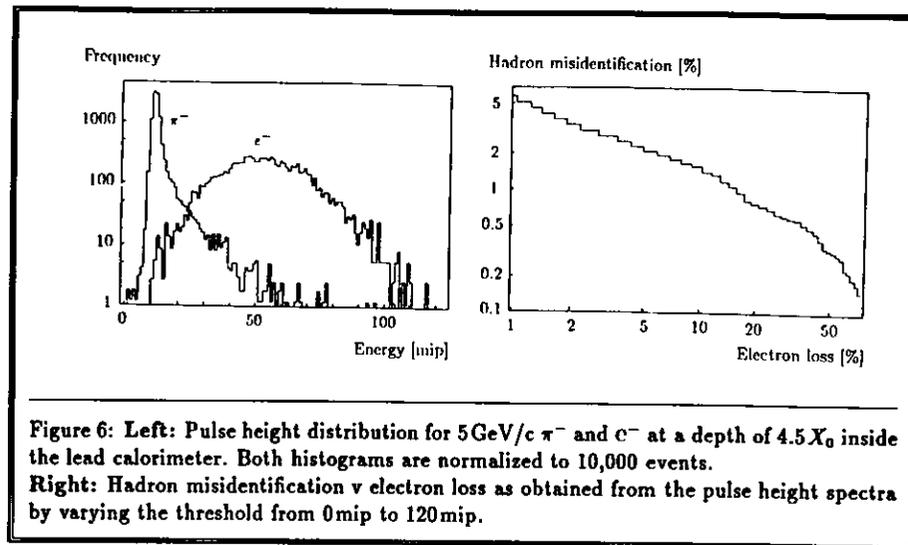


Figure 6: Left: Pulse height distribution for 5 GeV/c  $\pi^-$  and  $e^-$  at a depth of  $4.5X_0$  inside the lead calorimeter. Both histograms are normalized to 10,000 events. Right: Hadron misidentification v electron loss as obtained from the pulse height spectra by varying the threshold from 0 mip to 120 mip.

### 2.4.3 One-plane analysis and results

Figure 6 shows the pulse height distributions caused by 5 GeV/c electrons and pions, as identified by the two Čerenkovs, at a depth of  $4.5X_0$ , i.e. at the electron-shower maximum. The electron distribution peaks at 42 mip and is roughly Gaussian with a standard deviation of 37%. The hadron spectrum peaks at very small signal heights but extends to high signals—comparable with the highest electron signals.

A cut is applied to discriminate hadrons from electrons. Pulse heights bigger than the cut value are attributed to electrons; those being less to hadrons. Not all but only a fraction of the genuine electrons pass the discriminating cut; this fraction is called the electron efficiency of the cut. Some hadrons also pass the cut, and the odds for this to happen are called the hadron misidentification probability; its inverse is the hadron-rejection power.

It is clear from the shape of the hadronic pulse height spectrum, that some hadron contamination of the electron sample (as defined by the pulse height cut) is unavoidable. The contamination reduces with increasing thresholds at the price of reduced electron efficiencies.

Figure 7 shows the variation of the hadron misidentification at a fixed electron efficiency of 90% v the position of the silicon detectors. For data taken with the diodes being located deeper than  $2X_0$  into the calorimeter, the dependence of the hadron misidentification on the plane position can empirically be parametrized using parabolas. The position of the parabola's minimum provides an estimate of the optimum location of the HES plane. At 3, 5, and 9 GeV/c, one obtains 4.3, 4.8, and  $5.2X_0$ , respectively. A comparison with figure 2 shows that the optimum location closely traces the shower maximum<sup>5</sup>. This can be understood from the observation, that close to the shower maximum, the deposited energy is Gaussian-distributed with the smallest relative width and the biggest mean. Thus, the overlap with the hadronic spectrum is smallest in this case.

<sup>5</sup>The small differences between uranium and lead as absorber can be neglected.

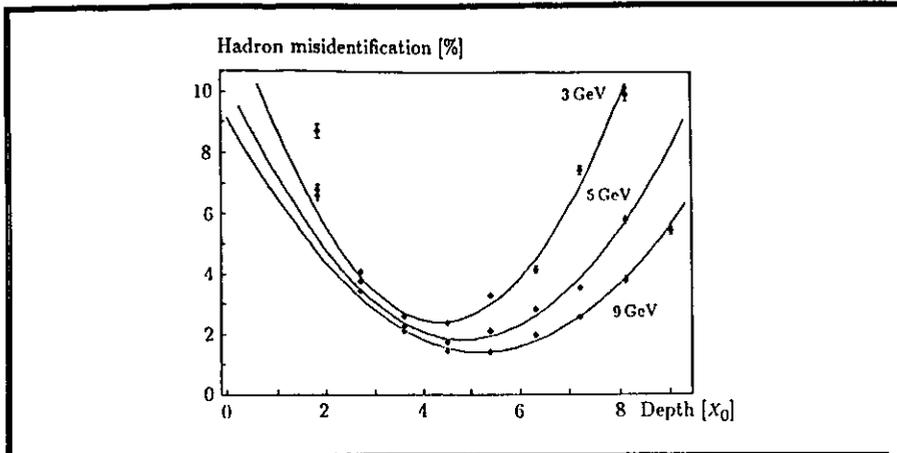


Figure 7: Hadron-misidentification probability at 90% electron efficiency as a function of the depths at which the diode plane was located, with the beam momentum as a parameter. The spacing from layer to layer in the calorimeter is  $1.8X_0$ ; the intermediate points were taken with an additional  $0.9X_0$  of Pb in front of the calorimeter.

#### 2.4.4 Two-plane analysis and results

If the part of the hadronic spectrum with big pulse heights is due to spallation protons, then the high hadronic pulse heights in two different HES planes, separated by more than the spallation-proton range, should be uncorrelated. The simultaneous use of two planes should increase the hadron-rejection power considerably. At best, the two-plane rejection power could be the squared value of the one-plane rejection power. The actual increase in performance, in general, is less than that for a couple of reasons.

- The two planes will both be located off the shower maximum, which means higher fluctuations in the deposited energy and smaller mean values. To ensure high electron efficiencies the thresholds have to be lowered, which increases the hadron contamination.
- At low energies, the second plane (the one located at a greater depth) cannot contribute much, since here the shower already fades away and shows large fluctuations.
- At very high energies, the shower is still in a premature phase in the first plane. Therefore, its contribution to the hadron-rejection power will level off and become constant at high energies.

The distance between the two HES layers should exceed the reach of the spallation protons. Otherwise, the high hadronic signals in both silicon layers would be strongly correlated, and there would be no advantage in using both of them. A rough estimate for the necessary separation is arrived at by folding the exponential energy distribution density of the spallation protons

$$p(E) = \frac{1}{E_0} e^{-E/E_0} \quad E_0 = 60 \text{ MeV for uranium}$$

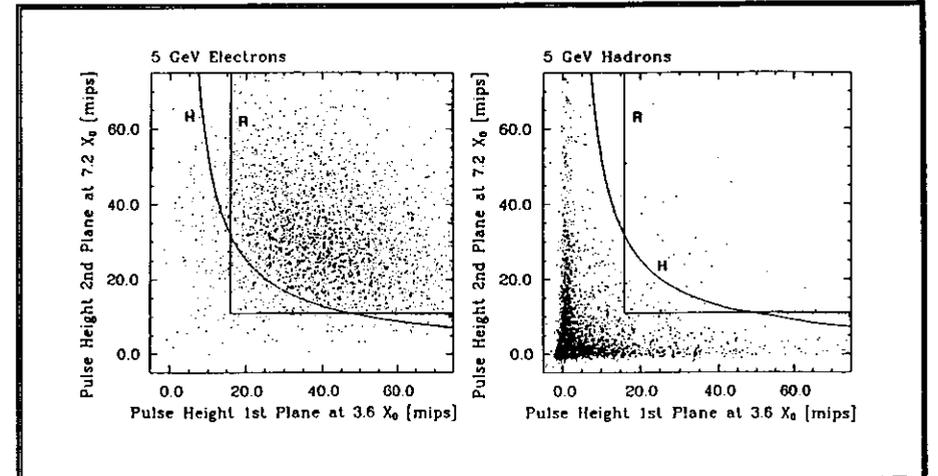


Figure 8: Scatter plot of the pulse heights obtained in plane 1 at a depth of  $3.6X_0$  and in plane 2 at  $7.2X_0$ . The beam momentum was 5 GeV/c. Large hadronic signals are in most of the cases observed only in one of the two layers.

with the mean practical range  $R(E)$  taken from [APS92]. In uranium, this gives:

$$R(E) = \int_0^\infty \frac{R(E)}{E_0} e^{-E/E_0} dE \approx 10 \text{ g cm}^{-2}$$

Using the same value for lead gives  $R \approx 1.6X_0$  as a rough approximation for the distance above which the correlation of the hadronic signals should tend to zero.

While the hadronic correlations call for bigger distances, the plane separation should be considerably smaller than the fwhm of the average shower profile. Otherwise, the electron signals become too small and fluctuate too much. We studied the performance of two different varieties of a two-plane HES. The interplane distance was either  $1.8X_0$  or  $3.6X_0$ , and the first plane was located at depths from  $2X_0$  to  $5X_0$ .

Figure 8 shows a scatter plot of the pulse heights in the first plane v those in the second plane. There are now different ways of applying a two-dimensional cut. If independent and constant thresholds are used, one obtains the 'rectangular' domain as shown by the rectangle marked R. As can be seen, it reaches into the hadron 'cloud' around the origin. A hyperbolic cut, marked H in figure 8, avoids this and leads to a better hadron rejection.

Figure 9 shows the performance when the hyperbolic cut is used. More elaborate procedures have been tried, but with essentially the same outcome; for details see again [Dw88]. The data favour a plane separation of  $3.6X_0$  for low beam momenta while at higher momenta the differences are small. Increasing the separation beyond  $3.6X_0$  again reduces the  $\pi$ -e separating power (not shown).

Along the mean spallation-proton range  $R$  of  $10 \text{ g cm}^{-2}$  the inelastic-reaction probability for an incident high-energy hadron is 5.1% in lead. Assuming that every inelastic reaction produces a signal high enough to be wrongly attributed an electron shower gives an estimate

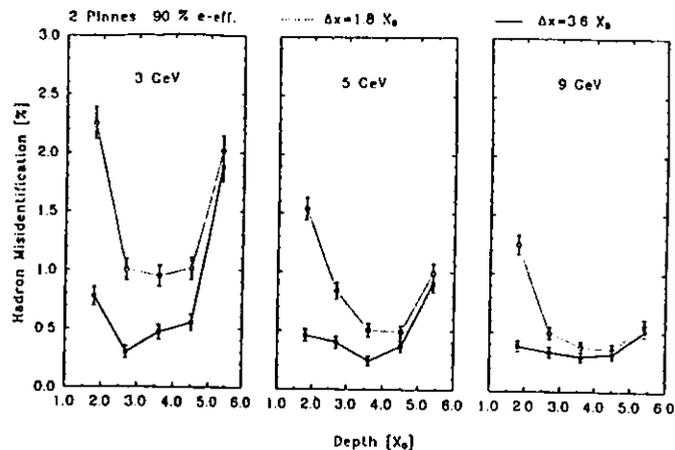


Figure 9: Performance of a two-plane HES. Shown are the hadron misidentification probabilities at 90% electron efficiency v the location of the first plane for two different plane separations ( $\Delta x = 1.8X_0$ ,  $\Delta x = 3.6X_0$ ) and for various beam momenta. A separation of  $3.6X_0$  is preferable, and there is an optimum first-plane location someplace between  $2.5X_0$  and  $4.5X_0$ .

for the combined performance of two planes of 0.25%. For the optimum placement, hadron misidentification probabilities of 0.2% to 0.3% are actually achieved.

#### 2.4.5 Conclusion

The experiment has shown that a calorimetric hadron–electron separator can be realized with  $3 \times 3 \text{ cm}^2$  silicon detectors built into a lead–scintillator calorimeter.

The operation principle can be understood on grounds of the differences between the energy deposition profiles of electron and hadron showers on one side, and the adverse effect of inelastic hadronic reactions (chiefly spallations) if they occur in the very first part of the calorimeter, on the other side. The best hadron–electron separation is achieved by installing the HES in vicinity to the electromagnetic–shower maximum, because there the distribution of deposited energy is Gaussian with a minimal relative width. Thus, its overlap with the spectrum of hadronic energy deposition is smallest, which ensures optimum performance. Using the results from the lead calorimeter, the decision at which depth a one-plane HES should be installed into the ZEUS-calorimeter could be made.

Electrons from  $B$ -meson decay that hit the ZEUS rear calorimeter (RCAL) mostly have momenta below 10 GeV/c. Silicon detectors will go behind the fourth scintillator layer in the RCAL. Together with some material in front of it, this corresponds to a depth of  $3.4X_0$ . A barrel HES is foreseen at the same depth in the barrel calorimeter. The electron momenta tend to be higher in this region, but due to the presence of the superconducting solenoid ( $1X_0$ ) the actual depth is also somewhat bigger, namely  $4.1X_0$ .

In the forward region, where the electron energies span a wide range, an improved HES using two consecutive layers of detectors is foreseen. High hadronic pulse heights, caused by nearby spallation of a uranium atom, are uncorrelated if the layer separation is bigger than the reach of the low-energetic disintegration products. A separation of  $3.6X_0$  yields good results:  $\sim 0.5\%$  hadron misidentification probability at 90% electron efficiency. As the two HES planes are necessarily located off shower maximum, their combined performance is a little worse than naively expected, because of increased fluctuations of the electron signal.

Such a double HES is envisaged for the FCAL. Space for it has been reserved behind the 3rd and 6th scintillator, which corresponds to depths of  $3.3X_0$  and  $6.3X_0$ .

## 2.5 Prototyping

The first experiment had been concerned mainly with a proof of principle for a silicon-detector-based HES operating inside a sampling calorimeter. It led to the design of a two-plane HES prototype. An extensive programme was devoted to evaluate its performance in a realistic environment. We intended to run a version with diode dimensions and separations as close as possible to the final one. It was to be installed in a uranium calorimeter that represented the ZEUS calorimeter in all important physical details. With this setup we determined the achievable hadron–electron separation and measured to what precision the electron impact point can be reconstructed from the HES data.

### 2.5.1 The HES prototype

This experiment has been performed in collaboration with the ZEUS-FCAL group who were extensively testing a (forward) calorimeter prototype. It consists of four identical modules placed besides each other. Each module has a depth of seven interaction lengths ( $=180X_0$ ), a width of 20cm, and a height of 80cm (active area). The optical readout of each module segments the front area into sixteen  $5 \times 20 \text{ cm}^2$  strips in the electromagnetic section and into four  $20 \times 20 \text{ cm}^2$  towers in the hadronic section. Gaps following the fourth and seventh scintillator layer had, in all four modules, been kept clear for the installation of the HES silicon detectors.

The two inner modules had been equipped with two planes of silicon detectors at a depth of  $3X_0$  and  $6X_0$ . Counted bottom up, the diodes covered all of the second and a part of the third HAC-tower; see figure 10. Each plane in each module held 46 diodes. They were distributed over three identical boards mounted besides each other into the gaps held free for the HES. Tungsten carbide spacers were used every 20cm to keep clear the HES gaps, and diodes had to be skipped from the pattern at their positions.

Together with the diodes, the preamplifiers and line drivers were mounted on the boards. The signals were led into the control room via 80m of  $100\Omega$  twisted-pair cables. Ferrite transformers were used to match the  $100\Omega$  differential-signal source to the  $50\Omega$  single-ended input of the charge-sensing ADC's. The ferrites also helped to filter out picked-up low-frequency noise. The neighbouring table summarizes the properties of the amplifier chain.

charge gain	2000
linear range	20 mip @ error <2%
saturation	above 120 mip
noise	0.8 mip fwhm
pulse length	75 ns fwhm

Performance of the amplifier chain with a detector ( $C_{det} = 350\text{pF}$ ) connected to its input.

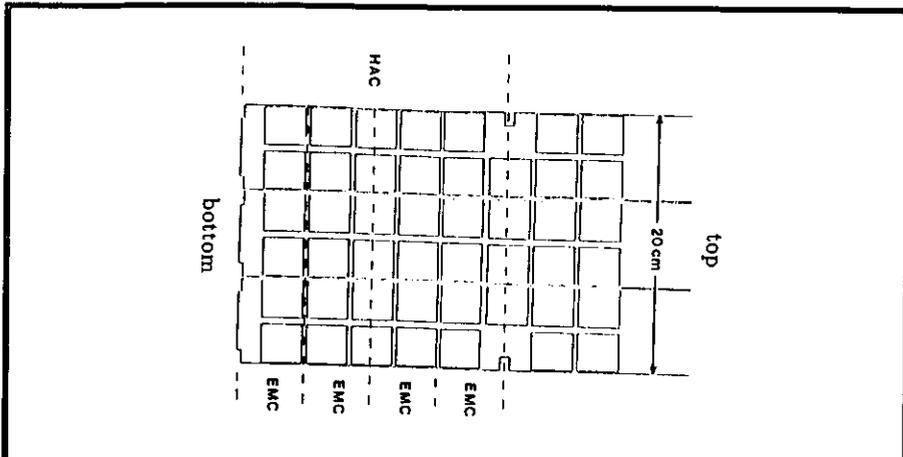


Figure 10: Position of the HES-diodes against the readout-strips of the electromagnetic calorimeter section (EMC) and against the hadronic section (HAC).

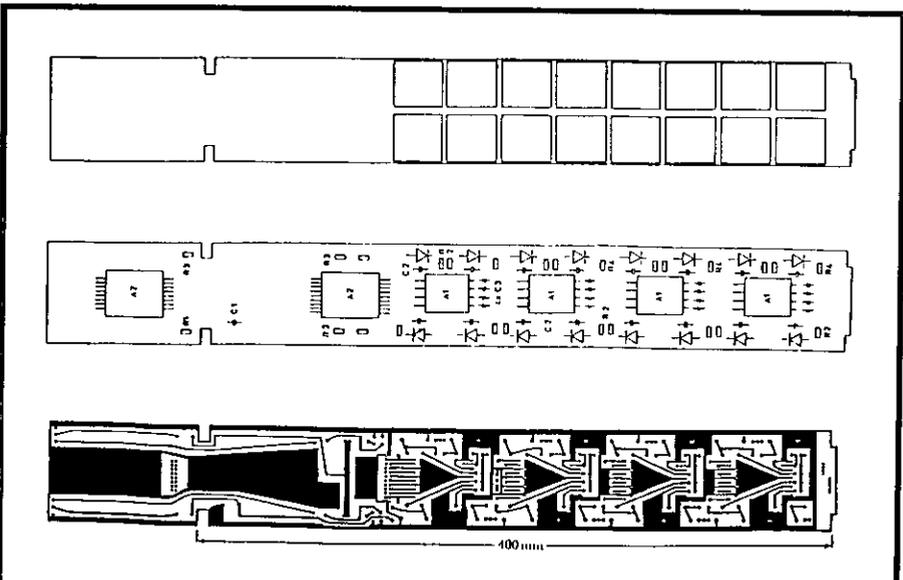


Figure 11: One of twelve diode boards with the 16 diodes mounted on one side and the amplifiers on the other. A1 denotes four MSDII preamplifiers on one ceramic board, and A2 denotes the line drivers, four on each ceramic board, two ceramic boards on top of each other. The passive elements are found again in the next figure.

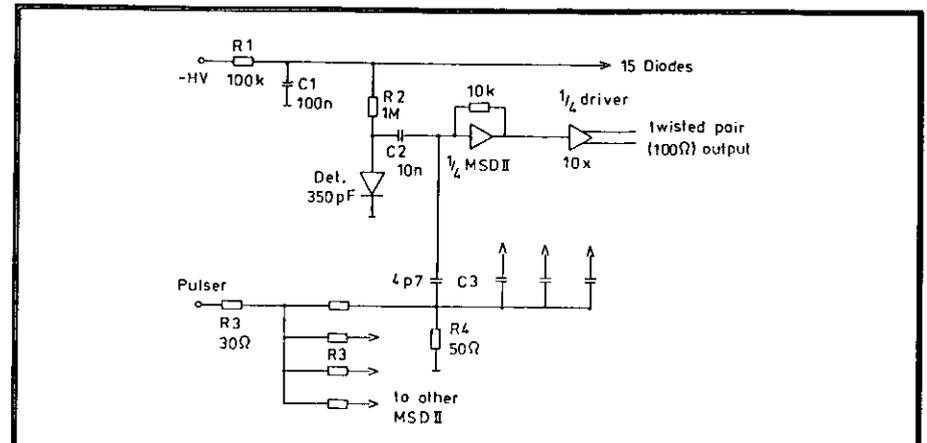


Figure 12: Electrical scheme of a diode board. Shown is the connection of the diodes to the high voltage supply (HV) and to the amplifiers (for one diode only). The calibration pulses are split via the five  $R_4$  resistors into four equal parts. Each part then drives all four test input capacitors  $C_3$  belonging to one MSDII card.

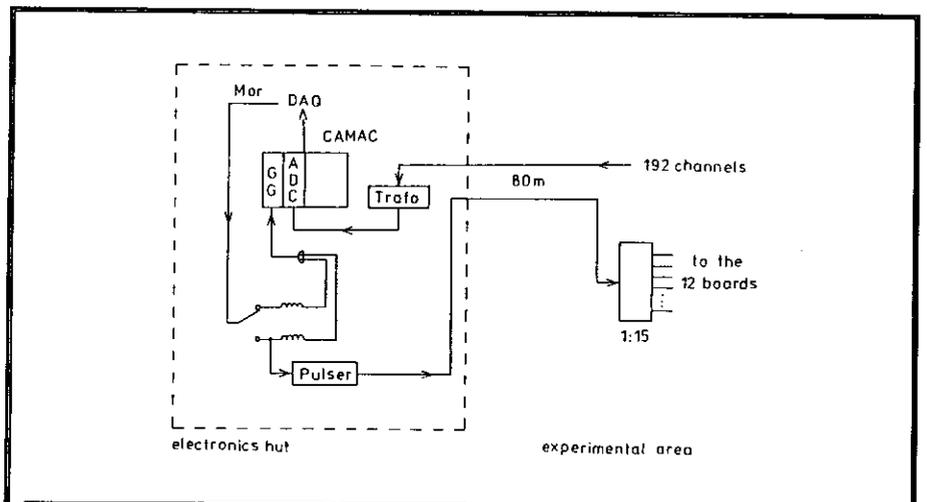


Figure 13: The data-taking logic, HES part only. Pulser: Variable-voltage pulser. Trafo: A 19" cassette housing the 192 ferrite transformers needed to adapt the signals to the ADC. ADC: LeCroy 2282B single-ended charge-sensing ADC. GG: A Camac-housed, programmable LeCroy gate generator, needed by the ADC. DAQ: Data acquisition system to which the data are sent. MOR: 'Master Or'; denotes the trigger signal indicating that an event has been accepted.

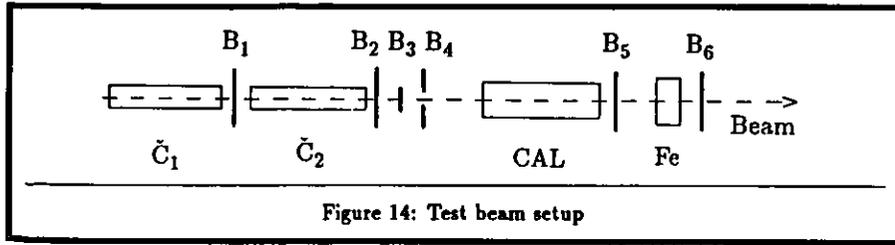


Figure 14: Test beam setup

All amplifiers were calibrated simultaneously with a pulser. Its output was split and routed to the twelve boards. On each board, the pulse was further divided and fed into the calibration inputs of the amplifiers, cf figure 12. The scheme allowed an absolute calibration to a precision of  $\sim 5\%$ .

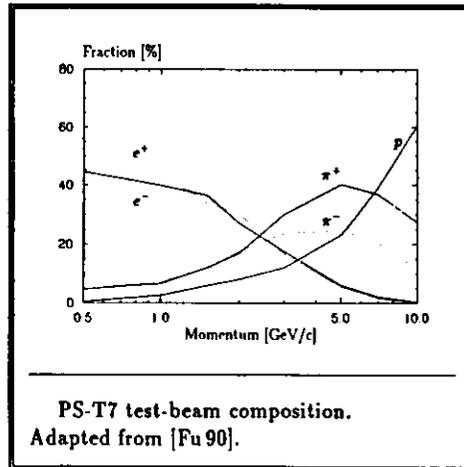
The diodes had been manufactured by Hamamatsu. The thickness and size were  $280\mu\text{m}$  and  $2.9 \times 3.1\text{ cm}^2$ , respectively. They became fully depleted at voltages at or below  $60\text{ V}$ ; all of them were operated at  $80\text{ V}$ . The dark currents at the temperature of operation ( $35^\circ\text{C}$  to  $40^\circ\text{C}$ ) were less than  $200\text{ nA}$  for each diode. After having been installed in the calorimeter prototype for about one year, no change in their characteristics could be detected.

### 2.5.2 Test beam setup

The arrangement at the test beam facilities is shown schematically in figure 14. A beam event was defined as a coincidence between the signals from two  $10 \times 10\text{ cm}^2$  paddle counters ( $B_1, B_2$ ) and an anti-coincidence with a large paddle counter with a  $2\text{ cm}$  hole ( $B_4$ ). The beam cross-section area was about  $1\text{ cm}^2$ . It could be further constrained by a  $3\text{ mm}$ -wide finger counter ( $B_3$ ) mounted either horizontally or vertically for measurements during which the beam was scanned across the gap between two diodes. The particle definition used the information of one of the Čerenkov counters ( $C_1$ ) defining an electron as  $\text{BEAM} \& C_1$  and a hadron as  $\text{BEAM} \& \bar{C}_1$ . In any case, the pulse heights of both,  $C_1$  and  $C_2$ , were always recorded.

To perform the experiments, pure hadron samples with a low contamination by other particles (electrons or muons) are needed. The admixture has to be less than  $0.1\%$ . For the electron samples the restrictions are not that severe. Admixtures of  $1\%$  can be tolerated for we only ask for electron efficiencies of  $90\%$ . The test beams available at CERN are secondary beams coming from a target onto which the primary proton beam is steered. They are a mixture of hadrons, electrons, and muons, and some effort is needed to get clean samples of each particle type.

At the CERN proton synchrotron, data were taken with particle momenta below  $10\text{ GeV}/c$ . The electron efficiency of the



PS-T7 test-beam composition.  
Adapted from [Fu 90].

two Čerenkov counters was very high, and a suitable cut on the their response retained practically all electrons and rejected all hadrons. The hadron sample was produced by requiring that there be no signal in any of the Čerenkovs. The particle composition of the T7 test beam has been measured by the ZEUS-FCAL group using the calorimeter as well as time-of-flight information; see the side figure on the previous page.

At the super proton synchrotron, data were taken up to  $100\text{ GeV}/c$ . The efficiencies of the Čerenkovs were down from nearly  $100\%$  at  $20\text{ GeV}/c$  to  $20\%$  at  $100\text{ GeV}/c$ , and they cannot be used to produce a clean hadron sample. But at this site, quite pure beams are available. The protons ( $450\text{ GeV}$ ) hit a primary target from which secondary particles emerge. These can be selected according to their momentum ( $135\text{ GeV}/c$  to  $200\text{ GeV}/c$ ) and are then guided to a secondary target which can be selected by the user group. Downstream follows the usual beam optics for momentum selection and to focus the tertiary beam into the experimental area.

Pure electron beams can be generated using the lead target—thickness:  $4\text{ mm} \approx 0.7X_0$ . There, electrons from the secondary beam produce hard bremsstrahlung gammas. Behind the target, all charged particles are swept away, and the gammas hit a lead converter ( $6\text{ mm}$ ) at which electrons and positrons are produced by pair creation. Of these the electrons were used for experimentation.

Hadron beams are produced by moving the copper target into the secondary beam. Its length is  $40\text{ cm}$ , which corresponds to  $2.66$  nuclear-interaction lengths and  $28$  radiation lengths. The copper target absorbs the electrons and gammas and enhances the number of hadrons in the beam. This works well if the energy of the secondary beam is chosen not too high for the desired hadron energies. At very low momenta, a further cleaning is needed using the Čerenkov counters. Details about the beam composition can be found in [Gr87].

Muons can be extracted from the particle samples by their known energy deposition in the whole calorimeter. At high momenta, several  $\text{GeV}/c$ , they behave like minimum-ionizing particles and deposit  $2.3\text{ GeV}$  on traversing the calorimeter. Low-momentum beams contain nearly no muons and the Čerenkovs are set to add them to the electron sample. At energies below  $100\text{ GeV}/c$  muons do not shower up, nor do they spallate atoms, ie they behave like gold-plated hadrons. Adding them to the hadron sample would fake an unrealistically good hadron-rejection power.

Note that calorimeter information cannot be honestly used to provide purer particle samples because this would introduce unwanted correlations. If, for instance, only those particles that deposit most of their energy in the hadronic sections of the calorimeter were called hadrons, there would be a severe bias in the data set. Obviously, these particles would be hadrons that showered up late in the calorimeter. They would then surely not have showered up in the very first part of the calorimeter where the HES is residing. The hadron misidentification would come out much too good.

### 2.5.3 Analysis and results

Figure 15 shows the hadron misidentification at  $90\%$  electron efficiency over the whole momentum range. The effectiveness of the first plane proves to be fairly constant for momenta of  $2\text{ GeV}/c$  and bigger. It gives a hadron misidentification of about  $4\%$ . This has to be attributed to the fact that a depth of  $3X_0$  corresponds to a quite early stage in the growth of the electromagnetic cascade, where the fluctuations are large at all energies. For  $1\text{ GeV}/c$  electromagnetic showers the relative standard deviation is never below  $50\%$ , at any depth, which necessarily spells trouble.

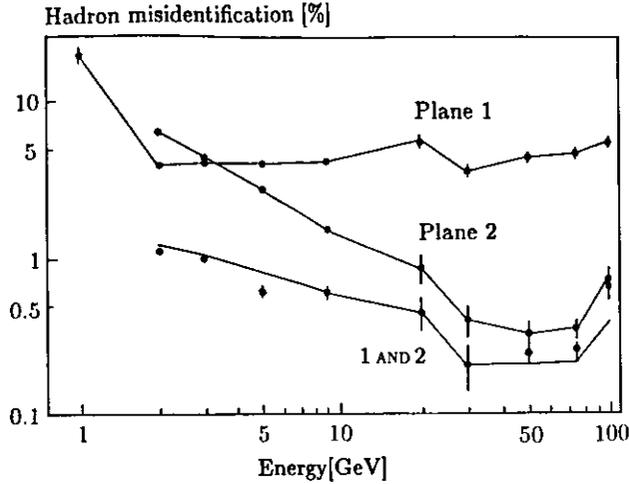


Figure 15: Hadron misidentification probability at 90% electron efficiency over the whole momentum range. The curves show how the effectiveness of each plane on its own as well as their combined action varies with the particle momentum. The lines connecting the data points from plane 1 and plane 2 are merely meant to guide the eye. The line for the two-plane performance (1 AND 2) shows the calculated values when using formula 1, see text.

The performance of the second plane improves with energy, because the maximum of the electromagnetic showers is moving towards it. At 20 GeV/c, the shower maximum has reached the second plane, and the improvement of its performance levels off. At 100 GeV/c, the shower maximum already lies  $1.3X_0$  past it, and its effectiveness degrades.

Over the momentum range 2 GeV/c to 75 GeV/c the combined action ( $R_{12}$ ) of the two planes can be calculated from the two single-plane rejection powers ( $R_1, R_2$ ), at 90% electron efficiency, by the empirical formula:

$$\sqrt{R_{12}} = \sqrt{R_1} + \sqrt{R_2} \quad (1)$$

This approximation has been used in figure 15 and it is indicated there by the solid line labelled '1 AND 2'.

#### 2.5.4 Scans across diodes

Ideally, the active detector areas would completely cover the HES planes, but due to the construction and the mechanical tolerances in mounting, some spacing between them is unavoidable. Their mutual separation is 3mm in the present setup. To study the influence of these gaps the particle beams were scanned in 3mm steps from the centre of one diode to the

next. The transverse beam size had been constrained to 3mm using a finger counter in the beam-defining coincidence. The topics under investigation were the signal loss due to the gaps, the spatial resolution, and the change in hadron-rejection power.

If the electron beam hits the centre of a diode, about 90% of the total signal is collected there. This fraction reduces for off-centre incidence. Figure 16 shows what happens when the beam is stepped across the gap. Both diode responses have been normalized to 100% for the centrally hitting beam. From the two half-height positions one reads that the effective separation of the active areas is 2.9mm. The sum of the two signals is shown in the upper curve. It reduces to 70% in the middle of the gap. The signal loss at other gap sizes is found from a gedanken-experiment: Shift (spatially) the detector response function of one diode with respect to the other and compute the sum function. The effective detector separation will be 0.9mm in the final HES realization. By the above method, one finds a signal loss of about 5% in the gap. This non-uniformity is small enough to have no consequences for the HES performance.

The transverse readout segmentation of the HES is  $3 \times 3 \text{ cm}^2$ . The spatial resolution of the HES is at least 3cm, but it is better for electrons and gammas. If the electron shower signal is shared among different diodes, the individual fractions can be used to reconstruct the electron impact co-ordinates. A simple approach for the one dimensional case (centre-to-centre direction only) is depicted in figure 17. One calculates the ratio:

$$R = \frac{E_1}{E_1 + E_2}$$

as a function of the position  $x$ . The two energy depositions are denoted by  $E_1$  and  $E_2$ , respectively. The data points show the average values of  $R$  at each position. The vertical lines give the r.m.s. fluctuations of  $R$  that originate from the event-to-event shower fluctuations. Inverting  $R(x)$  yields  $x(R)$ , i.e. the electron position as a function of  $R$ . The expected standard deviation of the reconstructed coordinate is then:

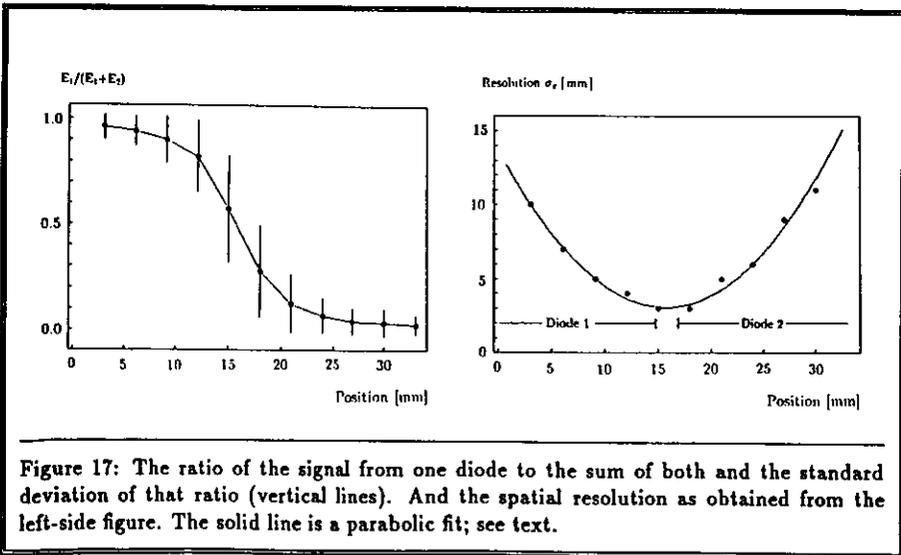
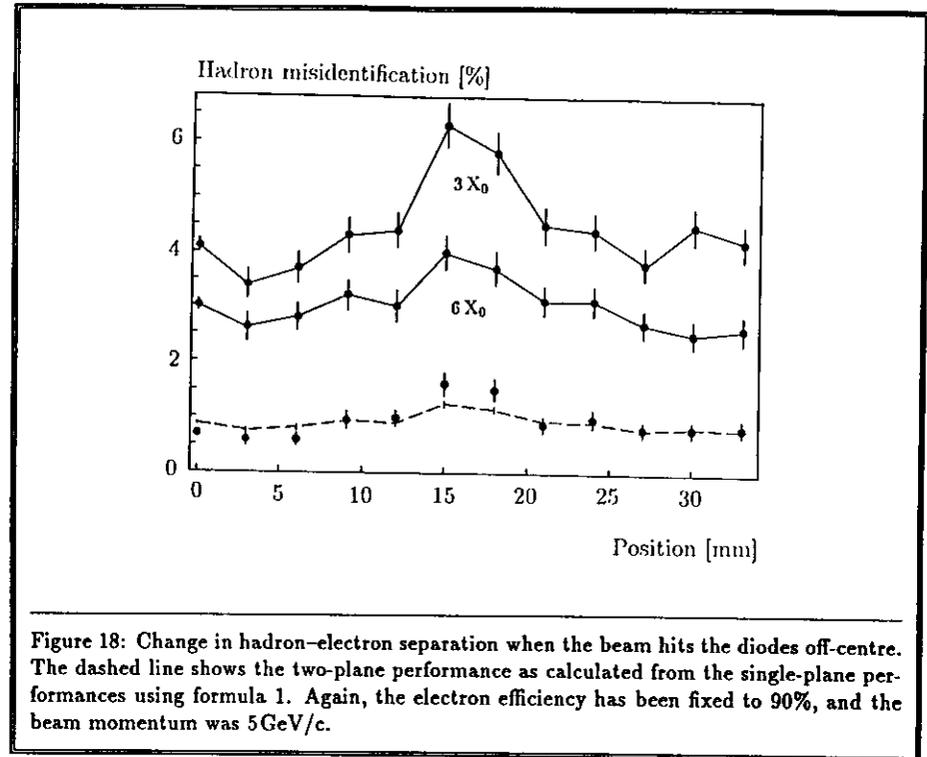
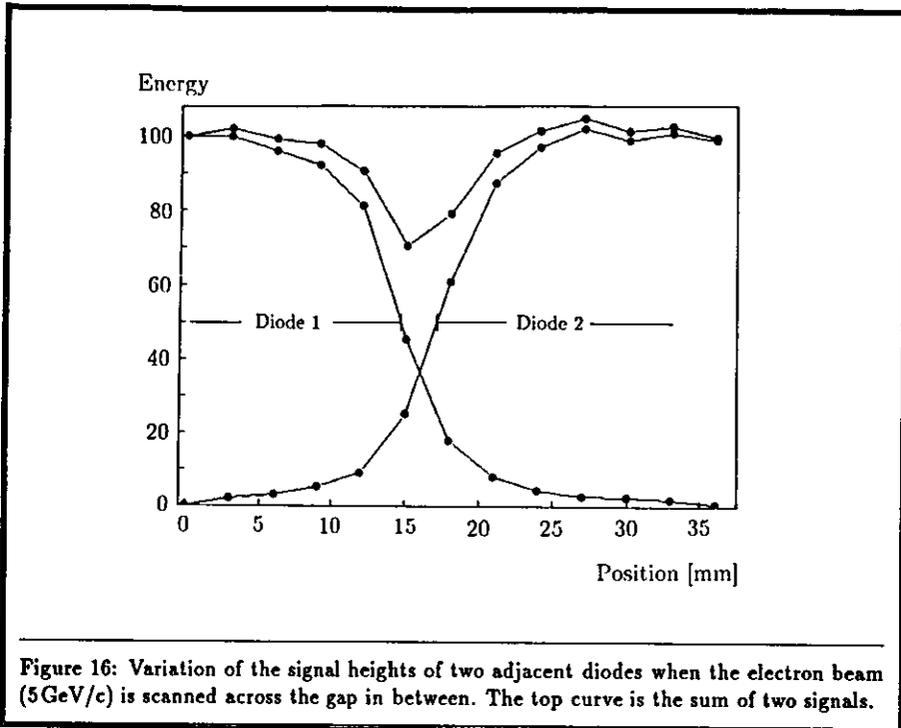
$$\sigma_x = \frac{dx}{dR} \sigma_R$$

In figure 17,  $\sigma_x$  is plotted against the beam position, in a scan where the beam momentum had been set to 5 GeV/c. The solid line is a parabolic parametrization:

$$\sigma_x = 3 \text{ mm} + 10.6 \text{ mm} \left( \frac{x - x_{gap}}{x_{gap}} \right)^2$$

$x_{gap}$  denotes the distance of the middle of the gap to the centre of the next diode. The beam width was restricted to 3mm by a finger counter. Assuming that the beam intensity was constant over this width gives a standard deviation due to its spatial extension of  $3 \text{ mm} / \sqrt{12}$ . This value should have been subtracted quadratically from the quoted  $\sigma_x$ , but it is too small to be of importance.

The spatial resolution is best in the gap because there the change of  $R$  with  $x$  is largest. Moving away from the gap, it worsens rapidly. This behaviour can be understood by considering that an electromagnetic shower consists of a dense core with a diameter less than  $R_M$  and a broad halo, that is subject to large fluctuations. For electron impacts not close to the gap, the position measurement relies on the energy content of the halo, and the position resolution is decreased.



Doing the same scans as before, we measured the HES performance as a function of the beam position. The upper and lower curve in figure 18 show the hadron misidentification probability at 90% electron efficiency when the two planes, at  $3X_0$  and  $6X_0$ , are used separately. The lowest points give the combined performance and the dashed line shows the calculated performance using again formula 1. It approximates the data well, except the two data points taken very close to the gap. The single-plane hadron-rejection power is degraded by a factor of 1.5, and the combined power goes down by a factor of 2 in or close to the gap.

There are two reasons for this degradation. In the gap, the electron signal is on average reduced by 30%, and the fluctuations in deposited energy (sum of both diodes) become larger, thus lowering the cuts needed to have the 90% electron efficiency. The lateral spread of the hadronic energy deposition is considerably larger than in the electron case. This is especially true for the high-energetic part of the hadron spectrum. Therefore, the average loss of the unwanted hadronic signal is smaller than the electron signal loss. Both effects tend to decrease the hadron-rejection power.

## 2.6 Summary

A calorimetric hadron-electron separator (HES) can be realized by exploiting the differences between the longitudinal (in depth) development of electromagnetic and hadronic showers in the ZEUS main calorimeter, which employs uranium as the absorber material. The HES consists of layers of  $3 \times 3 \text{ cm}^2$  silicon detectors used as particle counters. These are built into the ZEUS calorimeter at a suitable depth.

At the ZEUS experiment the task for the HES is to single out electrons, eg originating from heavy-quark decays, inside a dense jet of many hadrons. The challenge is to correctly identify most of the electrons with as little as possible a chance for the hadrons to be misidentified as electrons.

Electromagnetic showers are confined to the very beginning of the calorimeter while hadronic ones distribute their energy more evenly throughout the whole depth of the calorimeter. In order to recognize electrons, the HES has to be located close to the depth of maximum intensity of the electron-initiated showers. The intensity of a hadronic shower is at this shallow depth, on average and in the ZEUS calorimeter, very small compared to that of an electron shower. However, debris from hadron-induced spallation of an atom just in front of the calorimeter will occasionally produce a big energy deposition in the HES detectors thereby faking an electron shower.

For particle momenta between  $2 \text{ GeV}/c$  and  $10 \text{ GeV}/c$ , a discriminating cut in the energy deposition in the HES detectors that correctly identifies 90% of the genuine electrons, will mistake 5% of the hadrons for an electron. This best performance is achieved if the HES is located at a depth of three to four radiation lengths inside the calorimeter, as viewed from the interaction point.

A much bigger range of particle momenta is successfully covered if the HES consists of two layers of silicon detectors, at depths of four and seven radiation lengths. In that case, hadron misidentification probabilities are about 1% for particle momenta between  $2 \text{ GeV}/c$  and  $75 \text{ GeV}/c$  while the odds for correctly identifying the electrons are 90%.

The spatial resolution of the reconstructed impact co-ordinate of an electron is similar for a one-layer and a two-layer HES. Basically, it is set by the size of the HES detectors  $3 \times 3 \text{ cm}^2$ , but it improves to 3mm (r.m.s.) when the flight direction of the impacting electron points at the (small) gap in between two adjacent detectors.

## 3 Technical description of the HES detectors

### 3.1 Fabrication

The quality of a silicon detector depends crucially on the base material used, the fabrication process employed and its geometry. Special care is needed on all three items since the requirements for detector operation exceed those for commercial electronics.

Detector-grade base material is solely produced by the float-zone process. In this a polycrystalline silicon rod is mounted upright into a vacuum chamber. Using a thin radio frequency coil a slice of material is melted. The molten zone is slowly moved upwards allowing a single crystal to grow from bottom to top. The melt is suspended between the two solid half rods and is not in contact with any material but the solid silicon thus avoiding contaminations from the container. The crystal orientation is defined by a seed crystal at the lower end of the polycrystalline rod. If the temperature and the pulling-rate are chosen properly the silicon crystallizes free of dislocations and stacking faults.

The solubility of most of the impurities is larger in the silicon melt than in the solid. They accumulate in the melt and are pulled upwards to the end of the rod. The ratio of the solubilities, solid to liquid, is called the segregation coefficient. For the transition metals, which are feared as life-time killers, segregation coefficients range between  $10^{-4}$  and  $10^{-6}$ . Repeated zoning will therefore purify the silicon. To a lesser extent this is also true for the group-III and group-V elements, which act as dopants in silicon. Phosphorus and arsenic have the highest segregation coefficients (0.35, 0.3) of the group-V dopants and usually constitute the *n*-type doping, especially of the high-resistivity crystals.

Dopant compensation by boron becomes appreciable as the target resistivities enter the range  $> 5 \text{ k}\Omega\text{cm}$ . The boron content is not lowered by segregation, and its concentration profile across the rod diameter is different from the phosphorus profile due to phosphorus evaporation from the melt. The compensation ratio, acceptors to donors, has been reported by Wacker Chemitronic to reach 50% in  $5 \text{ k}\Omega\text{cm}$  material. The resistivity variations across the rod diameter are about  $\pm 10\%$  in three-inch diameter wafers, [Am 84].

The most abundant electrically inactive impurities in float zone silicon are oxygen, which dissolves better in the solid than in the melt, and carbon with typical concentrations of  $10^{16} \text{ cm}^{-3}$ .

The single-crystal silicon rod is cut into wafers of  $300 \mu\text{m}$  or  $400 \mu\text{m}$  thickness on which the detectors are fabricated by a planar process, [Kc 84]. The polished wafers are oxidized at  $\sim 1000^\circ\text{C}$  in steam and in dry oxygen at the end of the treatment. Layers with thicknesses of  $0.5 \mu\text{m}$  to  $1.0 \mu\text{m}$  grow within one to three hours. At the  $\text{SiO}_2/\text{Si}$ -interface a thin ( $\sim 2 \text{ nm}$ )  $\text{SiO}_2$  layer builds up containing excess silicon and immobile positive charges. Their surface density depends on the growth conditions and the crystal orientation. It is lowest for (100)-wafers, and the concentrations reached reproducibly are  $\sim 10^{11} \text{ cm}^{-2}$ , [Pl 80].

Using photolithographic techniques windows for the fabrication of the rectifying junction and the back contact are opened in the oxide. The junction can be produced by in-diffusion of the dopant from a gas atmosphere or by ion implantation. The latter has the advantage of avoiding further high-temperature ( $T > 1000^\circ\text{C}$ ) processing steps in which contaminants could diffuse into the wafer.

Usually, the  $p^+$  layer is formed by implanting boron with an energy of  $15 \text{ keV}$  up to a surface density of  $10^{15} \text{ cm}^{-2}$ . On average, the boron ions are stopped  $50 \text{ nm}$  behind the crystal surface; the r.m.s.-spread is  $25 \text{ nm}$ , [Gi 80]. The peak concentration amounts to  $2 \cdot 10^{20} \text{ cm}^{-3}$ , and the total depth of the boron layer is less than  $0.3 \mu\text{m}$ .

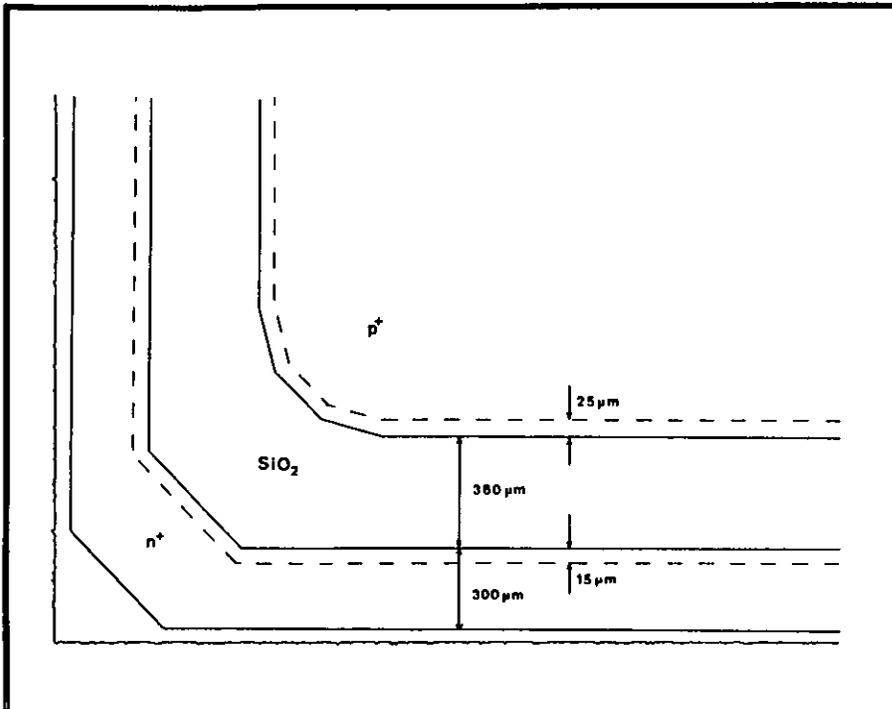


Figure 19: Mask layout of a corner of a Siemens diode showing the  $p^+$  area surrounded by the  $n^+$  fieldstop ring and the insulating  $\text{SiO}_2$  in between. The dashed lines show the extension of the  $\text{SiO}_2$  beneath the aluminium metallization. A cut-away view is shown in figure 23.

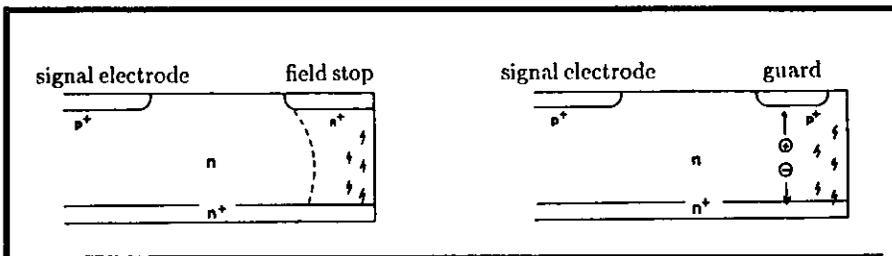


Figure 20: Two ways to avoid collecting unwanted reverse current from the detector edges: An  $n^+$  field stop ring prevents the field zone from extending to the edge. The fieldzone boundary is indicated by the dashed line. A  $p^+$  guard collects electron-hole pairs produced in the edge region. The aluminium metallization and the insulating oxide are not shown, cf figure 23.

#### Charge carrier concentration in a parabolic band:

All energies are measured in multiples of  $kT$ , and the reduced quantities are denoted by  $\xi$ , eg  $\xi_C = E_C/kT$ . The density of states available to the electrons is  $N_C$ ; for the holes it is  $N_V$ . The concentration of electrons and holes are given by the Fermi integrals:

$$n = N_C F_{1/2}(\xi_F - \xi_C) \quad p = N_V F_{1/2}(\xi_V - \xi_F)$$

The Fermi integral of order  $s$  is defined as:

$$F_s(\xi) = \frac{1}{\Gamma(s+1)} \int_0^\infty \frac{z^s}{1 + \exp(z - \xi)} dz$$

$$\approx \exp(\xi) \quad \text{for } \xi \ll -1$$

The inverse function is denoted by  $F_s^{-1}$  and is in the non-degenerate case approximated by:

$$F_s^{-1}(y) \approx \log(y) \quad \text{for } 0 < y \ll 1$$

The  $n^+$  layers are made using phosphorus at higher energies (30keV to 50keV) and doses ( $10^{16} \text{cm}^{-2}$ ). A dose of  $10^{15} \text{cm}^{-2}$  leaves the silicon amorphous. Any subsequently implanted phosphorus is stopped in the amorphous layer. Depth and thickness of the  $n^+$  implant are comparable to the boron values.

After the implantation crystallinity has to be restored. Annealing at  $600^\circ\text{C}$  for half an hour gives best results in terms of maximum electric activation of the dopants. This temperature is still low enough to avoid diffusion of the implanted dopants.

Now the wafer is metallized, and the wanted patterns are again produced by photolithography. Finally, the detector chips are cut from the wafer using a diamond saw.

The HES detectors will be fabricated on  $400 \mu\text{m}$  wafers with a resistivity of  $\sim 5 \text{k}\Omega\text{cm}$ . The metallization mask layout of a detector made by Siemens is shown in figure 19. Rounding off the edges of the  $p^+$  layer avoids high electric fields there, that could cause early breakdown of the insulation. The junction is surrounded by an  $n^+$  ring that is electrically connected to the backside  $n^+$  contact. This ring acts as a field stop. It prevents the depletion region from extending out to the very edges of the chip where the crystal is damaged by microscopic cracks from the sawing. At those cracks electronic states are introduced into the silicon bandgap enhancing greatly the thermal electron-hole pair creation rate. An electric field extending into this zone would collect these, and the dark current of the detector would be large. The cut-away view in figure 20 shows another possibility to keep this current away from the sensitive readout electronics. The  $p^+$  guard collects the current originating at the edges. But it needs an extra pin, which rises costs.

The field stop ring as well as the guard ring constitute inactive detector area. The width of the damaged edge region dictates the width of either type of protection ring. Cutting the detectors with a diamond saw allows for widths of  $300 \mu\text{m}$ . Laser-cutting necessitates widths of  $700 \mu\text{m}$  and is not usually employed. In forming arrays of such detectors one can place them with a precision of  $100 \mu\text{m}$ . The average fraction of inactive area is then  $\sim 6\%$ .

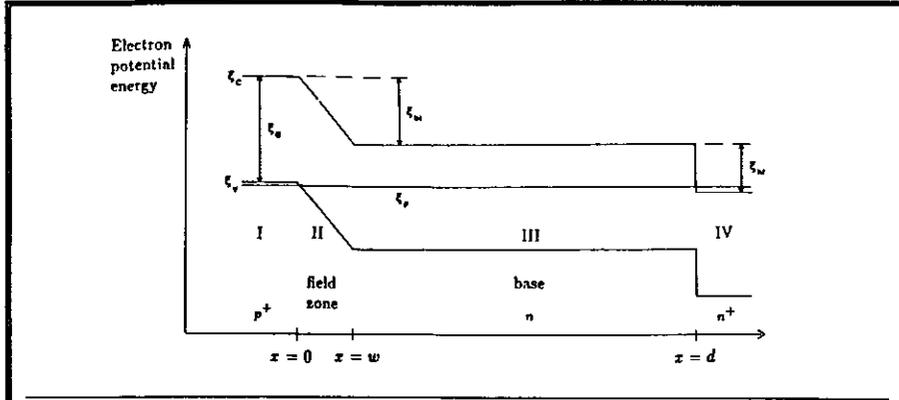


Figure 21: Energy band diagram of a  $p^+n^+$  diode in thermal equilibrium. The Fermi level ( $\xi_F$ ) is constant over all four regions. The conduction and the valence band are bent due to the different dopings, which gives rise to changes in the electrostatic potential energy.  $\xi_b$  and  $\xi_m$  are the (by  $kT$ ) reduced variations of electrostatic potential energy across the  $p^+n$  junction and the high-low  $n^+n$  junction, respectively.

## 3.2 Fields and potentials

Static electric fields and potentials are computed from Maxwell's equation for electrostatics:  $\text{div } E = \rho/\epsilon$ . To solve it, the dopant concentration, the details of the geometry, and the boundary conditions have to be specified. Using a computer it can then be solved numerically. But such a detailed calculation is not necessary here. Two simplifications allow for an analytical solution.

- **One dimension only:** The problem is restricted to one dimension, ignoring possible variations of the doping across the diode area and ignoring edge effects.
- **Depletion approximation:** It is assumed that there are clearly distinct regions in the diode. Field-free regions below the detector surfaces (with no net free charge) enclose the space charge region which is depleted of all mobile carriers (electrons and holes), leaving behind the ionized dopant atoms.

### 3.2.1 Equilibrium case

The driving force for a flow of particles is a spatial variation of the electro-chemical potential or Fermi level. The particle current density is proportional to the gradient of the Fermi potential. In thermal equilibrium there is no net flow of charge carriers in any part of the diode, hence,  $\xi_F(x) \equiv \text{constant}$ .

Figure 21 shows schematically the energy band diagram of a  $p^+n^+$  diode in thermal equilibrium. There are four distinct regions. Region I is heavily  $p$ -doped ( $p^+$ ); region IV is  $n^+$ -doped. They are the  $p^+$  and  $n^+$  layers formed by ion implantation. Their thicknesses are less than  $0.3\mu\text{m}$ . The regions II and III are weakly  $n$ -doped and account for almost all of the diode

thickness. The  $x$  coordinate has its zero at the metallurgical  $p^+n$  junction. At  $x = d$  there is a high-low ( $n^+n$ ) junction.

In the four regions the equilibrium concentrations of electrons and holes differ by many orders of magnitude: In the  $p^+$  layer there are  $10^{19}$  holes/ $\text{cm}^3$  and virtually no electrons:  $\mathcal{O}(1\text{cm}^{-3})$ . The situation is reversed in the  $n^+$  layer. In the field zone (II) there are neither electrons nor holes, and in the base (III) there are  $\sim 10^{12}$  electrons/ $\text{cm}^3$  (set by the  $n$ -type doping) and, at room temperature, about  $10^8$  holes/ $\text{cm}^3$ .

These concentration gradients cause diffusion currents to flow across each interface. Due to the built-in electric fields, drift currents flow in the opposite directions. In thermal equilibrium, the two types of currents cancel each other exactly.

The energy band diagram shows the electrostatic energy of an electron: Under the action of an electric field it moves 'downwards' while a hole moves 'upwards'. In both cases the electrostatic energy is lowered. The electro-chemical potential is the sum of the local electrostatic potential and a purely concentration-dependent chemical potential. In the regions I, III, and IV there is no electric field and no concentration gradient. There, the majority carrier concentrations equal the dopant concentrations ( $N_A$  in I,  $N_D$  in III, and  $N_D^+$  in IV). At the interfaces I-II and II-III there exist large gradients of the majority carrier concentrations, and the electric fields adjust such that the electro-chemical potential remains constant—i.e. no net current flows. The interface III-IV is not exactly abrupt as shown, but it is very thin. The field region associated with it has a thickness comparable to that of the  $n^+$  layer (IV). The equilibrium amount of change of the electrostatic potential across the two field regions (I-III and III-IV) can be read from figure 21. Using the relations of the previous box one can express the position of the Fermi level in each region by the majority-carrier concentration and the electrostatic potentials:

$$p^+ \text{ region (I):} \quad \xi_F = \xi_V - F_{1/2}^{-1} \left( \frac{N_A}{N_V} \right)$$

$$n \text{ region (III):} \quad \xi_F = \xi_C - \xi_b + F_{1/2}^{-1} \left( \frac{N_D}{N_C} \right)$$

$$n^+ \text{ region (IV):} \quad \xi_F = \xi_C - \xi_b - \xi_m + F_{1/2}^{-1} \left( \frac{N_D^+}{N_C} \right)$$

Combining the first two equations allows to calculate  $\xi_b$ ; the second and third give  $\xi_m$ .

$$\xi_b = \xi_C - \xi_V + F_{1/2}^{-1} \left( \frac{N_A}{N_V} \right) + F_{1/2}^{-1} \left( \frac{N_D}{N_C} \right)$$

$$\xi_m = F_{1/2}^{-1} \left( \frac{N_D^+}{N_C} \right) - F_{1/2}^{-1} \left( \frac{N_D}{N_C} \right)$$

For the non-degenerate semiconductor the inverse Fermi function can be set to  $\log(y)$ . The free-energy of the bandgap has the standard value of  $1.12\text{eV}$  at  $300\text{K}$ , and we obtain:

$$\xi_b = \xi_C - \xi_V + \log \left( \frac{N_D N_A}{N_C N_V} \right)$$

$$= \xi_i + \log \left( \frac{N_D N_A}{n_i^2} \right)$$

$$\xi_m = \log \left( \frac{N_D^+}{N_D} \right)$$

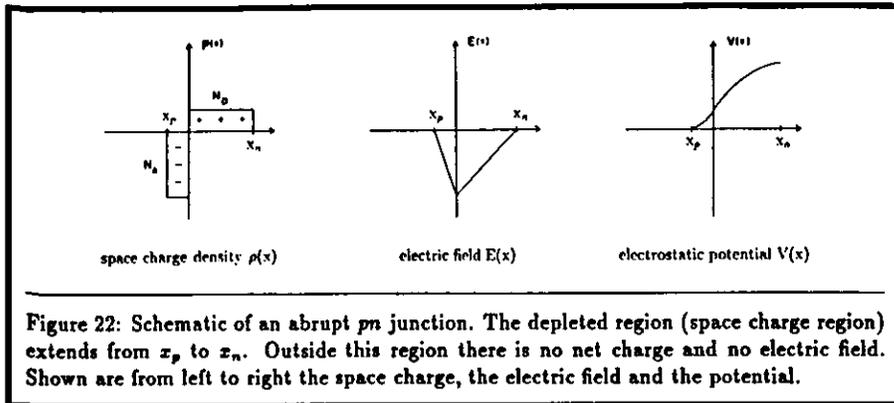


Figure 22: Schematic of an abrupt pn junction. The depleted region (space charge region) extends from  $x_p$  to  $x_n$ . Outside this region there is no net charge and no electric field. Shown are from left to right the space charge, the electric field and the potential.

One commonly calls  $V_{bi} = (kT/e) \cdot \xi_{bi}$  the built-in voltage of a diode. The built-in voltage of the high-low junction is  $V_{bi} = (kT/e)\xi_{bi}$ . Heavy-doping effects like bandgap-narrowing and degenerate Fermi statistics in the  $p^+$  and the  $n^+$  region reduce these values by about  $2kT/e$  to  $4kT/e$ . Values for  $N_A = 10^{19} \text{ cm}^{-3}$ ,  $N_D = 10^{13} \text{ cm}^{-3}$ , and  $N_D^+ = 10^{19} \text{ cm}^{-3}$  are  $V_{bi} = 0.56 \text{ V}$  and  $V_{bi} = 0.23 \text{ V}$ , at room temperature.

### 3.2.2 Backward bias

Figure 22 shows the charge, field and potential distribution for a pn diode. The space charge region is assumed to be free of mobile charge carriers and extends from  $x_p$  to  $x_n$ . Outside this region there is no electric field. This is possible only if the amount of negative charge on the p side matches the positive charge on the n side. This quasineutrality condition is expressed as

$$x_p \cdot N_A = x_n \cdot N_D$$

where  $N_A$  and  $N_D$  denote the acceptor and donor concentrations, which are assumed to be constant. In our case  $N_A$  is larger than  $N_D$  by more than six orders of magnitude making  $x_p$  negligibly small compared to  $x_n$ . Thus, the total width of the depletion zone equals  $x_n$ . The electric field and the potential are obtained from integrating Poisson's equation. The boundary conditions are: Zero field at the ends of the depletion zone and continuity at the metallurgical junction at  $x = 0$ . The maximum electric field then occurs at  $x = 0$ .

The depletion zone grows with the total electrostatic potential applied across the junction. Once it reaches the backside contact it cannot grow further. Any additional (excess) voltage applied gives rise to an additional electric field  $E_x = V_x/d$ , with  $d$  denoting the diode thickness. For a constant dopant density  $N_D$  in the base the results are:

$$-E(x) = \begin{cases} \frac{2V_w}{w} \left(1 - \frac{x}{w}\right) & \text{partially depleted } 0 < x < w < d \\ \frac{2V_d}{d} \left(1 - \frac{x}{d}\right) + \frac{V_x}{d} & \text{fully depleted } 0 < x < d \end{cases}$$

$$V(x) = \begin{cases} V_w \left[1 - \left(1 - \frac{x}{w}\right)^2\right] & \text{partially depleted } 0 < x < w < d \\ V_d \left[1 - \left(1 - \frac{x}{d}\right)^2\right] + V_x \frac{x}{d} & \text{fully depleted } 0 < x < d \end{cases}$$

In the above formulas  $V_w$  denotes the total electrostatic potential developed across the junction for a depletion zone of width  $w$ :

$$V_w = \frac{eN_D}{2\epsilon} w^2$$

The depletion potential  $V_d$  is obtained by substituting  $d$  for  $w$  in the above relation. The total electrostatic potential drop ( $V_t$ ), i.e. the amount of band-bending, equals  $V_w$  in the case of partial depletion. In the case of full depletion (with overbias),  $V_t$  equals the sum of the depletion voltage ( $V_d$ ) and the excess voltage ( $V_x$ ). Note that the voltage applied to the diode contacts is smaller than  $V_t$  by the built-in voltage  $V_{bi}$ .

### 3.2.3 Detector capacitance

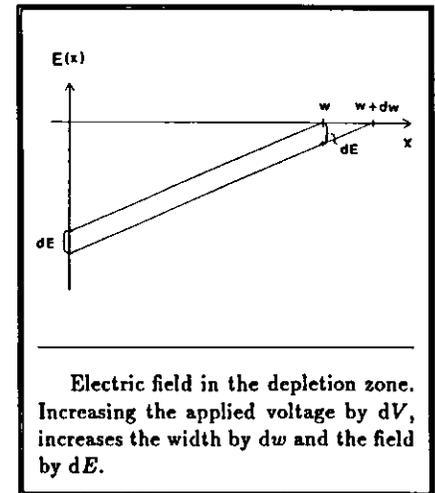
If a signal much smaller than  $V_t$  is applied to the terminals of the diode it will behave like a capacitance paralleled by a small conductance. The latter arises from imperfect insulation. If the reverse voltage is increased by a small step the negative charge on the  $p^+$  side, and the positive charge on the  $n$  side will increase by the same amount. But the diode is backward biased, so no charge carriers can cross the depleted zone. Only a displacement current like in a capacitor is observed. Observe the side figure to calculate the capacitance. The additional charge  $dQ$  produces an increased electric field, displacing the  $E(x)$  curve by a fixed amount  $dE$ . Now the capacitance follows:

$$\begin{aligned} dQ &= eN_D A dw \\ dV &= -w \cdot dE = w \frac{eN_D}{\epsilon} dw \\ C &:= \frac{dQ}{dV} = \epsilon \frac{A}{w} \end{aligned}$$

$A$  is the cross-section area of the diode, and  $\epsilon$  is the dielectric constant of silicon.

One recognizes that the small-signal capacitance of the backward-biased diode is the same as that of a plate capacitor with a plate separation  $w$ , though the field distribution is very different from that of a plate capacitor. The result is valid not only for constant  $N_D$  but for any  $N_D(w)$ . This is true because in the above derivation the increment in charge  $dQ$  as well as the increment in voltage  $dV$  are both proportional to  $N_D(w)$ .

This capacitance shall be called depletion capacitance in the following. It can be used to measure the depletion width regardless of the base doping profile.



Electric field in the depletion zone. Increasing the applied voltage by  $dV$ , increases the width by  $dw$  and the field by  $dE$ .

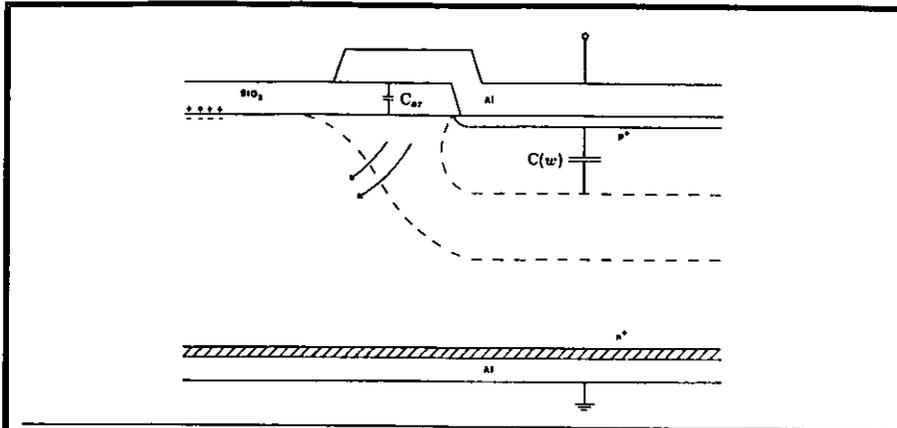


Figure 23: Cut-away view of the edge region of a Siemens diode; see fig. 19 for a top view. The overlap area of the aluminium metallization over the SiO<sub>2</sub> produces a plane capacitor with the e<sup>-</sup> accumulation layer at the Si/SiO<sub>2</sub> interface as the second electrode. When the applied voltage is large enough to offset the surface field of the fixed oxide charge, the accumulation layer is driven away, and the MOS-capacitance vanishes. The dashed lines indicate the extension of the depletion zone at voltages below (upper line) and above (lower line) the flat-band voltage.

Manipulating the above result for the capacitance we can obtain more information from a capacitance-voltage characteristic.

$$\begin{aligned} \frac{d A^2}{dV C^2} &= \frac{2w dw}{\epsilon dV} \\ &= \frac{2}{\epsilon e N_D(w)} \end{aligned}$$

The slope of the  $1/C^2$  vs  $V$  curve depends besides constants only on the dopant concentration at the n side end of the depletion region. A measurement of the depletion capacitance as a function of the reverse voltage is therefore well suited to determine the doping profile of the diode base.

### 3.2.4 MOS-capacitance

The above obtained relation involving the depletion capacitance is commonly used to measure doping profiles—i.e. establishing how the dopant concentration varies with the depth below the detector surface. However, it must not be overlooked that there may be other contributions to the total capacitance of the device. The most important is a parasitic MOS-capacitance.

The thermally grown thick oxide, that provides the insulation between the p<sup>+</sup> area and the n<sup>+</sup> field stop ring, always contains immobile positive charges close to the Si/SiO<sub>2</sub> interface. On the semiconductor side the resulting electric field is neutralized by a very thin (few nm) accumulation layer of electrons forming a conductive sheet just below the interface, [Lo 90].

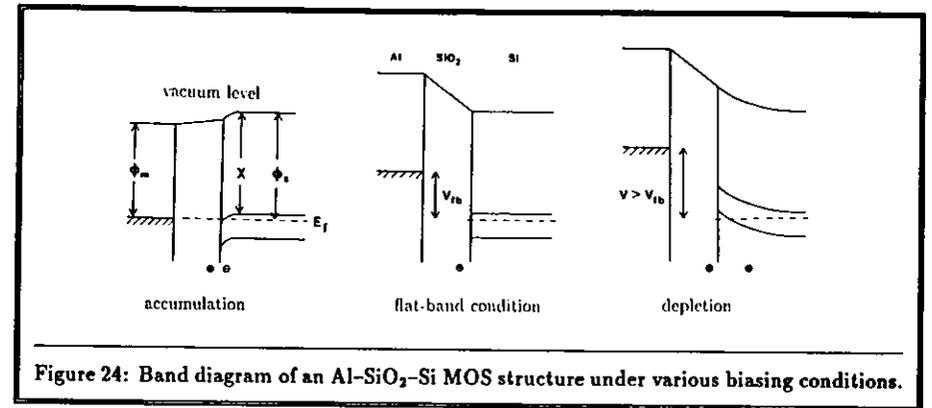


Figure 24: Band diagram of an Al-SiO<sub>2</sub>-Si MOS structure under various biasing conditions.

The overlap of the p<sup>+</sup>-area metallization over the SiO<sub>2</sub>, together with the accumulation layer, creates a ring capacitor.

When the aluminium electrode of the MOS structure in figure 24 is biased negatively with respect to the n-type semiconductor, an additional electric field, counteracting the field from the oxide charge, builds up, and the accumulated negative charge is reduced. The voltage needed to remove the accumulation is called the flat-band voltage,  $V_{fb}$ . At higher bias, the semiconductor below the oxide is driven into depletion. The MOS-capacitance is caused by the presence of an accumulation layer and vanishes at  $V \geq V_{fb}$ .

To arrive at flat-band conditions, the difference in the work functions  $\phi_s - \phi_m$  and the action of the surface charge density  $eN_{ox}$  at the oxide have to be compensated. Let  $t$  be the SiO<sub>2</sub> thickness.  $V_{fb}$  is:

$$\begin{aligned} V_{fb} &= \phi_s - \phi_m + \frac{Q_{ox}}{C_{ox}} \\ &= \phi_s - \phi_m + \frac{eN_{ox}t}{\epsilon_{ox}} \end{aligned}$$

The work function of a semiconductor ( $\phi_s$ ) is given by the difference between the vacuum level and the Fermi level; it therefore depends on the doping. The electron affinity ( $\chi$ ) is the work needed to remove an electron from the bottom of the conduction band, and it does not depend on the doping. The two quantities are related:

$$\phi_s = \chi - kT \cdot F_{1/2}^{-1} \left( \frac{N_D}{N_C} \right)$$

In a capacitance-voltage characteristics the vanishing of the MOS-capacitance at the flat-band voltage shows up as step-like decrease of the device capacitance, cf chapter 4 and 5 for examples.

To estimate this capacitance we treat it like a plate capacitor. Its lateral dimensions are known from the mask layout. For instance, in the Siemens diodes, cf figure 19, the aluminium overlap has a width of 25  $\mu\text{m}$ ; its perimeter is 11.7 cm. From the area and the measured

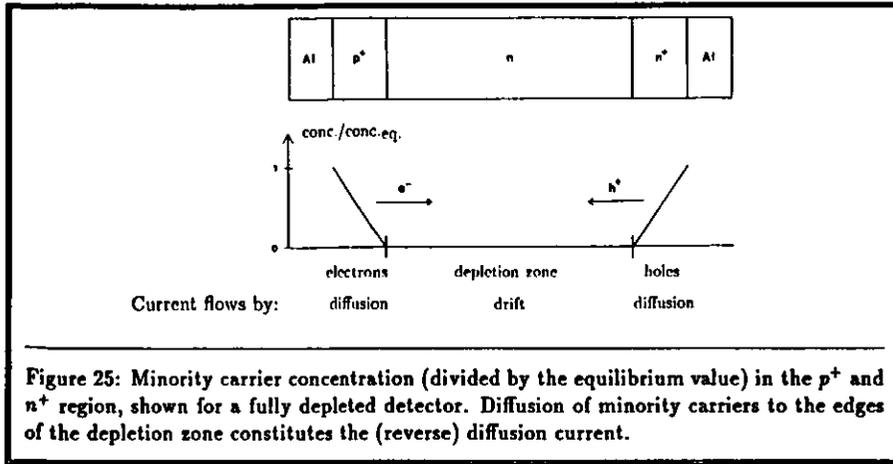


Figure 25: Minority carrier concentration (divided by the equilibrium value) in the  $p^+$  and  $n^+$  region, shown for a fully depleted detector. Diffusion of minority carriers to the edges of the depletion zone constitutes the (reverse) diffusion current.

capacitance steps— $180\text{pF} \pm 10\%$ —the oxide thickness has been calculated to be  $0.50\mu\text{m}$ , with a standard variation of 10% from wafer to wafer. A relative dielectric constant of 3.5 has been used for the  $\text{SiO}_2$ . Diodes manufactured on the same wafer have the same oxide thickness, and the average oxide thickness matches the specifications of the manufacturer.

From the measured flat-band voltages and the oxide thicknesses, a surface charge density of  $N_{sc} = 3.7 \cdot 10^{11}\text{cm}^{-2}$  with variations of 7% (r.m.s.) from wafer to wafer was calculated. This value is typical for (100)-oriented wafers, cf [P180], [Lo90].

### 3.2.5 Reverse currents

As will be shown in section 3.4, the continuous current flowing through a reversely biased detector constitutes an ever-present noise source. An understanding of the various origins of reverse current is vital in designing low-noise detectors. According to their origin, the currents are classified as leakage-, diffusion-, oxide-, and generation current. All of them can be reduced by means of design and technology.

The leakage current, due to insufficient insulation, does not usually present a problem since the  $\text{SiO}_2$  passivation ring between the  $p^+$  area, and the  $n^+$  field stop can be made rather wide ( $300\mu\text{m}$ ) without losing active detector area. Rounding off the edges of the  $p^+$  area as shown in figure 19 avoids locally-high electric fields, which could cause early insulation wearout. Finally, 40nm of  $\alpha$ -silicon are deposited on top of the  $\text{SiO}_2$  passivation to ensure field homogeneity across it. With these measures applied, the leakage currents are about 10nA or less in the HES detectors.

The diffusion current consists of minority carriers diffusing from the end regions into the field zone. In the fully depleted base their concentration is zero. In the heavily doped end regions there is no electric field, and the minority carriers move only by diffusion. Their concentration profile can be calculated from the diffusion equation and the boundary conditions: Equilibrium concentration at the metal contact and zero concentration at the edge of the depletion region. Due to this concentration gradient minority carriers diffuse into the depletion region where they are pulled to the other side by the electric field.

The minority carrier diffusion length in the heavily doped zones is determined by Auger recombination. With about  $10\mu\text{m}$  it is much larger than the zone thickness of  $0.2\mu\text{m}$ . Because of that the minority carrier concentrations can be approximated by the triangular function with constant slope shown in figure 25. The diffusion-current density of electrons coming from the  $p^+$  side is thus:

$$j = eD \frac{d}{dz} n(z) = e \frac{D}{0.2\mu\text{m}} \cdot n_{p0}$$

$D$  is the electron diffusion constant, and  $n_{p0}$  is the equilibrium concentration of electrons in the  $p^+$  region. Because of the strong doping,  $n_{p0}$  is vanishingly small, ( $n_{p0} = n_i^2/N_A \sim 2\text{cm}^{-3}$ ), which is reflected in a very small electron diffusion current. The same reasoning holds for the hole current coming from the  $n^+$  side. We see that by producing highly doped end regions the diffusion currents can be suppressed to negligibly small values of  $O(10\text{pAcm}^{-2})$ .

Oxide currents originate at the Si-SiO<sub>2</sub> interface. Here the periodicity of the crystal is interrupted, and there are many unsatisfied Si-bonds. These introduce electronic states into the bandgap, which enhances thermal pair-creation. As long as an accumulation layer exists below the surface the just created pairs recombine on the spot. When the applied voltage is high enough to deplete the semiconductor below the interface, they add to the reverse current. Correct processing during the growth of the oxide and keeping small the overlap of the  $p^+$  metallization with the oxide helped to reduce this type of current to a few nA at room temperature in the detectors manufactured by Hamamatsu and Siemens.

The by far dominant source of reverse current is the generation current caused by thermal creation of electron-hole pairs in the depleted detector volume. Direct band-to-band pair creation is strongly suppressed by the wide bandgap of silicon. Instead it occurs via intermediate electronic states, which are introduced by all structural defects and by transition-metal impurities. It is therefore essential to use very pure material for the fabrication of detectors and to avoid any contaminations of the bulk during processing. Especially the high-temperature treatments have to be reduced to the absolute minimum; during any such process contaminants readily diffuse into the bulk.

## 3.3 Detector performance

### 3.3.1 Energy deposition

The energy lost by a heavy charged particle on traversing a sheet of matter is well described by Landau's theory. The Landau spectrum of the energy loss of a minimum ionizing particle (mip) after having crossed  $400\mu\text{m}$  of silicon peaks at 104keV and has a full width at half maximum of  $\approx 40\text{keV}$ , [APS92]. The high-energy part of the spectrum is truncated because too energetic knock-on electrons are not stopped in the silicon. In such a case the deposited energy is smaller than the energy lost by the traversing particle. Using the tabulated values for aluminium as an approximation for the stopping-power of silicon, one finds that an electron with 350keV of kinetic energy is just stopped by  $400\mu\text{m}$  of silicon. The Landau spectrum is therefore truncated at energy deposits beyond an equivalent of 3.5mip.

Some part of the deposited energy goes into ionization; the rest is dissipated by acoustic shock waves (phonons). On average 3.6eV of deposited energy are needed for the creation of an electron-hole pair, [APS92].

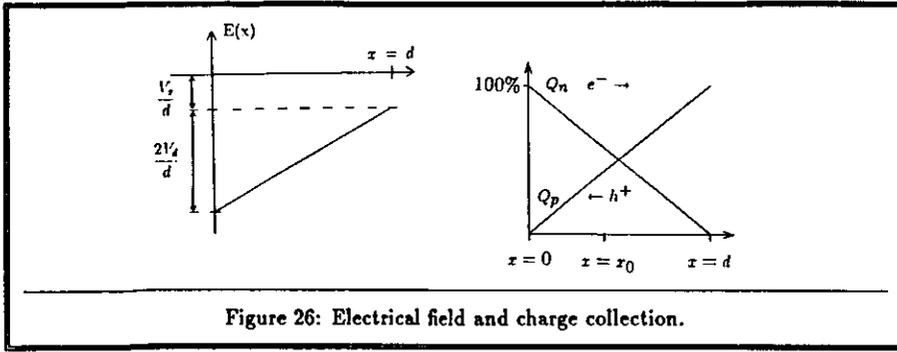


Figure 26: Electrical field and charge collection.

### 3.3.2 Charge collection

The electron-hole pairs are created along the particle track at an average rate of 108 pairs/ $\mu\text{m}$ . Under the action of the electric field they drift to the electrodes thereby inducing a current. The total electric charge induced on the diode capacitance by an electron or hole produced at a position  $x_0$  is:

$$Q_n = e \frac{d - x_0}{d} \quad Q_p = e \frac{x_0}{d}$$

The time it takes an electron to drift from the production point  $x_0$  to a position  $x_n$  is:

$$\begin{aligned} T(x_0, x_n) &= - \int_{x_0}^{x_n} \frac{dx'}{\mu_n E(x')} \\ &= -\tau_n \log \frac{x_n - d - \delta}{x_0 - d - \delta} \end{aligned}$$

where  $\delta := \frac{\epsilon V_s}{e N_D d} = d \frac{V_s}{2V_d}$

$$\tau_n := \frac{\epsilon}{\mu_n e N_D}$$

and  $\mu_n$  denotes the electron mobility. The maximum drift time for electrons is ( $x_0 = 0, x_n = d$ ):

$$T_{max} = \tau_n \log \frac{2V_d + V_s}{V_s}$$

If  $V_s$  is small, the drift times become large because in that case the electric field nearly drops to zero at the  $n^+$  side of the depletion zone. Thus, some overbias is needed to keep short the collection times. The scaling is given by  $\tau_n$  for electrons and by  $\tau_p = \epsilon / (\mu_p e N_D)$  for holes. Due to the smaller hole mobility,  $\tau_p > \tau_n$ . Because  $1/(e\mu N_D) = \rho$  is the specific resistance of the material, the  $\tau$ 's, which equal  $\epsilon\rho$ , are sometimes called the dielectric relaxation time. Typical values for  $N_D = 10^{13} \text{cm}^{-3}$  are 20 ns and 7 ns, respectively.

### 3.3.3 Incomplete charge collection

Some loss of signal might be expected if the electron-hole pair density could become large enough for Auger recombination of the just created pairs; that means recombination with the

assistance of a third charge carrier instead of a deep-level defect. The Auger lifetime  $1/\tau_A = \tau_A n^2$  becomes comparable to the charge collection time at electron-hole pair concentrations of  $2 \cdot 10^{19} \text{cm}^{-3}$ , cf App. A. One can estimate that this is orders of magnitude higher than any realistic concentration. The electrons and holes produced during ionization thermalize in a time of  $O(10 \text{ps})$  creating a tubular track. Even if the lateral extension of this was only due to diffusion during those 10 ps its diameter would be  $0.4 \mu\text{m}$ . A minimum ionizing particle produces on average 108 pairs/ $\mu\text{m}$  resulting in a concentration of  $10^{18} \text{cm}^{-3}$ —much too small for effective Auger recombination. For this reason, silicon detectors do not show saturation effects for any realistic ionization density.

A second source of signal loss is the trapping of mobile carriers by traps present in the detector. If the charge carrier is re-emitted at a time much later than the maximum collection time it has to be considered as lost. The fractional charge loss can be calculated as follows. According to the previous section, the charge induced by a liberated electron is less than the maximum possible if the electron comes to a standstill at  $x_t$ . The resulting deficit is:

$$Q_t = e \frac{d - x_t}{d}$$

Let  $p(x_0, t) dt dx_0$  denote the probability that an electron is produced in the interval  $(x_0, x_0 + dx_0)$  and is trapped in the time interval  $(t, t + dt)$ . Expressing the time  $t$  it takes the electron to drift from  $x_0$  to the trapping point  $x_t$  as  $T(x_0, x_t)$  then yields:

$$\begin{aligned} p(x_0, t) dt dx_0 &= p(x_0, x_t) \frac{dt}{dx_t} dx_t dx_0 \\ &= \frac{1}{d} \frac{1}{\tau_{trap}} \exp(-T(x_0, x_t)/\tau_{trap}) \frac{dx_t}{x_t} dx_0 \end{aligned}$$

Note that the probability density  $p(x_0, x_t)$  is normalized to unity. The total charge lost during the collection is obtained by weighting the charge collection deficit with the probability for its occurrence and integrating over the allowed range of  $x_t$  and  $x_0$ . Let  $Q_0$  be the total charge produced over the whole diode thickness  $d$ .

$$Q_t = Q_0 \int_0^d \int_{x_0}^d \frac{d - x_t}{d} p(x_0, x_t) \frac{dx_t}{x_t} dx_0$$

$$\text{and } x_t = \mu_n E(x_t)$$

This formula correctly takes into account the amount of charge that has already been induced on the diode capacitance by the movement of the electron from the production point to the trapping position. A similar calculation has to be performed for possible hole trapping. The integrations are elementary but lengthy.

At little or no overbias and trapping times  $\tau$  much larger than the relaxation times  $\tau_n, \tau_p$  from the previous section, the fractional charge loss due to trapping is well approximated by

$$\frac{Q_{tn}}{Q_0} = \frac{\tau_n}{\tau} \quad \frac{Q_{tp}}{Q_0} = \frac{\tau_p}{\tau}$$

for electrons and holes, respectively.

### 3.4 Noise

#### 3.4.1 Introduction

In many cases, the noise occurring in an electronic device can be modelled as a random pulse train. It is assumed that all pulses have the same shape  $f(t - t_k)$ , occur at random times  $t_k$ , and have random amplitudes  $a_k$ . The time dependence of the noise signal ( $X$ ) is under these circumstances described by the following sum:

$$X(t) = \sum_k a_k f(t - t_k)$$

If the times  $t_k$  are Poisson-distributed then Carson's theorem holds. It states that the power spectral density of the noise  $X$  is:

$$S_{XX}(\omega) = 2\nu\overline{a^2}|F(\omega)|^2$$

$\overline{a^2}$  mean-square of the pulse amplitudes  
 $\nu$  mean rate of occurrence  
 $F(\omega)$  Fourier transform of  $f(t)$

If the pulses are very short  $F(\omega)$  extends to very high frequencies and can be approximated by  $|F(\omega)| \equiv 1$ , ie white noise. The well-known results for the thermal noise voltage and current of an ohmic resistance  $R$  are:

$$S_{VV}(\omega) = 4kTR$$

$$S_{II}(\omega) = 4kT/R$$

For common shot noise, eg in a vacuum tube, caused by a dc current of magnitude  $I$  there holds:  $\nu = I/e$ , and all amplitudes are equal,  $a = e$ , [Bu 83].

$$S_{II}(\omega) = 2eI$$

#### 3.4.2 Shot noise in pn diodes.

The power spectral density of shot noise due to the generation current of a  $p^+n$  diode is different from the above stated result because the amplitudes are not the same from pulse to pulse. But employing Carson's theorem

$$S_{II}(\omega) = 2\nu\overline{q^2}$$

leads to the correct value, [Zi 75]. The pulse rate is as before  $\nu = I/e$ . The charges induced on the diode capacitance by an electron and a hole produced at  $x$  are:

$$Q_n = e\frac{d-x}{d} \quad Q_p = e\frac{x}{d}$$

The mean-square of the noise charge is then:

$$\begin{aligned} \overline{q^2} &= \overline{Q_n^2} + \overline{Q_p^2} \\ &= e^2 \frac{1}{d} \int_0^d \left(\frac{d-x}{d}\right)^2 + \left(\frac{x}{d}\right)^2 dx \\ &= \frac{2}{3}e^2 \end{aligned}$$

The spectral noise power density follows:

$$S_{II}(\omega) = 2\frac{2}{3}eI$$

Underlying the derivation above is the assumption that the electron and hole emission are uncorrelated. It would be violated if the electron emission followed the hole emission in a time short compared to the charge-collection time. In the latter case of full correlation, one would obtain the common shot noise result  $S_{II} = 2eI$ .

#### 3.4.3 Signal processing

The noise performance of a receiver is best described by an equivalent noise input signal. The response of the integrator-shaper chain of figure 27 to an input current  $I(\omega)$  is given by:

$$U_{out} = I_{in} \cdot Z_f \cdot v(\omega)$$

$$= \frac{I_{in}}{\omega C_f} \frac{\omega\tau}{1 + \omega\tau} v(\omega)$$

Here,  $\tau = RC_f \approx 20\mu s$  is the decay time of the integrator. The shaper is a bipolar shaper with a peaking time of 180ns. A voltage step of 100mV at its input (from the preamp) yields a peak output voltage of 89mV, ie  $v_{step} = 0.89$ .

The output noise voltage due to the shot noise input current is calculated from:

input-current noise-power density:  $S_{II}(\omega) = \frac{4}{3}eI$

output-voltage noise-power density:  $S_{VV}(\omega) = \frac{4}{3}eI |v(\omega) \cdot Z_f(\omega)|^2$

Now the mean-square of the output noise voltage follows as:

$$U_n^2 = \frac{4}{3}eI \int_{-\infty}^{+\infty} |v(\omega) \cdot Z_f(\omega)|^2 \frac{d\omega}{2\pi}$$

$$\approx \frac{I[\mu A]}{C_f^2} 1.69 \cdot 10^{-28} C^2 s^{-1}$$

The frequency dependence of the shaper amplification,  $v(\omega)$ , has been found by performing a numerical Fourier transformation of the shaper output pulse form. The output noise voltage is equivalent to an input noise charge:

$$Q_n^2 = C_f^2 U_n^2 / v_{step}^2$$

$$\approx (0.22 fC)^2 \cdot I[\mu A]$$

The total input-noise-charge squared is the sum of a constant noise term from the amplifier noise and an independent shot noise term. Figure 28 shows a measurement of the mean-square of the input noise charge  $v$  the diode reverse current. The measured slope of  $(0.22 fC)^2 / I[\mu A]$  is in fair agreement with the calculated value.

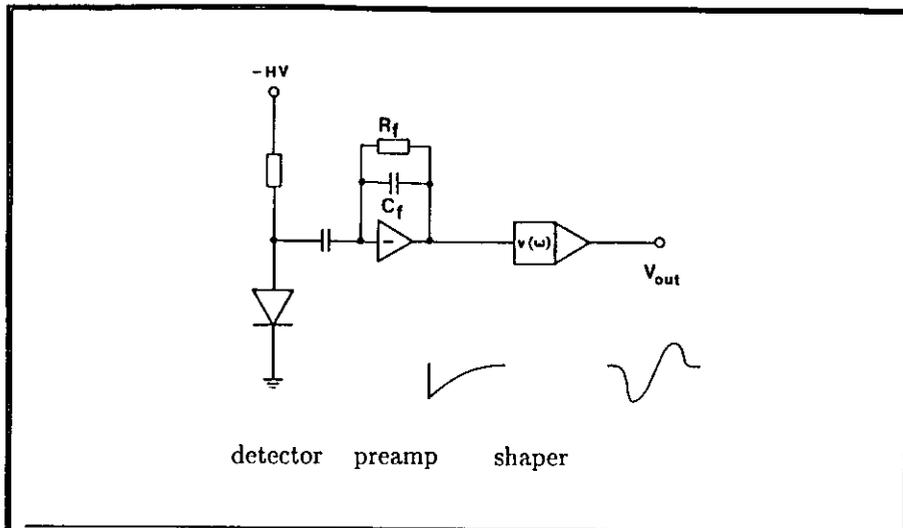


Figure 27: Analog signal processing for HES. The charge liberated by an ionizing particle ( $Q$ ) is fed into the preamplifier via a coupling capacitor. The preamplifier responds with a voltage step  $\Delta V = QC_f$ . Its output voltage decays exponentially with a characteristic time of  $20\mu s$ . The shaper converts the voltage step into a bipolar output signal as sketched. The positive peak voltage is  $V_{out} = v_{step}\Delta V$ .

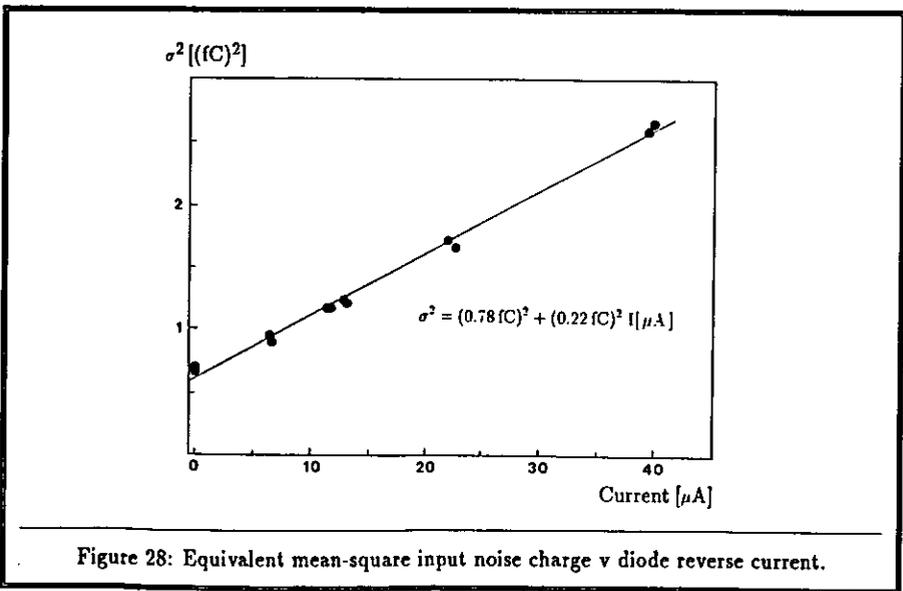


Figure 28: Equivalent mean-square input noise charge v diode reverse current.

## 4 Detector testing

### 4.1 Introduction

The full-sized HES will eventually involve some 50,000 silicon detectors; their costs alone will amount to some US\$3mill. Such quantities of large-area silicon detectors cannot be purchased off the shelf. It was decided, to have the prospective vendors produce 100 detectors, which should be subjected to close scrutiny. The testing at DESY had to have enough decisive power to give a general approval for the design chosen by the manufacturer.

The number of items under survey together with the number of diodes expected to be delivered by the different enterprises made semi-automatic testing a necessity. The test facility designed at DESY had to be flexible enough to allow for very different kinds of measurements and, on the other hand, had to be fast enough to handle a hundred diodes within a week.

During the mass-production and delivery of the detectors some simplified acceptance test, taking into account the unique characteristics of the chosen manufacturer's design, will be performed on every arriving diode. This will, however, be done on some other facility, more apt to handle a large number of detectors.

### 4.2 Items under survey

A general technical description of the detectors and an understanding of their performance have been developed in the previous chapter. Descriptions that are more precise on the mask layout, and other constructional details are not usually provided by the manufacturer. Where possible, a few of the diodes were subjected to an optical inspection, using a microscope, to ensure the correctness of the dimensions as well as to determine the area of the MOS-capacitance.

**Electric parameters:** For all diodes the reverse current, the capacitance, and the series resistance have been measured as a function of the reverse voltage. From the capacitance-voltage characteristic the depletion width, the depletion voltage, and the resistivity of the base material are extracted. The reverse current should consist chiefly of volume-generated current and should not exceed a specified value. An early break-down of the diode insulation is most easily recognized by a steepening increase of current with voltage at or beyond depletion.

In the HES all diodes will be fed the same supply voltage. Hence, the tolerances in the depletion voltage and break-down voltage have to be tight. In the DESY bidding, the specifications concerning the electric parameters were given as follows:

- The common operating voltage for all detectors must be less than 200V.
- At the operating voltage the depletion depths of the detectors must be within  $\pm 5\%$  of the nominal depth stated by the manufacturer.
- The nominal depth must exceed  $280\mu m$ .
- The minimum safe operation voltage must for any detector be larger than 110% of the common operating voltage.
- The reverse current at this voltage must not exceed  $2\mu A$  at  $20^\circ C$ .
- The series resistance of the depleted detectors must not exceed  $20\Omega$ .

**Noise:** The HES analog electronics is adapted to the high detector capacitances and to the requirements from the high bunch-crossing frequency at HERA, dictating a quite short peaking-time of 200ns for the bipolar shaper. With a qualifying detector being connected to the HES preamp, the noise at the output should not be larger than the amplified noise from the FET input transistor and amplified shot noise from the reverse current. Additional noise of comparable magnitude is a cause to reject the detector.

**Performance as detectors:** Detector operation, namely the complete collection of all liberated charge, is tested using a  $\beta$ -source with a high end-point energy,  $^{106}\text{Ru}$ . The relativistic electrons ( $E_{\text{max}} = 3.5\text{MeV}$ ) behave approximately like penetrating minimum ionizing particles. Their detection is a performance test at the highest level of sensitivity.

**Long-term stability:** As a protection against unfavourable environments the detectors need a coating. If it gradually wears out the reverse current can show a steady increase with time. To check this, a representative number of diodes are powered for two weeks with their reverse currents being monitored.

### 4.3 The test facility

The task profile divides in two parts: routine testing and special investigation. Standard measurements that repeat for every detector should be done semi-automatically, which means that the devices have to be inserted manually into the different test fixtures, but the measurements then proceed automatically in a predefined way. Running the tests should require no more than a day's training for non-physicist personnel. On the other hand, the facility should be flexible enough to allow for the introduction of new procedures by an expert. Such enhancements must not interfere with already existing procedures.

#### 4.3.1 General layout

The demands on flexibility and low costs were the guide line in choosing the hardware to be used. An ATARI 1040ST computer together with two 20MB hard disks was regarded as sufficient in computing power and storage capacity, and it offered an exceptionally good price-to-performance ratio. At the other end of the chain, most of the commercially available analog instruments are equipped with an IEEE-488 remote-control feature. The IEEE-488, or equivalently the IEC-625, bus standard allows to interconnect up to 31 devices and offers a data transmission rate in excess of 60kByte/s. The host computer communicates with the instruments via a controller. It acts as a master on the IEEE side issuing all the special command sequences required by the IEEE-488 protocol. The communication between the host and the controller is performed via the MIDI I/O-port. Its transmission rate of 4kByte/s is sufficient for this application.

The need for easy reprogramming and enlargement of the system necessitates the use of a high-level language. When development began, no professional C-compiler for the ATARI could be found, and the high-level language of choice had been FORTRAN. The software package installed on the station is organized in three hierarchical levels. Assembler-written routines providing access to the operating system constitute the bottom level. On an intermediate level repeating tasks are collected in three FORTRAN libraries. On the top level, menu cards are employed to guide the user through the different procedures.

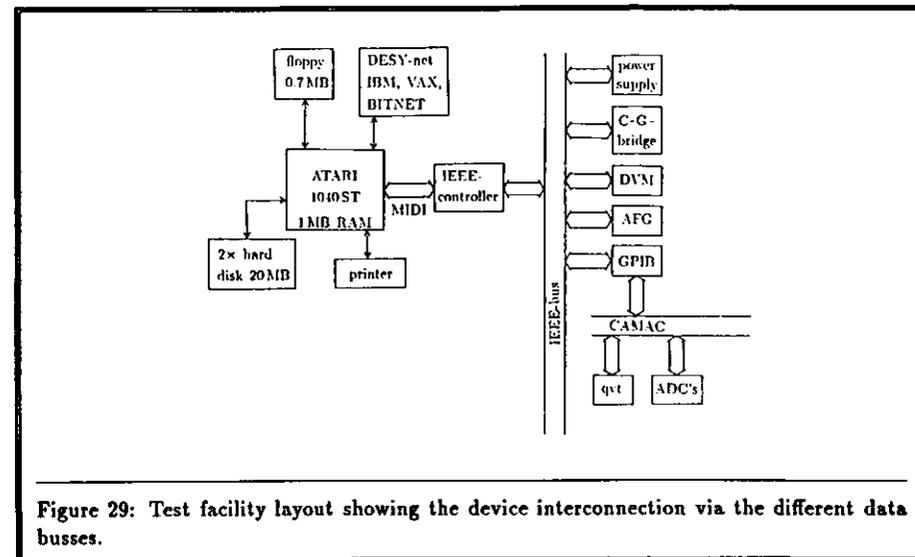


Figure 29: Test facility layout showing the device interconnection via the different data busses.

#### 4.3.2 Software

The software development consisted of linking the ATARI to the IEEE-bus, providing libraries for operating the instruments, and writing the menu-driven control program itself.

A controller bought from GTI<sup>6</sup> serves as a mediator between the host computer and the devices on the IEEE-bus. It supports the full set of IEEE-488 commands and is linked to the ATARI via its MIDI I/O-port. An operating-system enhancement, also delivered by GTI, implements the needed Xon/Xoff protocol on the ATARI side. Some safeguards and time-out features have been added by the author of this work. When the computer is booted, the driver is copied into its memory where it remains until the next cold start.

The host and the controller exchange data and commands in ASCII format. The subroutine to send and receive the ASCII strings has been written in assembler out of speed considerations. It exists in two versions to interface with the high-level languages FORTRAN 77 and C. Since the FORTRAN compiler did not offer access to operating-system functions, like renaming files or executing other programs, this has been made possible by assembler-written FORTRAN-callable SUBROUTINES.

A CAMAC crate has been hooked up to the IEEE-bus via a CAMAC controller equipped with an IEEE-interface. While the other components send and receive data or commands solely as ASCII strings, the older CAMAC sends its data in binary format. This is in conflict with the Xon/Xoff protocol, which reserves the numbers 10 and 13 (ie the bytes 1010, 1101) as data-flow control bytes. The conflict has been resolved by converting the binary data to an ASCII stream prior to transmission via MIDI and back-converting it to binary on the receiving end. This had to be implemented for both directions calling for a programmable controller with a remote-program execute facility.

<sup>6</sup>Berlin, Germany

TRUNK

- <6> CHARACTERISTICS
- <7> LONG-TERM
- <8> SPECTRUM
- <9> IEEE TEST
- <10> GRAPHICS
- <11> DUAL-DIODE CARD

YOUR CHOICE:

<3> COMMENT <4> RETURN

Figure 30: Main menu as shown on start-up. The left-hand numbers give the code that has to be entered for the sub-menu of choice. The bottom line shows the number code for the general commands that are active at this stage.

Three FORTRAN libraries which collect all the repeating tasks have been installed. Two of them are the device-command library and the CAMAC-command library in which tasks like initializing and performing the measurements have been bundled together. The other library comprises the features needed frequently, like reading and storing complete data sets in a standardized way as well as routines supporting the tree-like structure of the main program.

After start-up, the test station introduces itself to the user displaying the available features. On selecting one of them, the user is led into a sub-menu in which he can set the parameters of interest. All parameter settings are preserved throughout a session. On exit they are stored in a resource file which is read in on start-up. All acquired data for which this is sensible are stored in such a manner that they are readable by the included graphics package PLOTFIT<sup>7</sup>. The features appearing in the entrance menu, figure 30, are described in detail in the following paragraphs.

#### 4.3.3 The features

**Characteristics:** Measuring the electric characteristics is done using the setup shown in the figure on the next page. The detector under test (DUT) is biased by a programmable power supply of Rohde & Schwarz that can supply up to 300V in steps of 0.1V. Its output voltage is low-pass filtered by  $R_1$  and  $C_1$ . The large time constant of 0.6s ensures a soft switch-on for the detector, a feature found necessary for some of them. The resistors  $R_2$ ,  $R_3$  are needed to decouple the device under test from ground.

The capacitance-conductance bridge, an HP 4280A of Hewlett Packard, performs to an accuracy of better than one percent in its three ranges 10pF/100 $\mu$ S, 100pF/1mS, 1nF/10mS. It uses a measuring signal of 1MHz and 30mV r.m.s. At low bias the detector capacitance is about 3nF, and the bridge would sense an overflow if  $C_2 \approx 0.8$ nF was not included. The data-taking routine takes into account the action of  $C_2$  and calculates the true capacitance and series resistance of the detector while the bridge itself corrects for the errors introduced by the coaxial cables.

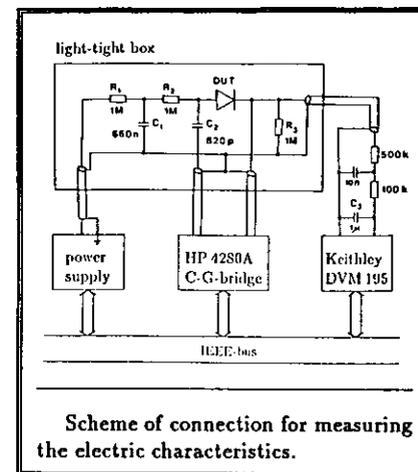
The dark current is determined from the voltage drop over the resistor  $R_3$ . Due to the very high input resistance of the Keithley voltmeter a 50Hz ripple is always present at its inputs,

<sup>7</sup>Wolfgang Koenig, Am Bruehl 1, D-W-6101 Messel, Germany. The program is of public domain.

even when they are left unconnected. The 1 $\mu$ F capacitor  $C_3$  reduces this considerably. To further suppress noise, the voltmeter performs a sixteenfold averaging, asynchronous to the 50Hz mains frequency. All in all, the maximum error in the reverse currents is below 0.1nA. Again, the raw data are manipulated. The finite input resistance of the voltmeter is taken into account, and the true voltage developed across the DUT is calculated. The latter is smaller than the applied voltage by the voltage drop across the three 1M $\Omega$  resistors.

The power supply was not stable enough to be run without an online gauge. On start-up and every few hours thereafter, the power supply was hooked to the voltmeter and a calibration run was started. The required voltage settings for the power supply are read in from a file, and the true voltages, as read by the voltmeter, are stored in a calibration file. The voltages obtained from this procedure were then used in the standard measurement runs for calculating the voltage drop across the DUT.

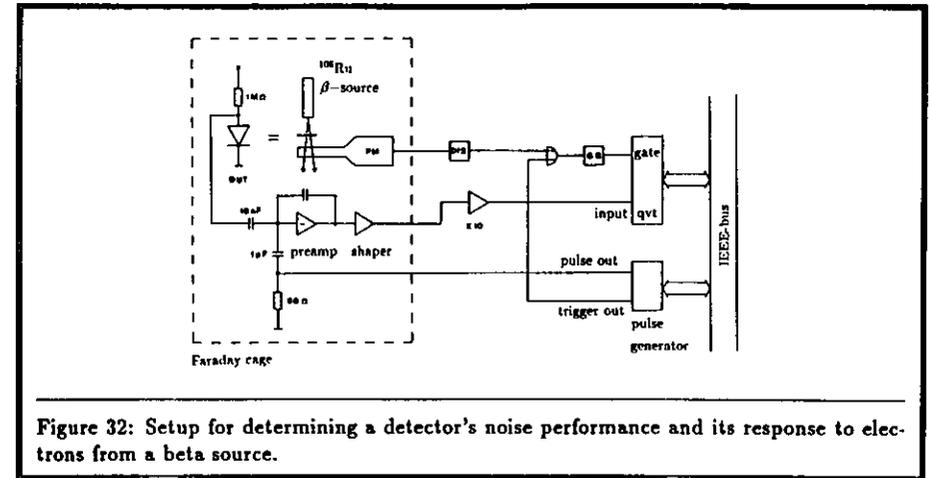
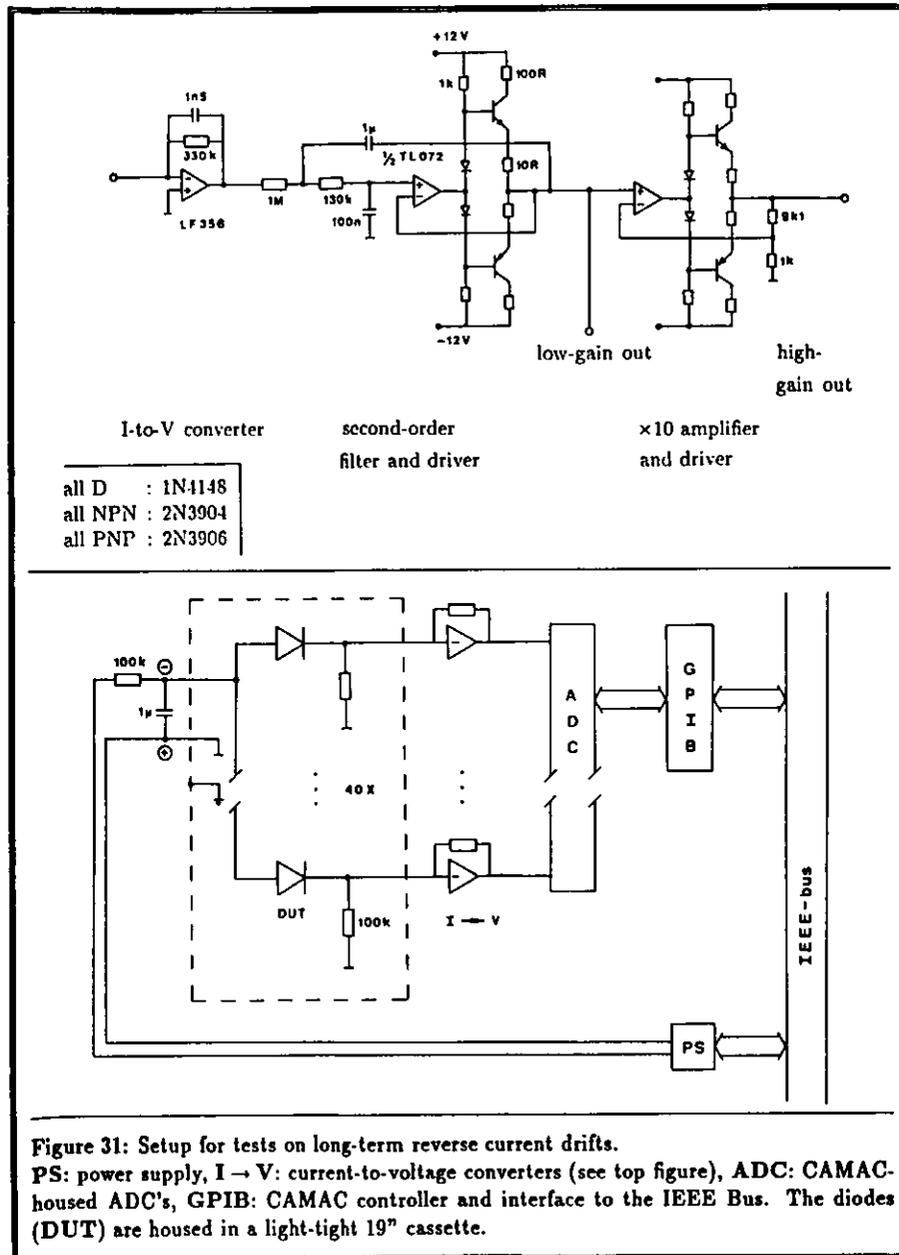
At the end of each measurement the reverse current, capacitance, and resistance of the detector as a function of the true bias are stored on the disk. Immediately afterwards they can be displayed graphically by selecting the graphics option in the main menu.



**Long-term:** The dark current of a diode may change with time. It was found that the reverse current of some diodes increased considerably during the first few days under bias. Because of this a setup to simultaneously monitor the dark currents of up to 40 diodes has been built. Its scheme is shown in figure 31. An RC-filter in the power lines ensures that the voltage on the diodes rises slowly after turn-on. The diodes together with the resistors and the temperature sensors reside in a light-tight box that fits into a standard 19"-rack. The current of each diode is converted to a voltage by its  $I \rightarrow V$  converter. Its output is sent to a low-gain buffer amplifier, which also acts as a second-order low-pass filter with a corner frequency of approximately 1Hz. Further amplification is available by a subsequent amplifier. The two sensitivities are 3V/ $\mu$ A and 30V/ $\mu$ A. Depending on the magnitude of the reverse current either the high-gain or the low-gain output is routed to voltage-sensing ADC's. The signals of the four temperature sensors inside the box are processed to give output voltages of  $V_{out} = (200 + T)$ mV with  $T$  being the temperature in centigrades. The ADC's are read out in programmable time intervals by the ATARI. The necessary gates are provided by a free-running clock ( $\sim 100$ Hz).

The long-term measurement can extend over any period. However, no other work can be done on the station during this period because the ATARI's operating system supports only one task at a time.

**Spectrum:** A light-tight Faraday cage has been built to house the sensitive part of the electronics needed to measure the noise performance of a detector and to determine its response to  $\beta$ -rays. As shown in figure 32 it contains the diode under test (DUT) a preamplifier, a shaper, and a scintillator paddle which is read out by a photomultiplier. The diode is mounted as close as possible to the scintillator paddle, and a  $\beta$ -source with a high endpoint energy is placed on



top of it ( $^{106}\text{Ru}$ ,  $E_\beta \leq 3.5\text{MeV}$ ). Electrons which deposit an energy beyond threshold in the scintillator, 1 MeV, trigger the multi-channel analyser, a qvt of LeCroy. These electrons nearly behave like minimum-ionizing particles when traversing the only  $300\mu\text{m}$ -thick diode. The  $\beta$ 's are collimated by a steel tube to a 3mm-wide beam with an opening angle of 300mrad, which limits the energy straggling due to different incident angles to 1%.

With no  $\beta$ -source present the apparatus can be calibrated using an external pulser connected to the test input of the preamplifier. From the width of the pulser peaks the noise is inferred.

Graphics: On selecting this option the user is led to a menu from which he can start the graphics package, PLOTFIT. It is written in GFA-BASIC and provides flexible graphics, extensive fitting facilities to predefined functions with an arbitrary number of fit parameters (polynomials, exponentials, power-laws, and Gaussians) as well as Fourier transformation. All but the spectrum data are stored in a format readable by PLOTFIT. Spectrum data are compressed, which reduces the amount of disk space needed by a factor of three. To be displayed graphically, they have to be 'expanded'—a feature also provided in this menu.

IEEE test: If a new device has to be introduced into the system this is a useful tool. All commands can be given here manually. The response (if any) will be displayed. All commands can thus be tested prior to writing any routine that will control the new device. Special commands allow the initialization of the known devices.

Dual-diode card: In its final version the HES building blocks will be modules consisting of two diodes and two preamplifiers mounted together on one ceramic board. These dual-diode cards can be tested using the previously described setups, with minor modifications. And this is the menu to do it. It is divided into four sub-menus:

The characteristics measurement in this case comprises only the combined reverse-current characteristic because both diodes receive the bias voltage from a common input pin. A capacitance measurement is impossible because the diodes are connected to their preamplifiers which provide a virtual ground at the input.

Each preamplifier is connected to a shaper whose output is, after additional amplification, routed to the multi-channel analyser (qvt). Using a pulser, 0fC, 5fC, and 10fC are injected

into the known test-input capacitance, and a preset number of pulses are recorded on the qvt. It is read out by the ATARI, and a peak-finding routine determines the peak positions from which the gain is calculated. From the gain and the widths of the peaks the noise is calculated in terms of an equivalent input noise charge.

The detector function of both channels is verified using a ruthenium source as described above. Again, the peak-finding routine determines the peak positions, and the pedestal-corrected value of the ruthenium peak is kept in memory.

The fourth sub-menu controls which data are to be filed away. The reverse-current characteristic, the gain, the noise, and the ruthenium-peak position of both channels are always stored on disk. It is also decided here whether or not to also write some or all of the spectra that have been recorded to the hard disk.

#### 4.4 Test results

Prototype diodes of three enterprises namely Ansaldo of Italy 36 pieces, Siemens of Germany, 113 pieces and Hamamatsu of Japan, 300 pieces have been available. All of them have been tested extensively, and the results are summarized below.

##### 4.4.1 Ansaldo

The detector design of Ansaldo differs crucially in two respects from what has been described in chapter 3. They employ neither an  $n^+$ -field stop nor a  $p^+$ -guard. Instead a 1mm wide inactive zone surrounds the  $p^+$ -area. No MOS-capacitance was detected, and no insulating  $\text{SiO}_2$  layer was seen under the microscope. The passivation needed to avoid surface currents seems to be missing. There is nothing but a polyimide layer covering the bare silicon to provide some protection.

As a consequence the reverse currents are not only high from the start, but they also increase considerably with time. Most of the detectors recovered after the bias was switched off, but some would not, becoming useless after a few days under power because their reverse currents by then exceeded  $50\mu\text{A}$ .

Noise measurements also yielded discouraging results. The diodes had been stored in the dark with no bias applied for at least 12 hours. At switch-on, the currents ranged from  $0.5\mu\text{A}$  to  $3\mu\text{A}$ , and large excess noise was found in about half of them.

Diodes whose reverse currents exceeded  $4\mu\text{A}$  after two days or showed excess noise of more than 1fC (r.m.s.) are indicated in black in chart 34. Obviously, no correlation exists between the two causes for rejection. Only a quarter of the diodes fulfilled the specifications. The manufacturer did not undertake the necessary changes in design and processing and eventually withdrew from the project.

##### 4.4.2 Hamamatsu

Hamamatsu delivered 300 diodes with a nominal thickness of  $300\mu\text{m}$ . They have been produced from two lots of wafers delivered by two different suppliers. In figure 35 the resistivity of the base material is plotted against the diode serial number. One recognizes the two lots having different resistivities.

According to a communication from Hamamatsu a gettering process that involved grinding off some material from the backside had been employed to reduce the concentration of electri-

cally active deep traps in the bulk volume. The procedure reduced the wafer thickness to less than  $280\mu\text{m}$  and caused a considerable spread of it in lot 1.

The reverse current and capacitance characteristics of a good Hamamatsu diode are shown in figure 36. With less than  $10\text{nA}$  at full depletion, the reverse current is extremely small. At the flat-band voltage a current step of  $\sim 0.5\text{nA}$  due to the oxide current is observed. The capacitance shows a similar step-like behaviour because the MOS-capacitance vanishes at flat-band conditions.

The charge collected per minimum ionizing particle was proportional to the diode thickness as expected—about 3.2fC (most probable value) for a  $280\mu\text{m}$  diode.

The HES prototype that was tested at CERN used 184 of these diodes. In a period of one and a half years they had been under power for about 20 weeks at operating temperatures of  $30^\circ\text{C}$  to  $40^\circ\text{C}$ . No degradation in their performance concerning reverse current, charge collection and noise could be detected. None of the 184 diodes showed excessive noise.

The diodes, though of exceptionally high quality, fell short of the specifications concerning the depletion widths and their allowed variation. Subsequently Hamamatsu changed or omitted the gettering process allowing for somewhat larger dark currents. 20 diodes, delivered later, had dark currents of  $50\text{nA}$  to  $100\text{nA}$ . But the variation of the thickness had reduced to a mere  $2\mu\text{m}$  r.m.s.

##### 4.4.3 Siemens

One hundred diodes made from  $380\mu\text{m}$  wafers have been delivered by Siemens. The thickness is reduced by polishing to  $350\mu\text{m}$  with a maximum variation of  $5\mu\text{m}$ . The electric parameters are well within the specifications, see table on page 59.

But the characteristics plot of a typical Siemens diode (figure 37) reveals a problem with the backside contact of the diode. For all diodes the volume-generated current is acceptably small. However, the dark current increases strongly when the diode is driven into full depletion.

When 40 diodes were kept under bias ( $120\text{V}$ ) for 14 days, it was found that the dark current of eight of them changed erratically with time increasing or decreasing by as much as a factor of two. In one diode the current had grown by 500% in the first 100 hours. In a meeting with representatives from the enterprise these effects were attributed to an improper annealing procedure following the backside ( $n^+$ ) ion implantation. Siemens proved they had solved the problem when they sent in a new sample of diodes (15 pieces). The diodes had reverse currents between  $100\text{nA}$  and  $500\text{nA}$  and showed no strong current rise at depletion. Also none of the diodes showed drifts in the reverse current over a period of one week.

The charge collected for the new as well as the old diodes was 4.2fC per mip as expected for a depletion depth of  $355\mu\text{m}$ . 25 diodes from the first batch and all 15 of the second have been subjected to a noise test; none showed excessive noise.

#### 4.5 Conclusion

Testing the basic parameters of a medium-sized sample of detectors made it possible to point out weaknesses in the design and to give hints to the manufacturers about the nature of the problem. Two enterprises capable of producing  $3 \times 3\text{cm}^2$  silicon detectors at acceptable costs and with the required quality have been found.

resistivity	(5...7) kΩcm
depletion depth	(400 ± 4) μm
depletion voltage	(110 ± 10) V
reverse current @ 120 V, 20°C	(1...3) μA

Electric parameters of the Ansaldo diodes.

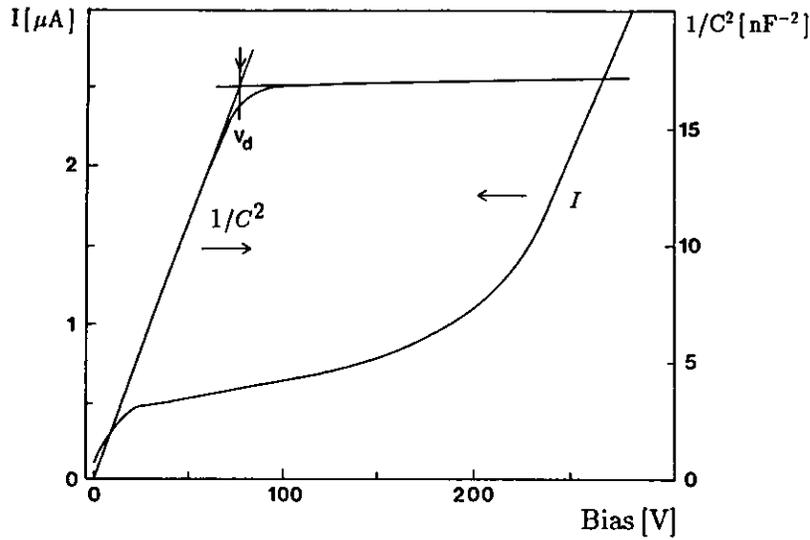


Figure 33: Reverse current and capacitance characteristic of a typical Ansaldo diode. One notices that the reverse current (left-hand scale) keeps rising strongly after the diode has been depleted, at  $V = V_d$ .

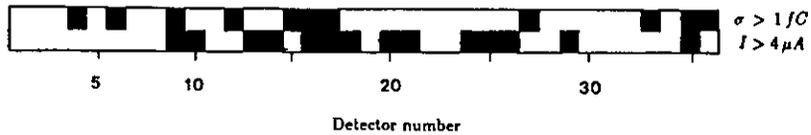


Figure 34: Rejection (■), acceptance (□) plot. The limits are  $\sigma > 1fC$  (upper row) and  $I > 4\mu A$  (lower row) for the noise and the reverse current, respectively.

	lot 1	lot 2
resistivity	(4.0...5.5) kΩcm	(7...9) kΩcm
depletion depth	(245...280) μm	(275 ± 5) μm
depletion voltage	(50...55) V	(32...38) V
reverse current @ 85 V, 20°C	(10 ± 2) nA	(10...40) nA

Electric parameters of the two lots of Hamamatsu diodes.

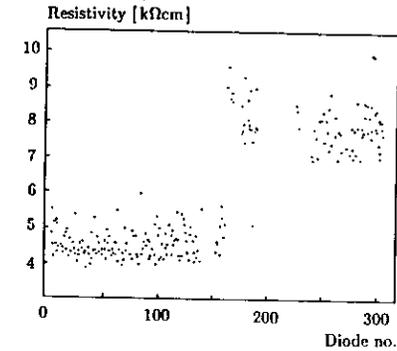


Figure 35: Resistivity of the Hamamatsu diodes v the serial number. One clearly recognizes the two wafer lots.

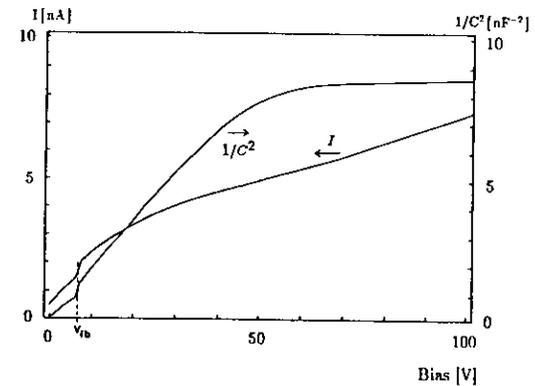
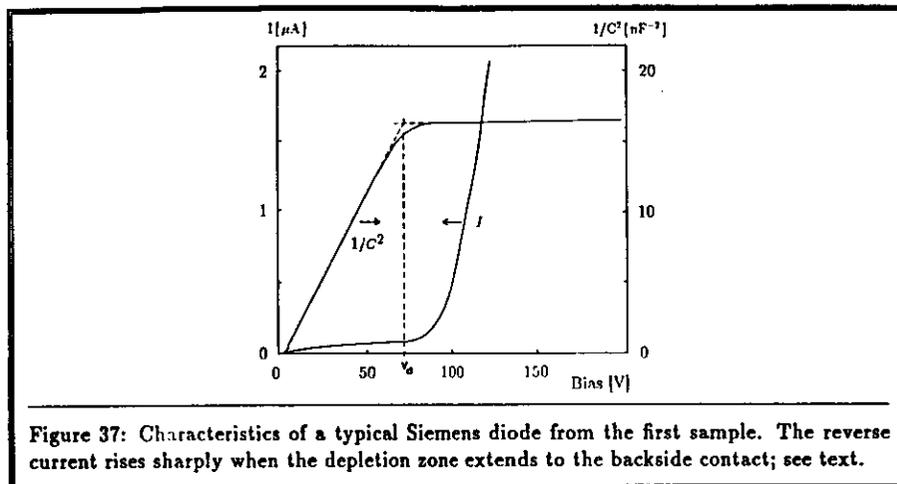


Figure 36: Characteristics of a good Hamamatsu diode. Note the capacitance step at the flat-band voltage  $V_{fb}$ . The step-like current increase is due to the oxide current. At 300V the reverse current is only 15nA (not shown).

resistivity	(5...6.5) k $\Omega$ cm
depletion depth	(350...360) $\mu$ m
depletion voltage	(70...85) V
reverse current @ 100 V, 20°C	(0.1...0.8) $\mu$ A

Electric parameters of the Siemens diodes from the first sample.



## 5 Phenomenology of radiation damage

The silicon detectors to be used at ZEUS will in their final version have an area of  $3.4 \times 3.0 \text{ cm}^2$  and a nominal thickness of  $400 \mu\text{m}$ . A resin will provide protection for the  $p^+$  side, and they will be glued with their  $n^+$  side onto a ceramic board. One board will hold two detectors and their preamplifiers. Such a dual-diode card will form the smallest repetitive unit.

Two detector components of ZEUS will be equipped with these. By far the biggest is the ZEUS hadron-electron separator (HES) with its 50,000 detectors. The HES will be installed inside the ZEUS uranium-scintillator calorimeter. It will be mounted in planes behind the third uranium plate throughout most of the calorimeter. In the forward section another HES plane is foreseen behind the sixth uranium plate.

The other component is a small tungsten-silicon electromagnetic calorimeter, located close to the beam pipe, which is designed to measure the rate of elastic e-p scattering. It will incorporate 36 detectors on 18 cards plus two strip detectors to determine the impact point of the scattered electron.

In both cases the detectors will be exposed to radiation from ionizing particles and neutrons. Their tolerance to this is investigated in this chapter.

### 5.1 Sources of radiation

#### The uranium calorimeter

The HES detectors continuously receive radiation from the enclosing uranium plates. These are hermetically sealed with a 0.2mm thick steel foil which absorbs the alpha particles. The cladding also substantially reduces the dose from  $\beta^-$  particles originating from the decay of the uranium daughter isotopes. At the surface of a cladded plate one measures a  $\beta^-$ -induced dose rate of 50 Gy/y, [ZEUS 86, sect.5.12.2]. The dose rate from the  $\gamma$ 's, following the beta decay, is 10 Gy/y, [ZEUS 86, sect.5.12.2].

The spontaneous-fission half-life of  $^{238}\text{U}$  is  $8.0 \cdot 10^{15}$  years. On average two neutrons are produced in each fission. The resulting flux of neutrons at the surface of a single uranium plate is  $1.5 \cdot 10^6 \text{ cm}^{-2} \text{ y}^{-1}$ .

#### Synchrotron radiation

When the electron beam is bent into the interaction region it emits synchrotron radiation with a critical energy of 70keV. Most of this is caught by a system of collimators and absorbers, but some of it is reflected into the inner clearing of the calorimeter. The HES detectors are well shielded by three radiation lengths ( $3X_0$ ) of uranium in front of them plus another radiation length of material, eg from the superconducting coil. The dose they can receive from synchrotron radiation should therefore be negligible, [ZEUS 86, sect. 5.12].

The beam-pipe calorimeter views the interaction region through an iron flange ( $1X_0$ ) and a tungsten absorber ( $2X_0$ ); like the HES diodes it is well shielded against the synchrotron radiation, [Lo 87].

#### High-energy particles

The rate of high-energy particles entering the ZEUS calorimeter is believed to be much too small to cause radiation damage in the HES detectors. At their position the estimated flux is at most  $2 \text{ cm}^{-2} \text{ s}^{-1}$ . The resulting dose rate of 0.1 Gy/y is utterly negligible, [SR 89, sect.13.4.5].

This is different for the beam-pipe calorimeter. The primary rate of electrons and photons (with an energy comparable to 30 GeV) that hit its 40 cm<sup>2</sup> front surface is estimated to be  $50 \cdot 10^3 \text{ s}^{-1}$ , [Lo 87]. The dose delivered by electromagnetic cascades at the depth of maximum shower intensity (here occurring at  $4X_0$  to  $8X_0$ ) has been measured to be  $3.2 \cdot 10^{-9} (E/\text{GeV})^{0.83} \text{ Gy}$  per incident particle per cm<sup>2</sup>, [Hi 88]. If a beam-time year is taken to be  $2 \cdot 10^7 \text{ s}$  one calculates a dose of 2 kGy/y around the shower maximum.

During the development of the electromagnetic cascade photon-neutrons are produced in  $(\gamma, n)$  and  $(\gamma, 2n)$  reactions. Lead, tungsten, and uranium exhibit large cross-sections for these reactions ( $\sim 0.5 \text{ b}$ ) at  $\gamma$  energies around 12 MeV (giant resonances). It was found that in lead about 0.2 neutrons per GeV of incident energy are produced; the value for uranium is 0.35 n/GeV. Compared to the profile of energy deposition, the longitudinal profile of the photon-neutron flux is shifted deeper into the calorimeter by about two radiation lengths. At 10 GeV its peak occurs at  $7.5X_0$ , and the fwhm is  $9X_0$ , [Hi 88].

## Neutrons

There are several sources of neutrons in the ZEUS detector. The only one of importance for the HES detectors is neutron emission from the beam pipe along the inner clearing of the ZEUS calorimeter. These are produced by 800 GeV protons that accidentally spallate rest-gas atoms. At the inner surface of the barrel calorimeter (1.2 m from the beam pipe) the total neutron flux consisting of direct and albedo neutrons is estimated to be  $1.6 \cdot 10^{10} \text{ cm}^{-2} \text{ y}^{-1}$ , and we will use this as an estimate of the flux received by the HES diodes. Those that are closest to the beam pipe may receive up to four times as much, [SR 89].

The diodes of the beam-pipe calorimeter will reside at an average distance of 7 cm from the proton beam. The direct neutron flux from the beam pipe is estimated to be  $6.5 \cdot 10^{10} \text{ cm}^{-2} \text{ y}^{-1}$ . In addition there are two strong neutron sources nearby. The proton beam will be surrounded by a halo of off-momentum protons with an average energy of 250 GeV. A tungsten collimator ( $C_1$ ) of thickness  $9X_0$  is the first obstacle they encounter inside the ZEUS detector. It has been estimated that about  $10^9$  neutrons per second will be generated there. As the calorimeter is located 20 cm away from  $C_1$ , the neutron flux at its position becomes  $4 \cdot 10^{11} \text{ cm}^{-2} \text{ y}^{-1}$ .

The remaining energy of the hadrons emerging from  $C_1$  is absorbed in the backside end of the rear uranium calorimeter (50 cm from the beam-pipe calorimeter) which produces an even stronger source of neutrons:  $3 \cdot 10^9 \text{ s}^{-1}$ . Taking it to be a point source and disregarding albedo neutrons, it adds another  $2 \cdot 10^{12} \text{ cm}^{-2} \text{ y}^{-1}$ , [SR 89]. The energy-weighted spectrum of these neutrons ( $E \times dN(E)/dE$ ) peaks at 1 MeV kinetic energy and extends up to several MeV. All in all a yearly fluence of  $2.5 \cdot 10^{12} \text{ cm}^{-2}$  is to be expected.

## 5.2 Radiation effects in the detectors

In this chapter we are concerned with macroscopic effects of radiation damage, chiefly the increase in dark current and a change of doping. A description of the microscopic structure and the characteristics of individual defects is deferred to the next chapter.

Charged high-energy particles that traverse a silicon detector create electron-hole pairs along their path and can displace silicon atoms from their lattice sites. In a semiconductor the first effect is a temporary disturbance only, but the latter causes persistent damage.

The energy needed to displace a silicon atom from its lattice site is about 20 eV. The corresponding momentum transfer is 1.0 MeV/c. For kinematical reasons this places a lower

limit on the energy an incoming particle needs to create displacement damage. For electrons the threshold energy is 0.20 MeV while for the much more massive neutrons 150 eV suffice.

The cross-section for displacement by electrons rises sharply at the threshold and soon levels off to a constant value ( $\sim 50 \text{ b}$ , [La 83]). The silicon recoils have low energies in most of the cases since the Coulomb interaction strongly favours small momentum transfers.

This is different with neutrons in two respects. The scattering is more isotropic which for MeV neutrons results in very energetic recoils, which readily displace other atoms creating short cascades and heavily damaged regions. Secondly, the scattering cross-section above neutron energies of 160 keV is no smooth function of the energy but shows a multitude of narrow resonances. The total scattering cross-section is on average 2b to 3b. The total inelastic cross-section peaks at 6 MeV where it reaches 0.8b. Most of this is due to neutron exchange reactions and nuclear excitation. Reactions that change the chemical identity of the silicon atom involved only play a minor role, [McL 88].

In contrast to the bulk material, the insulating oxide is mostly damaged by the primary ionization. Though SiO<sub>2</sub> is an insulator, once electrons and holes are created they are freely mobile. When the detector is under bias during irradiation the mobile charge carriers are swept out of the oxide by the electric field. In contrast to electrons, holes can be trapped at certain defects in the crystal lattice of the oxide. The energy needed to re-emit the hole is 4.3 eV, which means that at room temperature the trapping is permanent. As a result, a positive charge builds up in the oxide. Its limiting value is the number of hole traps present.

## 5.3 Experiments

A number of diodes have been exposed to gamma rays, electromagnetic particle showers, and neutrons. All diodes (made by Hamamatsu) were of the same type. They had a surface area of 9.02 cm<sup>2</sup> and a thickness of 270  $\mu\text{m}$ . Their  $n^+$ -doped backside was glued onto a ceramic plate, and the top was covered with a black plastic foil. Prior to irradiation the  $n$ -type doping was  $0.9 \dots 1.1 \cdot 10^{12} \text{ cm}^{-3}$ , and the dark currents, at full depletion and 20°C, were about 20 nA.

Two detectors (H157 and H158) were for three hours exposed to gamma radiation from an 8,500 Ci <sup>60</sup>Co source at a hospital in Krakow, Poland<sup>6</sup>. The total dose received was measured by a hospital staff member and was found to be 11.7 kGy of human-tissue equivalent. This can be converted into a gamma fluence using standard tabulated conversion factors. At an average gamma energy of 1.25 MeV the corresponding fluence is  $2.1 \cdot 10^{15} \text{ cm}^{-2}$ , [ICRP 73].

Three diodes (H154, H155 and H156) were exposed to electromagnetic showers. To this end they had been installed into a lead-scintillator calorimeter. Its sampling frequency was one radiation length, and the four detector diodes followed the 6<sup>th</sup>, 8<sup>th</sup>, and 10<sup>th</sup> lead plate. At the electron synchrotron in Bonn, Germany, the setup was exposed to a beam of bremsstrahlung gammas produced by 2 GeV electrons hitting a tungsten target. From fluence measurements and EGS Monte Carlo calculations the total dose received by the diodes was calculated to be 16, 10, and 5.4 kGy, [Dw 88a].

The induced radioactivity disallowed the back transport of the calorimeter and the diodes for four weeks, and the first measurements could only be made with some delay.

The neutron irradiation was performed at the radiological department of the Hamburg University Hospital. The facility runs a deuterium-tritium source that produces 14 MeV neutrons. It exploits the nuclear reaction  $^2\text{D} + ^3\text{T} \rightarrow ^4\text{He} + \text{n}$ . A beam of deuterium ions is accelerated

<sup>6</sup>performed by A.Dwurażny

to 0.5 MeV and directed onto a copper cylinder loaded with tritium where the reaction occurs. The strength of the source varies with the beam current and the tritium loading. Such a source is an ideal tool for experimentation: it features a high intensity ( $10^{13} \text{cm}^{-2} \text{h}^{-1}$ ), and, in contrast to a radioactive source, it can be switched off.

Two diodes (H137 and H138) were exposed to neutrons from this source. Mounted at different distances, they received  $2.5 \cdot 10^{13} \text{cm}^{-2}$  (H138) and  $0.72 \cdot 10^{12} \text{cm}^{-2}$  (H137) within 90 minutes. The fluence measurement was provided by a hospital staff member and is believed to be correct to 10%.

## 5.4 Results

### 5.4.1 Uranium activity

For experiments concerning the HES performance, 188 detectors were installed into a uranium calorimeter prototype that except for its size resembled the ZEUS calorimeter in all important detail. After one and a half years ten detectors were removed, and their characteristics were measured again. To a precision of two percent no change in the doping, and to a precision of a few nanoampere no change in the dark current was found. Hence the conclusion that the uranium activity will not noticeably affect the detectors over the life time of the experiment.

### 5.4.2 Results on dark current

The irradiation by neutrons and electromagnetic showers produced a big increase of the dark currents. The virgin diodes had had dark currents around 20 nA at room temperature; afterwards, values at the level of microamperes were observed.

Figure 38 shows the dark current as a function of the received neutron fluence for the two diodes H137 and H138. During the irradiation the reverse current is the sum of two components. One is due to the so far produced defects, and the other comes from ionization by the silicon recoils. Every fifteen minutes the neutron source was switched off, and the persisting current was recorded together with the fluence accumulated so far. The latter quantity was deduced from the integrated deuterium current. It was found that the dark current rises in strict proportion to the received fluence.

For the three diodes exposed to electromagnetic cascades it was not possible to measure the dark current as a function of the received dose during the irradiation. But, if one compares the in total accumulated dose to the dark current, one also finds a linear relationship; cf the inset in figure 39.

The dark currents of all five diodes (measured at a fixed temperature) decreased in the first ~100 days after irradiation and remained constant thereafter, but the reduction has not been followed in detail. The limiting value was 40% for the two neutron irradiated diodes and  $(65 \pm 2)\%$  for the three shower irradiated diodes. Note, however, that the first measurements on the latter could be made only four weeks after the exposure.

The dark current of a detector strongly depends on its temperature. It will be shown in the next chapter that the expected dependence is

$$I(T) \propto T^2 \exp\left(-\frac{A}{kT}\right)$$

where  $k$  is Boltzmann's constant, and  $T$  is the absolute temperature. The quantity  $A$  is called the activation energy and its connection with the defect that causes the current will also be discussed in the next chapter.

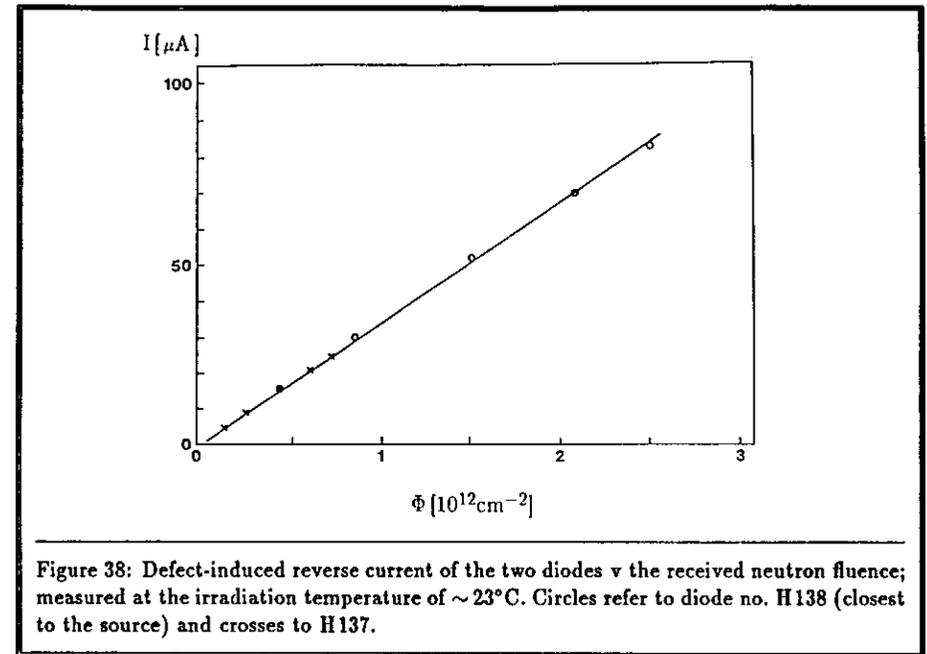


Figure 38: Defect-induced reverse current of the two diodes  $v$  the received neutron fluence; measured at the irradiation temperature of  $\sim 23^\circ\text{C}$ . Circles refer to diode no. H138 (closest to the source) and crosses to H137.

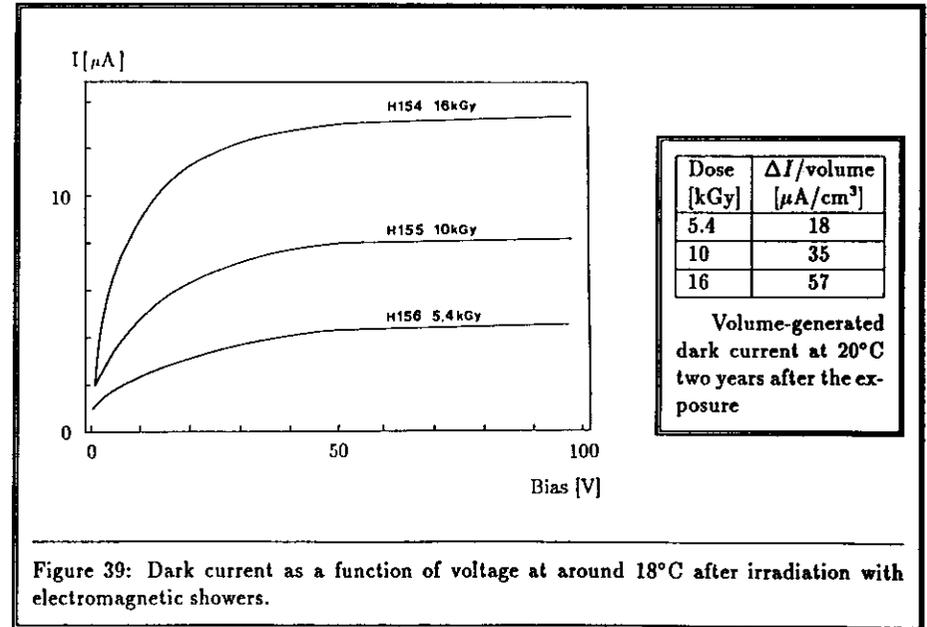


Figure 39: Dark current as a function of voltage at around  $18^\circ\text{C}$  after irradiation with electromagnetic showers.

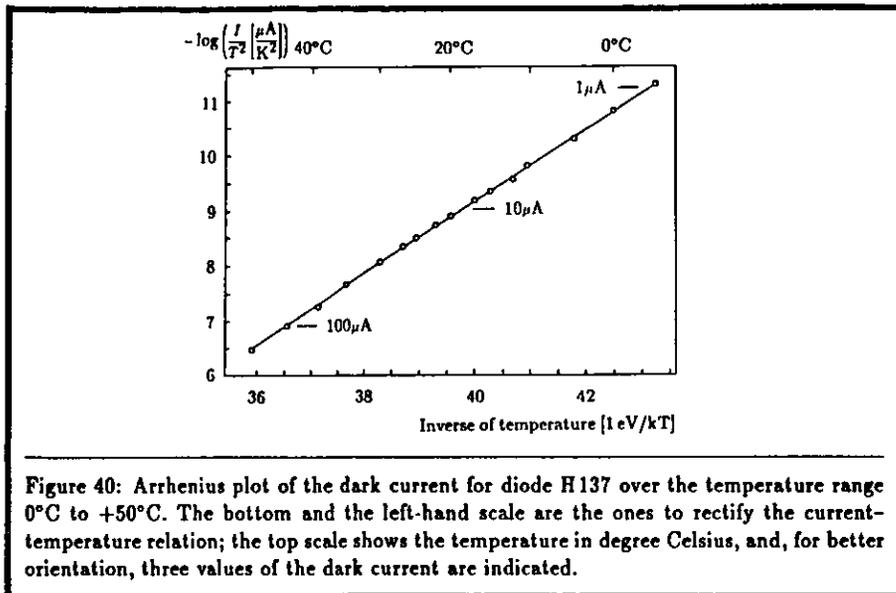


Figure 40: Arrhenius plot of the dark current for diode H137 over the temperature range 0°C to +50°C. The bottom and the left-hand scale are the ones to rectify the current-temperature relation; the top scale shows the temperature in degree Celsius, and, for better orientation, three values of the dark current are indicated.

The above relationship can be rectified. Plot  $Y = -\log(I/T^2)$  as a function of  $X = 1/kT$  (Arrhenius plot), and the data should fall on a straight line. Its slope is the activation energy. Figure 40 shows an example.

To within errors the activation energy of the dark current is the same in all diodes, whether irradiated with neutrons or electromagnetic showers:  $A = (0.65 \pm 0.01)\text{eV}$ . At 300K the dark current approximately doubles every 8°C.

Prior to irradiation the dark currents at 20°C were about 20nA, and the activation energy was  $(0.80 \pm 0.02)\text{eV}$ .

Let us summarize. To obtain comparable quantities, the currents are quoted as current per depleted detector volume and at 20°C. The currents are those measured two years after the irradiation:

- Electromagnetic cascades, induced in a lead-scintillator calorimeter by bremsstrahlung gammas from 2GeV electrons, yield

$$\frac{\Delta I}{\text{volume} \times \text{dose}} = (3.3 \pm 0.3)\text{nA/cm}^3\text{Gy}$$

- 14MeV neutrons yield

$$\frac{\Delta I}{\text{volume} \times \text{fluence}} = 4.6 \cdot 10^{-17}\text{A/cm}^3\text{cm}^{-2}$$

To compare these values to what is reported in literature is not easy. In the diodes exposed to electromagnetic showers the damage is caused by electrons and positrons with a wide range of

energies. To this is added the damage caused by an unknown, but probably sizable, fluence of photon neutrons also of a wide energy spectrum. Since no similar experiment has been reported, no comparison can be made.

Moreover, the reader must be warned that applying the above-given rate of dark current increase with electromagnetic dose may be inappropriate if the experimental situation is different. The reason is that the relative importance of the photon neutrons is unknown. If the electromagnetic calorimeter employs iron as the absorber their contribution is probably negligible. On the other hand, with uranium as the absorber they are 75% more numerous than with lead, [Hi 88]. Independent of the production rates, the actual fluence of neutrons very much depends on the calorimeter structure, especially its readout material. Scintillator layers used for the readout act as a moderator and absorber; the neutron fluence is higher without them.

To compare with literature the dark current increase per unit fluence of 14MeV neutrons should be more straightforward. But, again, one encounters problems.

In this study the dark currents were observed to decrease with time. After about 100 days they reached a floor level at about 40% of the value measured a few hours after the irradiation that itself took ninety minutes. In most cases the authors do not report whether they observed the same phenomenon or not, and at what time they performed their measurements.

Besides this one should be aware that the rate of dark current increase is not a universal constant. If the defect that causes the dark current involves some impurity of the host crystal then the production rate will be material dependent. However, measurements done on high-resistivity detector-grade material yield similar results. E.Fretwurst et al., [Fr 90], find  $\alpha = 6.0 \cdot 10^{-17}\text{Acm}^{-1}$  at 20°C, measured a year after the irradiation. Using a  $\text{PuBe}$ -source that emits neutrons chiefly in the range from 3MeV to 6MeV, H.Kraner et al. [Kr 89] found  $\alpha = 6.1 \cdot 10^{-17}\text{Acm}^{-1}$  with a scatter of  $\pm 25\%$  after irradiations that lasted for 400 to 800 hours. Hasegawa et al., [Ha 89], find  $\alpha = 6.1 \cdot 10^{-17}\text{Acm}^{-1}$  after irradiations with reactor neutrons lasting seconds only (thermal plus  $\sim 1\text{MeV}$ ). They noticed a gradual reduction of the dark current over a period of 50 days but do not give details. Given the differences in neutron energies, experimentation details, and the fact that the induced dark currents in general change with time, the numerical similarities among the  $\alpha$  values should be regarded as purely coincidental.

In most of the cases, the activation energy of the dark current is not reported. Exceptions are Hasegawa et al., who give  $A = 0.58\text{eV}$ , and Srour et al.,  $A = 0.64\text{eV}$ .

Proton irradiation also causes a strong increase in dark current, and its activation energy matches the value from neutron irradiation: 0.65eV by Dietl et al., [Di 87], 0.64eV by Nakamura et al., [Na 88], and 0.62eV by Ohsugi et al., [Oh 88].

H.Kröger has reported on irradiating detectors very similar to the ones used for this work with neutrons from a  $^{252}\text{Cf}$  fission source. The neutrons have a Maxwellian spectrum with a temperature  $kT = 1.42\text{MeV}$ , [Fr 90a]. With a decay time of  $\sim 15$  days the dark current reduced to 42% of the immediate post-irradiation value. Measured three months after the exposure the dark current introduction rate is  $\alpha = 2.4 \cdot 10^{-17}\text{Acm}^{-1}$ . The measured activation energy was 0.63eV, [Kr 90].

We see that the dark current is caused by a defect with an energy level very close to the band-gap centre—a mid-band trap. Its rate of production is fairly insensitive to the neutron energy. It shall receive further attention at the end of chapter 6.

### 5.4.3 Results on doping changes

The defects introduced by radiation can affect the net doping in the detector volume. This was investigated for those diodes exposed to electromagnetic showers. The doping concentration of the virgin diodes was derived from the capacitance-voltage characteristic as explained in chapter 3. One plots  $1/C^2$  as a function of the reverse bias and obtains the effective doping from the slope straight-line part of the curve.

But this method fails with the irradiated diodes. Due to carrier trapping at the crystal defects the measured capacitance now strongly depends on the temperature and frequency. No more is it related in a simple way to the width of the depletion zone and the doping.

An alternative is to measure the dark current as a function of the applied bias. If currents due to surface effects are sufficiently small and if the net dopant density is constant throughout the detector volume, the current should be proportional to the square root of the sum of the applied bias and the diode built-in voltage. The depletion voltage  $V_d$  can thus be extracted from the  $I^2$  vs bias characteristic in the same manner as from a  $1/C^2$  measurement. Together with the known detector thickness ( $d$ ) the net doping ( $N_D$ ) can be calculated (cf chapter 3):

$$V_d = \frac{eN_D}{2\epsilon} d^2$$

The measurements gave that the net doping is reduced in proportion to the received dose. The side table gives an account on this. Measured at around 20°C the reduction per dose is:

$$\frac{\Delta N_D}{Dose} = 4.5 \cdot 10^7 \text{cm}^{-3} \text{Gy}^{-1}$$

For the neutron-irradiated H137 the reduction was:

$$\eta := \frac{\Delta N_D}{fluence} = \frac{0.45 \cdot 10^{12} \text{cm}^{-3}}{0.72 \cdot 10^{12} \text{cm}^{-2}} = 0.63 \text{cm}^{-1}$$

Again, for the shower-irradiated diodes no comparable measurements are reported in literature. There are but a few for neutrons. Lindström et al., [Li 89], report  $0.23 \text{cm}^{-1}$  for 14 MeV neutrons with a scatter of 30%, but they have abandoned the idea of a constant rate in a later publication, [Wu 91]. H.Kröger, [Kr 90], found  $\eta = 0.17 \text{cm}^{-1}$  after irradiation with the lower-energetic  $^{252}\text{Cf}$  neutrons. Edwards et al., [Ed 91], give  $\eta = 0.033 \text{cm}^{-1}$  for  $\sim 1 \text{MeV}$  neutrons. Using reactor neutrons and fluences of up to  $10^{14} \text{cm}^{-2}$  Hasegawa et al., [Ha 89], found type inversion in Hamamatsu n-type photodiodes. From their data one can infer an introduction rate of  $\eta = 0.06 \text{cm}^{-1}$ .

It is tempting to attribute the differences to the different neutron energies, but there may be material dependencies as well.

Dose [kGy]	$N_D$ , before [ $10^{12} \text{cm}^{-3}$ ]	$\Delta N_D$ [ $10^{12} \text{cm}^{-3}$ ]
5.4	0.98	0.27
10	0.92	0.47
16	0.92	0.74

Net doping of the virgin diodes and its reduction ( $\Delta N_D$ ) measured two years after the irradiation. Errors margins are 5% for  $N_D$ (before) and 10% for  $\Delta N_D$ .

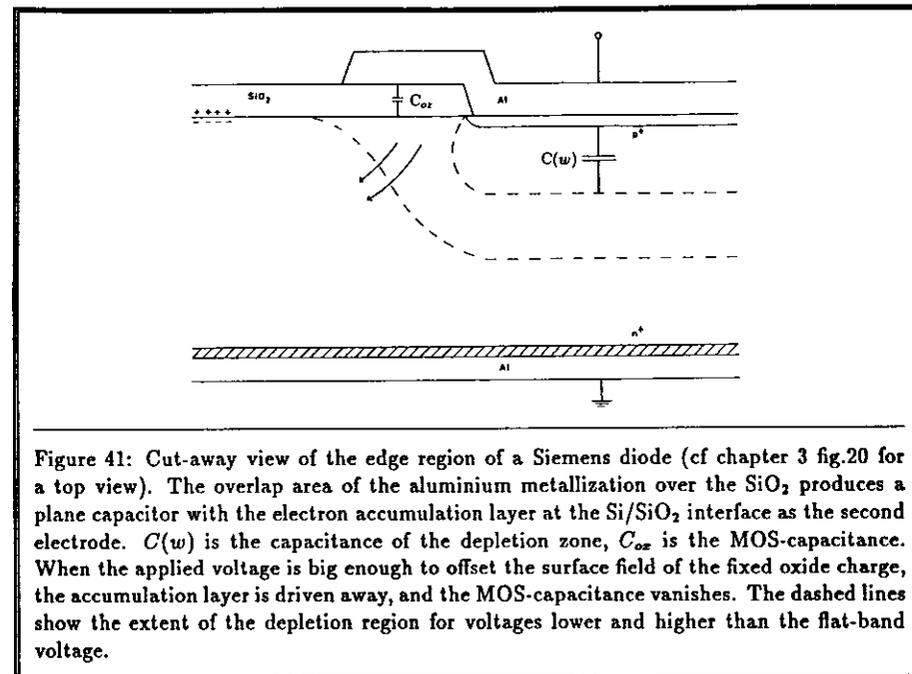


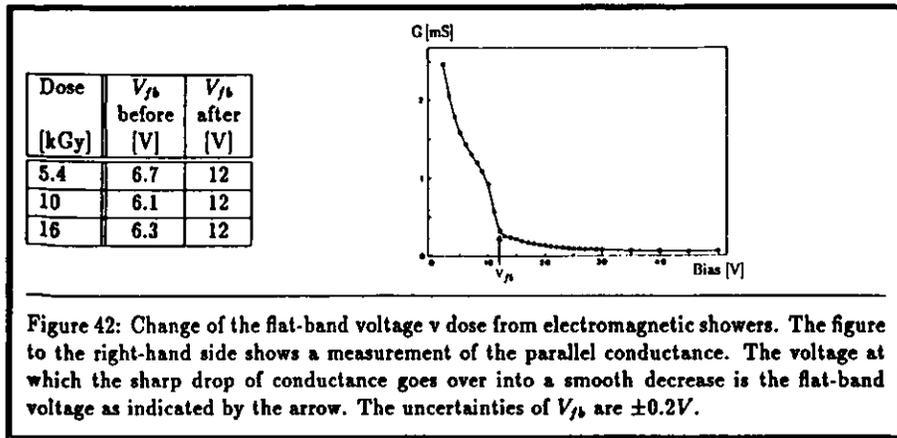
Figure 41: Cut-away view of the edge region of a Siemens diode (cf chapter 3 fig.20 for a top view). The overlap area of the aluminium metallization over the  $\text{SiO}_2$  produces a plane capacitor with the electron accumulation layer at the  $\text{Si}/\text{SiO}_2$  interface as the second electrode.  $C(w)$  is the capacitance of the depletion zone,  $C_{ox}$  is the MOS-capacitance. When the applied voltage is big enough to offset the surface field of the fixed oxide charge, the accumulation layer is driven away, and the MOS-capacitance vanishes. The dashed lines show the extent of the depletion region for voltages lower and higher than the flat-band voltage.

### 5.4.4 Results on oxide charging

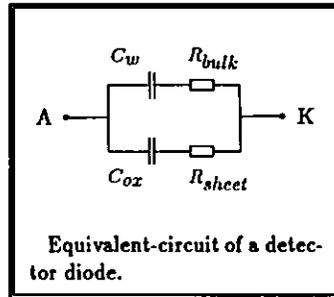
To clarify the effect we are dealing with, figure 41 shows a cut-away view of the edge region of a Siemens diode<sup>9</sup>; compare with figure 20 in chapter 3. On top, the contact covering the  $p^+$  layer is seen. The aluminium (Al) overlaps the silicon dioxide that insulates the  $p^+$  area from the surrounding  $n^+$  field stop ring (which would be outside of the figure to the left-hand side). The oxide contains positive charges which on the semiconductor side attracts electrons to form a very thin (a few nm) conductive  $n^+$  sheet below the  $\text{Si}-\text{SiO}_2$  interface. This sheet connects to the  $n^+$  field stop. With no or little bias applied, the sheet extends under the region where the aluminium covers the oxide. The  $p^+$  layer is always isolated from the  $n^+$ -type sheet by the depletion zone of the  $p^+n^+$  diode which these two form.

If the  $p^+$  contact is biased negatively to reversely bias the diode, an electric field is produced in the oxide below the aluminium that tends to offset the influence of the positive oxide charge, which reduces the number of electrons attracted to the silicon-oxide interface. Once the applied bias is big enough to sweep away all electrons, the depletion zone extends below the aluminium overlap. The voltage that is just sufficient to completely drive away the electron sheet is called the flat-band voltage  $V_{fb}$ ; cf chapter 3.2.4.

<sup>9</sup>With the Hamamatsu diodes the structure seen under the microscope was more complex, and the manufacturer did not give details. However, the principle is the same in both designs.



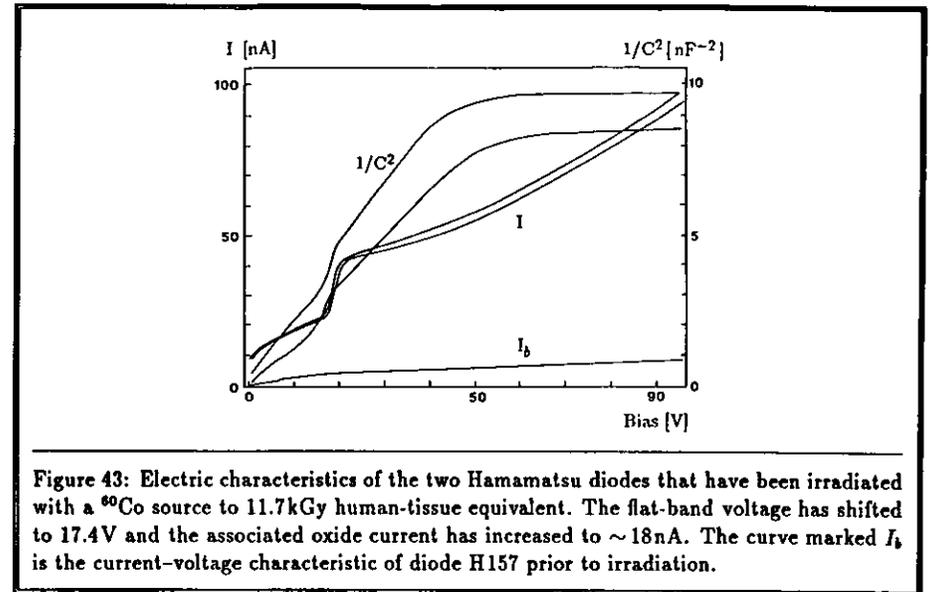
The impedance characteristics of the detectors reflect this in various ways. An equivalent-circuit model for the impedance measured between the anode (A) and cathode (K) is given in the side figure. The two capacitances are the capacitance of the depletion zone  $C(w)$  and the oxide or MOS-capacitance  $C_{ox}$ .  $R_{bulk}$  is the resistance of the undepleted bulk material between the  $n^+$ -layer and the end of the depletion zone.  $R_{sheet}$  is the resistance between the  $p^+$ -layer and the  $n^+$  guard ring which in turn is (deliberately) connected to the  $n^+$  back contact.  $R_{sheet}$  is small as long as there is an electron accumulation layer below the part of the oxide which is covered by the  $p^+$ -metallization. It becomes big when the electron layer is driven away.



The device capacitance shows a step-like reduction by  $C_{ox}$  at this point as has been demonstrated in chapter 4. The effect on the conductance (real part of the admittance) is even more pronounced, and the conductance-voltage characteristic is better suited to determine the flat-band voltage. The figure besides table 42 shows the behaviour of the parallel conductance in a bias range around the flat-band voltage for one of the shower-irradiated diodes (H155). The transition of the sharp drop in conductance to the region of smooth decrease with voltage is quite pronounced. It shall serve as the experimental definition of the flat-band voltage.

The irradiation generally reduced the conductance, but except for the shift in the flat-band voltage the characteristics changed little. For the diodes exposed to electromagnetic showers one finds that the flat-band voltage has shifted (from different starting values) to 12V in all three cases, though the received doses differ by a factor of three; cf table 42. This indicates a saturation of the hole traps present in the oxide.

Surface defects can be investigated in more detail in gamma-irradiated diodes. At the silicon-oxide interface the periodicity of the crystal is abruptly interrupted, and there is a high surface density of unsatisfied bonds from silicon and oxygen atoms. This introduces a continuous distribution of electronic states into the band gap. As a result, thermal pair creation is strongly



enhanced at the interface. As long as the applied bias is smaller than the flat-band voltage there is an accumulation layer of electrons below the oxide (the conductive sheet) and hence no electric field; thermally created electron-hole pairs recombine on the spot. But if the conductive sheet has been driven away, the pairs are separated by the now present electric field. They are collected on to the detector terminals and contribute to the dark current. We call the current originating from surface states at the silicon-oxide interface the oxide current  $I_{ox}$ . It is not large, and in the case of the shower and neutron-irradiated diodes it is hidden among the large volume-generated dark current.

This is different for the  $^{60}\text{Co}$ -irradiated diodes. The averaged energy of the two  $\gamma$  lines is 1.25MeV. They produce ionization only indirectly by Compton scattering, and the Compton electrons produce only very little damage in the bulk. This renders possible a study of the surface effects. As figure 43 shows, the dark currents are still in the nanoampere range after the diodes received about 10kGy of dose. The  $1/C^2$  curves exhibit the characteristic steps as the bias approaches the flat-band voltage. Once it is reached, one observes a step-like rise of the current. The height of this is the oxide current  $I_{ox}$ . The side table summarizes the results of measurements made around 20°C. Again we find that  $V_{fb}$  has approached a fixed value independent of the starting value. The irradiation has also increased the surface density of states, which is reflected by a growth of the oxide current.

	H157	H158
$I_{ox}$ before	0.6nA	0.6nA
$I_{ox}$ afterwards	19nA	17nA
$V_{fb}$ before	$6.8 \pm 0.2V$	$7.7 \pm 0.2V$
$V_{fb}$ afterwards	$17.4 \pm 0.2V$	$17.4 \pm 0.2V$

Results concerning surface effects of 11.7kGy of  $^{60}\text{Co}$  irradiation measured at around 20°C.

## 5.5 Operation at ZEUS

In estimating how long the detectors will remain operable at ZEUS, let us first concentrate on the HES detectors. It was shown that the detectors can easily tolerate 10kGy of hard  $\gamma$ 's and several kGy from electromagnetic showers. Thus, neither the radiation from the uranium radioactivity nor the dose received from high-energy particles are a problem. All the damage will be done by spallation neutrons which are produced by 800GeV protons hitting rest gas atoms in the beam pipe along the inner clearing of the ZEUS calorimeter. The diodes in the barrel calorimeter will receive an estimated yearly fluence of  $1.6 \cdot 10^{10} \text{cm}^{-2}$ . The prototype detectors could tolerate a fluence of  $2.5 \cdot 10^{12} \text{cm}^{-2}$  of 14MeV neutrons. These are known to produce two times more damage than  $^{252}\text{Cf}$  neutrons whose spectrum is quite close to the spectra that will be encountered at HERA, [Kr90], [Te90]. It is thus safe to conclude that the HES diodes will survive through the life time of the experiment.

This is less clear for the diodes in the beam-pipe calorimeter. They receive a yearly dose of 2kGy from electromagnetic showers and a sizeable, but rather badly known, yearly fluence of spallation neutrons of  $\sim 2.5 \cdot 10^{12} \text{cm}^{-2}$ . Adding together the effects of both, one expects for  $10 \text{cm}^2 \times 400 \mu\text{m}$  diodes, at 20°C, a yearly increase of dark current of 13 $\mu\text{A}$ . As long as the electronic noise induced by it can be handled and as long as the preamplifiers and line drivers tolerate this level of radiation, the beam-pipe calorimeter should remain operable for a few years—provided clean machine running.

## 6 Point defects and their characterization

About fifteen years ago, silicon counters operating at room temperature were added to the experimenter's tool box in the field of high-energy physics. The first detectors had their readout electrodes patterned into strips, hence strip detectors. Two such detectors placed in series, with their strip directions crossed, provided spatial resolution in two dimensions. Strip detectors have evolved since and are now manufactured in quantities. They come on chips as large as  $5 \times 5 \text{cm}^2$  and with as many as 200 strips per cm. It is even possible to pattern both sides of the chip into mutually orthogonal readout strips so that a single detector can provide  $x - y$  position measurements. Their field of application is micro-vertexing. They form the position sensitive detector located closest to the interaction point in a collider experiment.

As the price of detector-grade silicon fell and the production yields increased, chip prices plummeted, and physicists considered building electromagnetic calorimeters with silicon detectors for the readout. In such an application the good position resolution of strip detectors is unnecessary; silicon detectors for calorimetry have no segmentation, and chips of  $3 \times 3 \text{cm}^2$  or even  $5 \times 5 \text{cm}^2$  in size are read out as a single unit.

These two types of detectors, the micro-strip and the large-area detector, are the ones most commonly used in the 'big' experiments of high-energy physics. Besides them a number of other devices like the silicon drift chamber, pixel detectors, and fully depleted CCD's have been introduced in the recent years. But to make them fit the needs of experimentation at the forthcoming hadron colliders still requires a great deal of R&D.

From the beginning it was clear that the radiation which was to be detected (ionizing particles) would damage the detectors. In the conventional devices the main source of damage is that incoming particles readily displace silicon atoms from their lattice sites thereby creating crystal defects. These affect detector operation; with time they accumulate, and the detector will eventually cease to function properly.

There are several well-established methods which allow to characterize the various crystal defects—with the important restriction that the defects must not outnumber the dopant atoms. In the latter case, which we will refer to as high-level radiation damage, no such methods exist, and this work suggests the use of a new method to then replace the others.

This chapter is organized as follows: The first section is an introduction to the relevant theory of crystal defects. There it is shown how they can be incorporated into the thermodynamics of the host crystal—ie how to calculate the equilibrium concentrations of charge carriers and of defects in their various charge states. Stationary and non-stationary processes involving the emission of electrons and holes from a deep defect are studied at the end.

Based on this theory, an experimental technique to record in real time the electric current due to non-stationary emission of charge carriers from a deep-level defect is introduced in section 2. It is explained how relevant information about the defect can be extracted from the current transients. The experimental setup is described in detail.

Section 3 covers the results obtained with four detectors. One of them was irradiated with 14MeV neutrons and the others were exposed to electromagnetic showers. In all of them large concentrations of divacancies<sup>10</sup> have been found. The first acceptor level of the divacancy could be studied in detail including its enthalpy and entropy of ionization.

The macroscopic radiation damage, ie dark current increase and an apparent reduction of doping, is traced back to the production of an acceptor-like defect, or defect cluster, with an energy level close to the bandgap centre.

<sup>10</sup>If the silicon atoms of two neighbouring lattice sites are missing, the defect is called a divacancy.

## 6.1 Theory of point defects

### 6.1.1 Crystallography

Silicon is the second element in the fourth main chemical group, following carbon and preceding germanium. All elements of this group crystallize in the diamond lattice. This may be thought of as two interspersed face-centred cubic lattices of which the one is shifted with respect to the other by  $\sqrt{3}/4$  lattice constants along the cube diagonal. There are two atoms per unit cell, and the lattice constant is 0.357 nm (for silicon). Each atom is bound to its four nearest neighbours by covalent bonds.

The wafers on which the detectors are fabricated are cut from rods very precisely along some crystallographic plane. Most commonly, this coincides with the (1,0,0) plane because this minimizes the surface density of unsatisfied bonds.

**Impurities:** The single crystals are perfect in the sense that they are grown free of dislocations and stacking faults. But they cannot be free of contaminants. The most abundant are carbon and oxygen. In the purest material that is commercially available (float-zone refined silicon) their concentration is of the order of  $10^{16} \text{cm}^{-3}$ , [Am 74]. During the thermal oxidation step in the detector manufacturing process the oxygen concentration rises to  $O(10^{18} \text{cm}^{-3})$ .

Belonging to the same chemical group as silicon, the carbon atom fits well into the host lattice and exhibits no electrical activity. Oxygen, in contrast, is always at an interstitial position, but remains electrically inactive; it acts neither as a donor nor as an acceptor.

**The vacancy:** Creating a vacancy in silicon leaves behind four unsatisfied (dangling) bonds from the neighbouring atoms. If these remained in their places the defect would have the tetrahedral symmetry of the host crystal. But in addition to its one dangling bond each next-neighbour atom has three back bonds which mix with the broken one. Depending on the charge state of the vacancy, symmetry-lowering lattice distortions occur. In the neutral vacancy ( $V^0$ ), for example, the nearest neighbours pull together tending to form two satisfied bonds from the four broken ones. This lowers the defect symmetry to the tetragonal group  $D_{2d}$  (Jahn-Teller distortion), [Sc 82], [Wa 84]. It is these lattice distortions that make it so far impossible to calculate the electron wave functions produced by the defect.

The vacancy is mobile even at liquid nitrogen temperature. The migration process, however, is athermal and not understood, [Fa 89]. Migrating through the crystal, the vacancies may be captured by other impurities forming immobile vacancy-impurity complexes. Some may eventually reach the crystal surface where they lose their identity. When silicon is irradiated at room temperature one afterwards detects no vacancies but only vacancy-related defects. Vacancies have only been detected in samples irradiated and measured at  $\sim 4 \text{K}$ .

**The interstitial:** Very little is known about the interstitial. It introduces no extra energy level into the bandgap, [So 82], and cannot be detected with DLTS or EPR<sup>11</sup>. Only recently a model for silicon self-interstitial migration and defect formation has been proposed, [As 87]. The authors suggest that interstitial silicon is solely captured at substitutional carbon with which it exchanges places in the Watkins replacement mechanism. The now interstitial carbon is later captured by oxygen, carbon, or phosphorus giving rise to low-lying donor states of which  $C_i - O_i$  at 0.36 eV and  $C_i - P_i$  at 0.27 eV above the valence band are the most prominent; see

<sup>11</sup>DLTS: Deep Level Transient Spectroscopy, see end of this chapter;  
EPR: Electron Paramagnetic Resonance.

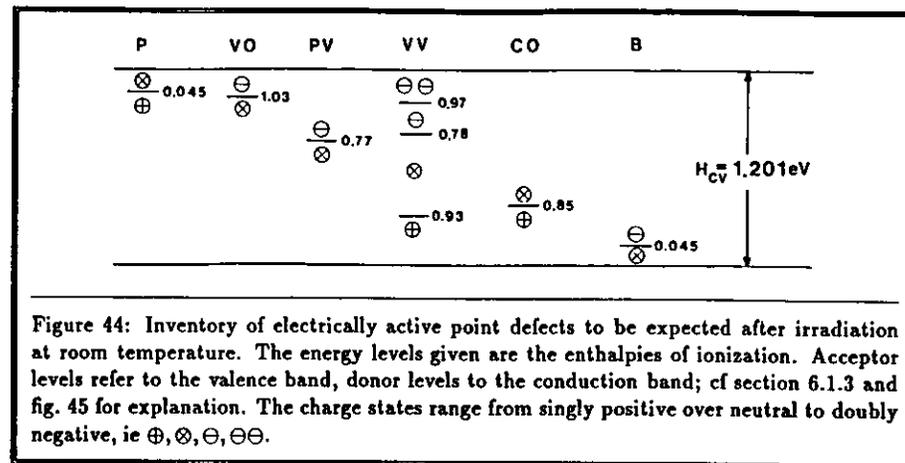
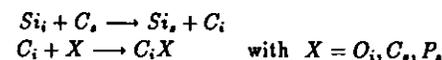


Figure 44: Inventory of electrically active point defects to be expected after irradiation at room temperature. The energy levels given are the enthalpies of ionization. Acceptor levels refer to the valence band, donor levels to the conduction band; cf section 6.1.3 and fig. 45 for explanation. The charge states range from singly positive over neutral to doubly negative, ie  $\oplus, \otimes, \ominus, \ominus\ominus$ .

[Si 76], [Wa 73] and references therein. The reaction is pictured as:



**The divacancy:** Removing two adjacent atoms from the lattice leaves behind a divacancy. It is stable and immobile at room temperature. Besides the neutral charge state it has three charged ones:  $VV^+$ ,  $VV^-$ , and  $VV^=$ . The divacancy is called an amphoteric defect: Depending on the Fermi level it is either a low-lying donor or a high-lying double acceptor.

**Vacancy-impurity:** The migrating vacancies combine with either oxygen or group-V dopants (donors) to form stable, immobile complexes. All of them are high-lying acceptors with a single (0/-)-level in the upper half of the bandgap. In the low-doped, n-type materials, used for particle detectors, the oxygen atoms outnumber the phosphorus atoms by four orders of magnitude. Diffusing vacancies are therefore predominantly captured by oxygen. At room temperature vacancies do not bind to group-III dopants, which are shallow acceptors.

The inventory of electrically active defects to be expected after room temperature irradiation is given in figure 44. The cited energy levels are the enthalpies of ionization; see section 6.1.3, Ionizing a defect.

### 6.1.2 Coulombic defects

Isolated defects in an otherwise perfect crystal add extra states to the set of pure-crystal electronic (and phononic) states. Energy levels within the bandgap give rise to bound states. In the group-IV covalent semiconductors structural defects (vacancies etc.) and most of the substitutional defects form bound electronic states localized within two or three elementary cells. As an exception, substitutional group-V and group-III elements form hydrogen-like bound states extending over many elementary cells. These shallow donors and acceptors can be treated by the effective-mass theory, an account of which shall be given below. For deep non-hydrogenic energy levels no simple theory exists. The change in the density of states has been calculated for a number of simple point defects using methods based on the extended Hückel theory

or using self-consistent Green's functions methods. While these techniques reproduce well the macroscopic quantities of the semiconductor (heat of formation, modulus of elasticity, heat conductivity, and the like) they fail to describe the electronic properties of deep defects. Therefore, most of our knowledge comes from experiments.

Effective-mass theory is applicable to the hydrogen-like states of shallow donors and acceptors. The bound-carrier wave function is constructed from Bloch states in momentum space or Wannier functions in real space. The free-electron (continuum) solutions of the Schrödinger equation do not propagate with the free electron mass but with a direction-dependent effective mass  $m^*$ , which is a second rank tensor. Its inverse is usually defined via the dispersion relation  $M_{\alpha\beta}^{-1} = \partial^2 E(k)/\partial k_\alpha \partial k_\beta$ , in which  $k$  is the vector of crystal momentum, and the derivatives are taken at the band extremum. If there is more than one energy band involved the individual contributions to  $M_{\alpha\beta}^{-1}$  sum up. This is called the parabolic-band approximation. It is valid as long as the free-carrier concentration is small compared to the density of states at the band extremum and as long spin-orbit coupling is small.

The zero-point of the energy scale is the lowest energy level above which continuum solutions exist. For a donor this is the minimum of the conduction band  $E_C$ . The polarization of the medium is taken into account by writing the electric potential as  $-e^2/\epsilon r$  with  $\epsilon$  denoting the static dielectric constant. As long as the wave function extends over many elementary cells the use of a macroscopic material constant is justified, and the potential varies slowly over the region of interest. Only the 1s ground state,  $\Psi = Ae^{-r/a}$ , has a non-vanishing density at  $r = 0$ , where the potential changes rapidly over the unit cell. Under the restrictions noted above, the effective-mass theory yields for the energy levels the hydrogen values scaled with  $m^*$  and  $\epsilon$ :

$$E_n = E_C - 13.6 \text{ eV} \frac{m^*}{m} \frac{1}{n^2}$$

The Bohr radius scales as:

$$a = 0.53 \text{ \AA} \frac{\epsilon}{m^*/m}$$

The complexity of the energy bands (anisotropy of the effective mass, multiple band minima, spin-splitting etc.) is absorbed into  $m^*$ . It is calculated as an average value from  $M_{\alpha\beta}^{-1}$  and will be different for the various excitations of the defect. The degeneracy of the p-states, for instance, is resolved due to the ellipsoidal effective mass in silicon. The ground-state binding energy in silicon is 46 meV for phosphorus. It varies by  $\sim 10\%$  for As and Sb. The experimental values for excited-level spacings are in agreement with theory to better than 1%, [Pa 79], [Fa 68].

While the thermodynamics of shallow acceptors and donors can be derived from the effective-mass theory, the thermodynamics of deep defects has been developed heuristically because no theory of the bound states exists for the deep defects.

### 6.1.3 Thermodynamics involving deep defects

There is a great number of defects in silicon that may under suitable conditions exist in more than one charge state. The 'natural' state is the neutral one, and by capturing a hole or an electron the defect can become ionized. Which charge state is prevalent in a chunk of semiconductor depends on conditions to be described below.

Many methods used to characterize defects (including CTR) disturb the stationary situation suddenly and observe by some means how the system returns to stationary conditions. The

initial disturbance may, for example, be a flash of light that ionizes the defects (by photoexcitation), and the quantity that is monitored afterwards is the conductivity, which reduces at the same rate at which the freed charge carriers are again captured by the defects.

If there is a diode fabricated on the silicon chip there is another way to create the sudden disturbance. The silicon detectors used in this work are  $p^+nn^+$  diodes in which the weakly doped  $n$  region has a thickness of  $300 \mu\text{m}$  while the  $p^+$  and the  $n^+$  layers are very thin ( $< 0.3 \mu\text{m}$ ). Recall from chapter 3 that there is a field zone extending from the  $p^+$  layer into the crystal; how far depends on the bias: The bigger the bias the wider the field zone. The electric field is strongest at the  $p^+ - n$  junction and reduces to zero over some fraction of the distance to the  $n^+$  layer. In the remaining part of the diode (from the end of the field zone to the  $n^+$  layer) there is no electric field. That part (called the base) is in thermal equilibrium while the field zone (with bias applied) certainly is not.

With this difference in environment, consider a kind of defects that is charged in the base but neutral in the field zone. We can dictate the width of the field zone by varying the reverse voltage. Pulsing the diode farther into depletion brings charged defects into the field zone where they now discharge by carrier emission. One can monitor the concomitant change of the diode capacitance (as in DLTS) or monitor the electric current (as in CTR) to measure the rate of emission.

In order to see what can be learned from such measurements we have to develop some theory that answers the following questions:

- Which is the predominant charge state of a certain defect in thermal equilibrium?
- Which charge state prevails in the field zone?
- What is the time law governing the return to stationary conditions after a sudden disturbance?

As a starter we introduce into the basic physics and provide the necessary definitions.

**Pair creation:** As the silicon atoms bind together to form the crystal the single-atom electron states merge together into bands. The band of highest energy which is almost completely filled is called the valence band. The naming reflects the fact that the electron states of this band form the covalent bonds (valences) that hold the crystal together. In a dispersion diagramme, the valence band attains its maximum is at the zero point—ie at the centre of the Brillouin zone (the  $\Gamma$ -point). The bands above the valence band are called conduction bands. In silicon—a semiconductor—they are almost, but not completely, empty. In terms of crystal momentum the conduction band which is lowest in energy has six equivalent minima. These lie along the (1,0,0) axis and the symmetrically equivalent directions, the  $\Gamma - X$  directions. The minima occur at a distance of about three quarters of the maximum Brillouin momentum away from the  $\Gamma$  point. Electron wave functions belonging to this band are antibonding.

Lifting an electron from the valence band maximum to the conduction band minimum is referred to as electron-hole pair creation. Removing the electron from the valence band leaves behind an unoccupied electron state in an almost completely filled band, which very much resembles a positively charged particle; this is called a hole. Placing the removed electron into the conduction band produces a freely mobile electron. The process is called electron-hole or pair creation, and we have to examine its effect on the lattice and how much energy is needed.

At the absolute zero of temperature that energy is 1.20 eV. At finite temperatures, removing an electron from a bonding band and placing it into an antibonding band weakens the lattice

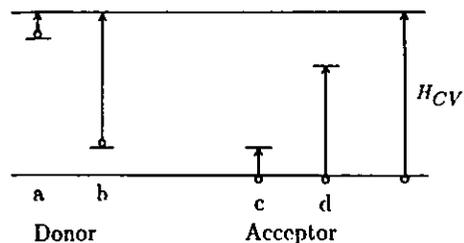


Figure 45: The circle (o) symbolizes the electron involved in ionizing the defect in question at the absolute zero of temperature. There is a shallow donor (a), a shallow acceptor (c), a deep donor (b), and a deep acceptor (d). Donor level energies are given with respect to the conduction band, acceptor level energies with respect to the valence band.

and causes the vibrational entropy to grow. The enthalpy of the reaction is the energy that were necessary if the growth in entropy could be left unaccounted for. The actual energy needed—the free energy—is smaller than the enthalpy because of the gain from the growing disorder in the crystal. It is expressed as:

$$E_{CV} = H_{CV} - TS_{CV}$$

Here,  $E$  denotes the free energy and  $H$  the enthalpy. The index  $CV$  indicates the initial state (valence band,  $V$ ) and the final state (conduction band,  $C$ ). At room temperature the entropy  $S_{CV}$  is  $2.96k$ , with  $k$  being the Boltzmann constant:  $k = 86.17 \mu\text{eV/K}$ . Note that  $S_{CV}$  can, and indeed should, be split into two parts: one that accounts for the growth in entropy caused by the creation of a hole in the valence band, and another caused by the creation of an electron in the conduction band,  $S_{CV} = S_h + S_e$ . There is no a-priori reason why these two contributions should be equal, and experiments suggest they are not. Measurements of the entropy changes in transitions from the valence band to various conduction bands have yielded the estimate  $S_h = (3.6 \pm 1.0)S_e$ . At room temperature this translates into

$$S_h = 2.3k \pm 0.15k \quad S_e = 0.6k \pm 0.2k$$

and we shall use these values in the remainder of this work.

**Ionizing a defect:** We proceed by defining the term 'enthalpy of ionization'. Recall that at the absolute zero of temperature all electrons are in the valence band. As a result, all defects are then neutral—even the dopant atoms. Now, we call a defect a donor if it can be ionized by emitting an electron into the conduction band (eg by photoabsorption) and if the energy needed for this to happen is smaller than the bandgap enthalpy of  $1.20\text{eV}$ . The corresponding enthalpy is the enthalpy of ionization. Figure 45 shows two such impurities: a) Phosphorus, a Coulombic defect, is a shallow donor. Its enthalpy of ionization is  $H_I = 0.045\text{eV}$ , and this places its energy level close to the conduction band. b) Some deep donor whose enthalpy of ionization is  $H_I = 0.9\text{eV}$ . This places its energy level closer to the valence band than to the conduction band. Please note, donor levels are cited as energy differences to the conduction band.

An acceptor can at the absolute zero of temperature be ionized only by lifting an electron from the valence band to the defect, which is equivalent to saying that the defect emits a hole into the valence band. The right-hand side of figure 45 illustrates this: c) Boron is a Coulombic defect and with its ionization enthalpy of  $0.045\text{eV}$  its level is placed close to the valence band. d) Some deep acceptor with  $H_I = 0.8\text{eV}$ . Its level lies  $1.20\text{eV} - 0.80\text{eV} = 0.40\text{eV}$  below the conduction band. Thus, acceptor levels are stated as energy differences to the valence band.

By definition, we attribute the enthalpy  $H_I$  to the ionized defect while attributing zero energy to the neutral one.

To calculate the free energies of ionization, or, more generally, of any reaction involving changes in the charge state of a defect, one has to take into account the changes in entropy as well. Besides the contribution from electron and hole production there may be an ionization entropy due to changes in the atomic and electronic configuration of the defect when it becomes ionized. Let us denote that contribution by  $S_I$ . What, then, is the free energy needed to lift an electron from the valence band to an acceptor-like defect?

$$E_{TV} = H_I - T(S_I + S_h)$$

The index  $T$  will in general be used to refer to a defect. It stands for 'trap', a term used to distinguish a deep defect from a shallow (Coulombic) one. Next, the free energy needed to lift the electron from the ionized acceptor to the conduction band is

$$E_{CT} = H_{CV} - H_I - T(-S_I + S_e)$$

And the sum is, as it must be, the free energy of the bandgap:

$$E_{CT} + E_{TV} = H_{CV} - T(S_e + S_h) = E_{CV}$$

The ionization entropies are not usually known and are difficult to measure. Coulombic defects are the only exception to this. They are formed by group-III elements (shallow acceptors) and group-V elements (shallow donors) when they substitute silicon atoms in the lattice. In the donor case (eg phosphorus) four of the five electrons from the outer electron shell are engaged in bonds with the four neighbouring silicon atoms. The fifth electron, however, is free to move in the residual Coulomb field of the surplus core charge. Because of the high dielectric constant of silicon ( $\epsilon_r = 11.7$ ) the Coulomb potential is greatly reduced, and the hydrogen-like orbit of the fifth electron stretches out over a distance of many lattice spacings. This electron therefore very much resembles a conduction band electron, and when it is freed from the donor the entropy of the crystal cannot grow. The same argument applies for a shallow acceptor, only that the bound charge carrier is a hole. In the above formalism the ionization entropy  $S_I$  of a shallow donor is  $S_I = -S_e$ , and that of an acceptor is  $S_I = -S_h$ , thus that  $\Delta S$  of the reactions is zero, [Ve 76] [Ve 80].

### Equilibrium thermodynamics

This section aims to answer the question: What are, in equilibrium, the concentrations of electrons, holes, ionized and neutral defects, and how do they depend on the temperature?

Statistical mechanics offers an easy way to incorporate defects into the thermodynamics of the semiconductor. The most probable state of a macroscopic system is found by the following recipe: Calculate the number of indistinguishable ways to realize a macroscopic state. Multiply this by the Boltzmann factor containing the free energy of the state in question. Maximize this expression by varying the concentrations of the electrons, holes, etc.

By using the fact that no more than two electrons can at the same time occupy one state, we will derive the Fermi statistics for defect occupation. We shall treat the intrinsic and the doped semiconductor as well as the case when deep acceptors are present.

### The intrinsic semiconductor

We write the unnormalized probability to find the system in a state  $j$  as:

$$p_j = \left( \begin{array}{c} \text{number of} \\ \text{possibilities} \end{array} \right) \times \left( \begin{array}{c} \text{Boltzmann} \\ \text{factor} \end{array} \right)$$

$$= W_j \exp\left(-\frac{E_j}{kT}\right)$$

Instead of maximizing  $p_j$ , we will find it convenient to maximize

$$kT \log p_j = -(E_j - kT \log W_j)$$

Throughout the remainder of this section calligraphic letters shall denote absolute numbers and roman letters concentrations.  $N_C, N_V$  are the total density of states available to the unbound electrons and holes, respectively. These densities include all spin and band degeneracies.  $\mathcal{N}_C$  and  $\mathcal{N}_V$  are the corresponding absolute numbers. In the end, all results can be expressed by ratios of absolute numbers, which then can be replaced by ratios of concentrations.

In the intrinsic semiconductor the number of electrons must equal the number of holes because of charge neutrality and we shall use  $\mathcal{N} = \mathcal{N}_e = \mathcal{N}_h$ . The number of ways to distribute  $\mathcal{N}$  electrons and holes over the available places is:

$$W(\mathcal{N}) = \binom{\mathcal{N}_C}{\mathcal{N}} \cdot \binom{\mathcal{N}_V}{\mathcal{N}}$$

The total free energy of a state with  $\mathcal{N}$  e-h pairs is  $\mathcal{N} \cdot E_{CV}$  and the probability to find the system in that state becomes:

$$p(\mathcal{N}) = \binom{\mathcal{N}_C}{\mathcal{N}} \binom{\mathcal{N}_V}{\mathcal{N}} \exp\left(-\frac{\mathcal{N} \cdot E_{CV}}{kT}\right)$$

The most probable number  $\mathcal{N}$  is found from solving:

$$\frac{d}{d\mathcal{N}} \log p(\mathcal{N}) = 0$$

Using Stirling's formula, which is asymptotic for large  $\mathcal{N}$

$$\log \mathcal{N}! = \mathcal{N} \log \mathcal{N} - \mathcal{N} \quad \frac{d}{d\mathcal{N}} \log \mathcal{N}! = \log \mathcal{N}$$

we find the well known result:

$$\mathcal{N} = \sqrt{N_C N_V} \exp\left(-\frac{E_{CV}}{2kT}\right)$$

$$n = p = n_i = \sqrt{N_C N_V} \exp\left(-\frac{E_{CV}}{2kT}\right)$$

The electron and hole concentration ( $n, p$ ) equal the intrinsic concentration  $n_i$ .

### The doped semiconductor

We proceed in the same way as above, but introduce the total number of donors ( $D$ ) their concentration ( $N_D$ ) and the free energy needed to ionize them ( $E_D$ ). In addition, we assume all shallow donors to be ionized. Similar as before we obtain the probability to find  $\mathcal{N}_e$  electrons:

$$p(\mathcal{N}_e) = \binom{N_C}{\mathcal{N}_e} \binom{N_V}{\mathcal{N}_h} \exp\left(-\frac{\mathcal{N}_e(H_{CV} - TS_e) - \mathcal{N}_h TS_h + DE_D}{kT}\right)$$

The number of holes is restrained by the neutrality condition:  $\mathcal{N}_h = \mathcal{N}_e - D$ . Insert this in the above equation and solve

$$\frac{d}{d\mathcal{N}_e} \log p(\mathcal{N}_e) = 0$$

to find:

$$n(n - N_D) = N_C N_V \exp\left(-\frac{E_{CV}}{kT}\right)$$

$$np = n_i^2$$

In the result, again, absolute numbers have been replaced by concentrations.

For the non-degenerate semiconductor (doping  $< 10^{18} \text{ cm}^{-3}$ ) we can relate to each other the electron concentration ( $n$ ) and the chemical potential or Fermi level ( $E_F$ ):

$$n = N_C \exp\left(-\frac{E_C - E_F}{kT}\right)$$

or  $E_F = E_C + kT \log \frac{n}{N_C}$

The Fermi level that corresponds to the intrinsic electron concentration ( $n_i$ ) is called the intrinsic level  $E_i$ . It is located very close to the centre of the bandgap and shows little temperature dependence.

### A simple deep acceptor

Now, add a simple deep acceptor. The concentrations are  $A$  (total),  $A_0$  (neutral ones) and  $A_1$  (ionized ones). The corresponding absolute numbers are  $\mathcal{A}, \mathcal{A}_0$  and  $\mathcal{A}_1$ . Let  $g$  be the degeneracy factor of the ionized acceptor. There are now two independent variables, eg  $\mathcal{N}_e, \mathcal{A}_1$ , and the two constraints:

$$\mathcal{N}_h = \mathcal{N}_e + \mathcal{A}_1 - D$$

$$\mathcal{A} = \mathcal{A}_0 + \mathcal{A}_1$$

The joint probability for finding  $\mathcal{N}_e$  electrons and  $\mathcal{A}_1$  ionized acceptors is:

$$p(\mathcal{N}_e, \mathcal{A}_1) = \binom{N_C}{\mathcal{N}_e} \binom{N_V}{\mathcal{N}_h} g^{\mathcal{A}_1} \binom{\mathcal{A}}{\mathcal{A}_1} \times$$

$$\times \exp\left(-\frac{\mathcal{N}_e(H_{CV} - TS_e) - \mathcal{N}_h TS_h + DE_D + \mathcal{A}_1(H_1 - TS_1)}{kT}\right)$$

The most probable values of  $N_e$  and  $A_1$  are found from solving:

$$\begin{aligned}\frac{\partial}{\partial N_e} \log p(N_e, A_1) &= 0 \\ \frac{\partial}{\partial A_1} \log p(N_e, A_1) &= 0\end{aligned}$$

The first of these two equations again leads to  $np = n_i^2$ . Writing the free energy of ionization of the acceptor as

$$E_1 = H_1 - T(S_1 + S_h)$$

yields a simple result from the second equation:

$$\begin{aligned}\frac{A_1}{gA_0} &= \frac{N_V}{p} \exp\left(-\frac{E_1}{kT}\right) \\ &= \frac{n}{N_C} \exp\left(\frac{E_C - E_1}{kT}\right) \\ &= \exp\left(\frac{E_F - E_1}{kT}\right)\end{aligned}$$

The three right-hand sides are completely equivalent. They show how the occupancy ratio of the acceptor can be calculated if either the electron or the hole concentration is given, or if the Fermi level is known.

The last expression makes clear that the Fermi level is a dividing line: Electronic levels (far) above it remain empty while those below it are all filled. At the Fermi level itself, half of the available states are filled.

### A double acceptor

The last case we are going to consider is the presence of an acceptor that can trap up to two electrons. Let  $A_0, A_1, A_2$  be the concentration of neutral, singly and doubly ionized acceptors, respectively. The degeneracy ratios are  $g_1, g_2$ , and calligraphic letters again denote absolute numbers. We now have three independent variables and two constraints:

$$\begin{aligned}N_h &= N_e + A_1 + 2A_2 - D \\ A &= A_0 + A_1 + A_2\end{aligned}$$

The probability we seek to maximize is:

$$\begin{aligned}p(N_e, A_1, A_2) &= \binom{N_C}{N_e} \binom{N_V}{N_h} g_1^{A_1} g_2^{A_2} \frac{A!}{A_0! A_1! A_2!} \times \\ &\times \exp\left(-\frac{N_e(H_{CV} - TS_e) - N_h TS_h - DE_D - A_1(H_1 - TS_1) - A_2(H_2 - TS_2)}{kT}\right)\end{aligned}$$

Setting  $E_2 = H_2 - T(S_2 + 2S_h)$  we obtain the equilibrium values in the same way as before:

$$\begin{aligned}np &= n_i^2 \\ \frac{A_1}{g_1 A_0} &= \frac{N_V}{p} \exp\left(-\frac{E_1}{kT}\right)\end{aligned}$$

$$\begin{aligned}\frac{A_2}{g_2 A_0} &= \left(\frac{N_V}{p}\right)^2 \exp\left(-\frac{E_2}{kT}\right) \\ &= \left(\frac{n}{N_C}\right)^2 \exp\left(\frac{2E_C - E_2}{kT}\right) \\ &= \exp\left(\frac{2E_F - E_2}{kT}\right)\end{aligned}$$

Again, three equivalent expressions are given for the third equation. Note that  $H_2$  and  $E_2$  are the sum energies of both electrons trapped. For the divacancy, for example, the enthalpy to ionize it is  $H_1 = 0.78\text{eV}$ . To lift the second electron from the valence band to the defect  $0.97\text{eV}$  are needed. The quantity  $H_2$  becomes  $H_2 = 0.78\text{eV} + 0.97\text{eV} = 1.75\text{eV}$ . This notation is in accordance with that of Landsberg, [La80], but no derivation was given there.

### Fermi level and trap occupancy v temperature

To actually compute the Fermi level and the occupancy of the traps one has to insert the above found relations into the equation of charge neutrality and solve this for the Fermi level ( $E_F$ ). All concentrations can be computed once  $E_F$  is known. We shall not embark on doing model calculations, but only discuss the values the Fermi level can attain in the presence of substantial concentrations of defects.

At high temperatures the Fermi level is always at the bandgap centre, because the high intrinsic carrier concentration ( $n_i$ ) then renders defects, impurities, and doping unimportant. When the temperature is lowered  $n_i$  will decrease, eventually becoming negligible compared to the doping. The semiconductor is said to be extrinsic now. The electrons from the donor dopants tend to bind to any present electron acceptor, the more lasting the deeper the temperature drops. Just how many of them will be trapped and the exact position of the Fermi level, depends on type and concentration of the deep acceptor.

We begin with a deep acceptor that has one level above the bandgap centre. If its concentration is less than the ( $n$ -type) doping, all acceptors can at low temperatures capture an electron. There remains an electron concentration of  $n = N_D - A$ . From  $E_F = E_C + kT \log(n/N_C)$  we see that  $E_F$  approaches the conduction band as the temperature goes down. So, on lowering the temperature, the Fermi level sweeps through the entire upper half of the bandgap.

When the concentration of acceptor-like traps exceeds the doping this is no longer true. From the condition of charge neutrality it is clear that the concentration of trapped electrons cannot exceed that of the  $n$ -type dopants ( $N_D$ ), which provide the electrons. So, the concentration of ionized acceptors ( $A_1$ ) will at low temperatures approach  $N_D$ . And the Fermi level will be pinned at the acceptor level:

$$E_F = E_1 + kT \log\left(\frac{A_1}{gA_0}\right) \rightarrow E_1 + kT \log\left(\frac{N_D}{g(A - N_D)}\right)$$

No other defect level, located above the acceptor level, can then be occupied.

One obtains similar results for a double-acceptor trap. For concentrations less than half the doping, all acceptors will be doubly ionized at sufficiently low temperatures. Due to the presence of a residual free-electron concentration the Fermi level will rise up to the conduction band. If  $N_D/2 < A < N_D$  all acceptors can trap at least one electron. The remaining electrons will also be trapped producing some doubly-ionized acceptors. As a result, the ratio  $A_2/A_1$  will

approach a non-vanishing value at low temperatures, and the Fermi level will be pinned at the second acceptor level.

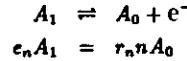
Similarly, if  $A > N_D$  not all acceptors can trap an electron, and the Fermi level will be pinned at the first acceptor level.

Summing up, we see that the position of the Fermi level crucially depends on the defect concentration. High-enough concentrations will limit the Fermi level to an intermediate position in the bandgap, preventing it from rising all the way up to the conduction band. In the extreme case, any defect level higher than the lowest acceptor level of the most abundant defect remains empty at any temperature.

#### 6.1.4 Non-equilibrium but stationary processes

If we want to answer the question about trap occupancy in the field zone of a reversely biased diode, equilibrium thermodynamics is of no use because the field zone, which contains no mobile charge carriers but a strong electric field, is certainly not in equilibrium.

Non-equilibrium processes are treated by introducing reaction rates for both directions as it is suggested by the stoichiometry of the reaction equation:



The two coefficients are called the electron emission coefficient ( $e_n$ ) and the electron recombination coefficients ( $r_n$ ). In the simplest models, they are expected to be independent of the carrier concentrations. But they do depend on the temperature and may depend on other extrinsic parameters like eg the electric field. Their ratio can be determined from the observation that the two reaction rates must be equal in thermal equilibrium.

$$\begin{aligned} \frac{e_n}{r_n} &= n \frac{A_0}{A_1} = \frac{1}{g_1} n \exp\left(\frac{E_1 - E_F}{kT}\right) \\ &= \frac{1}{g_1} N_C \exp\left(\frac{E_1 - E_C}{kT}\right) \end{aligned}$$

In the same way, a relation between the hole emission and recombination coefficient is found:

$$\begin{aligned} A_0 &\rightleftharpoons A_1 + h^+ \\ e_p A_0 &= r_p p A_1 \\ \frac{e_p}{r_p} &= p \frac{A_1}{A_0} = g_1 N_V \exp\left(\frac{E_V - E_1}{kT}\right) \end{aligned}$$

The usual approach to determine  $e_n$  and  $r_n$  separately is to consider the temperature dependence of the recombination coefficients. Figure 46 shows two possible variations of the potential energy of the electron as it approaches the trap. In the first case (l.h.s.) the incoming electron is not hindered to fall down into the potential well of the trap located at  $x = 0$ . The recombination coefficient is written as

$$r_n = v_{th} \cdot \sigma_n$$

where  $v_{th} = \sqrt{3kT/m}$  denotes the electron thermal velocity varying as  $T^{1/2}$  and  $\sigma$  is the trapping cross-section which is assumed to be a constant. The electron emission rate becomes:

$$e_n = \sigma_n v_{th} \cdot \frac{1}{g_1} N_C \exp\left(-\frac{E_{CT}}{kT}\right)$$

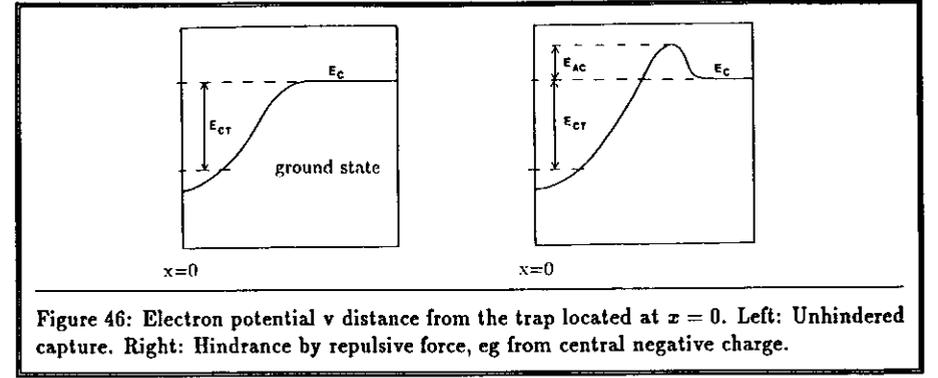


Figure 46: Electron potential  $v$  distance from the trap located at  $x = 0$ . Left: Unhindered capture. Right: Hindrance by repulsive force, eg from central negative charge.

$$= \frac{1}{\tau_0} \cdot \left(\frac{T}{300 \text{ K}}\right)^2 \exp\left(-\frac{H_{CT}}{kT}\right)$$

Here  $E_{CT}$  is short for  $E_C - E_1$ . In the last line all pre-exponential factors including the entropy term  $\exp(S_{CT}/k)$  have been absorbed into  $\tau_0$ . The temperature dependence of the pre-exponential factors has been factored out from  $\tau_0$  making it a constant. A measurement of  $e_n$  v temperature yields  $\tau_0$  and  $H_{CT}$ . One usually plots  $\log(e_n/T^2)$  v  $1/kT$  which gives a straight line. This is called an Arrhenius plot, and we have already encountered it in chapter 5.

The other possibility shown in figure refn is that the electron has to overcome some repulsive force before it can be trapped. The expected temperature dependence of  $r_n$  is then:

$$r_n = v_{th} \sigma_n \cdot \exp\left(-\frac{E_{AC}}{kT}\right)$$

That of  $e_n$  becomes:

$$e_n = \frac{1}{\tau_0} \cdot \left(\frac{T}{300 \text{ K}}\right)^2 \exp\left(-\frac{H_{CT} + E_{AC}}{kT}\right)$$

An effect like this has been observed for example for the (-/=)-level of the divacancy, [Br 82]. The incoming electron feels the repulsive force of the central negative charge of the  $VV^-$  prior to being trapped.

By the same reasoning the (unhindered) hole emission rate  $e_p$  becomes:

$$\begin{aligned} e_p &= \sigma_p v_{th} \cdot g_1 N_V \exp\left(-\frac{E_{TV}}{kT}\right) \\ &= \frac{1}{\tau_0} \cdot \left(\frac{T}{300 \text{ K}}\right)^2 \exp\left(-\frac{H_{TV}}{kT}\right) \end{aligned}$$

We are now prepared to calculate trap occupancies in the field zone. Under stationary conditions the time-averaged charge of a trap remains constant. The net emission of electrons must equal that of the holes because of charge conservation. There is no recombination taking place because the field zone is practically free of mobile carriers. Thus, we get the stationarity condition for an acceptor trap

$$e_n A_1 = e_p A_0$$

from which it follows that:

$$\frac{A_1}{A_0} = \frac{e_p}{e_n} = g_1^2 \frac{\sigma_p N_V}{\sigma_n N_C} \exp\left(\frac{E_{CT} - E_{TV}}{kT}\right)$$

The factor preceding the exponential is a constant of the order  $\mathcal{O}(1)$ . Whether the acceptor is ionized or not depends chiefly on its ionization energy ( $E_{TV}$ ). If it is less than the intrinsic energy  $E_i$  by at least a few  $kT$  the acceptor is ionized, regardless of the temperature. If it is bigger than  $E_i$  by a few  $kT$  the acceptor remains neutral at all temperatures.

This is true in general. Any trap level located below the intrinsic level is occupied while those above remain empty. The reason for this lies with the dynamics of carrier emission. Take, eg, an acceptor level in the upper half of the bandgap. When it is uncharged it takes a long time until an electron is lifted from the valence band to the defect from where it is almost immediately transferred to the conduction band. Therefore, the defect is unoccupied for most of the time.

### Dark current

We can use the above relations to calculate the dark current generated by a trap. From

$$\frac{A_1}{A_0} = \frac{e_p}{e_n} \quad \text{and} \quad A_0 + A_1 = A$$

we find the trap occupation:

$$\frac{A_0}{A} = \frac{e_n}{e_n + e_p} \quad \frac{A_1}{A} = \frac{e_p}{e_n + e_p}$$

And from that the electron-hole generation rate follows:

$$G = G_n = G_p = A \frac{e_n e_p}{e_n + e_p} = A \left( \frac{1}{e_n} + \frac{1}{e_p} \right)^{-1}$$

The dark current per unit volume is

$$\frac{I}{\text{volume}} = qG$$

where  $q$  denotes the elementary charge.

We see from this that the electron-hole pair-creation is governed by the slower process. If the trap level is located above  $E_i$  then  $e_p \ll e_n$ , and the dark current is proportional to  $e_p$ ; were it below,  $I$  would be proportional to  $e_n$ .

To get an idea about the orders of magnitude involved, let us calculate  $G$  as a function of the trap level. Let us set the capture cross-sections to  $\sigma_n = \sigma_p = (5\text{\AA})^2$  which corresponds to the typical size of a point defect. We chose  $T = 300\text{ K}$ , and the other values needed can be found in the appendix A. The side figure shows a plot of the electron-hole pair-generation rate (at 300K)  $v$  the position of the trap level. Depending, of course, on the capture cross-sections, the pair generation rate per trap attains  $100\text{s}^{-1}$  at its maximum. A dark current of  $1\mu\text{A}$  would require the presence of  $6 \cdot 10^{10}$  of these traps.

The temperature dependence of the dark current is governed by that of  $e_n$  and  $e_p$ . For an acceptor level in the upper half of the bandgap we have:

$$I \propto T^2 \exp\left(-\frac{H_{TV}}{kT}\right)$$

That is,  $I(T)$  should indeed be rectified by an Arrhenius plot as was experimentally verified in chapter 5. The activation energy in this case turns out to be the ionization enthalpy  $H_{TV}$ .

### 6.1.5 Non-stationary processes

In the preceding sections it was shown that in the undepleted part of the diode, which is in thermal equilibrium, any trap level below the Fermi level will be occupied. Furthermore, the Fermi level can be swept through the upper half of the bandgap by varying the sample temperature. In the field zone, which is not in equilibrium, any trap level above  $E_i$  will be empty. So, trap levels between the intrinsic level and the Fermi level are occupied in the base of the diode but empty in the field zone. We can thus produce charged traps in the field zone by suddenly pulsing the diode into depletion; cf figure 47. Now, we observe how the trap occupation there attains the stationary state. To be specific, we again study an acceptor-like defect. The rate of change of ionized acceptors is:

$$\frac{d}{dt} A_1(t) = -e_n A_1(t) + e_p A_0(t)$$

The limit value at infinity is:

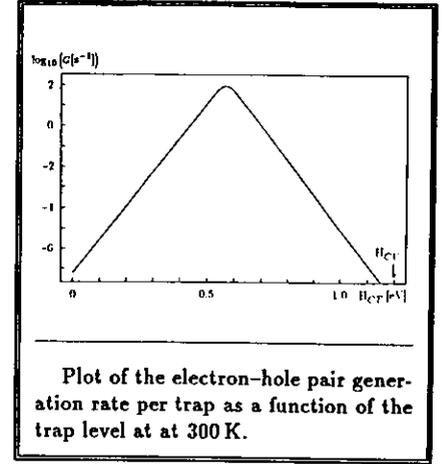
$$A_1(\infty) = A \frac{e_p}{e_n + e_p}$$

We have  $A = A_0 + A_1$  at any time, and the differential equation is solved by:

$$A_1(t) = [A_1(0) - A_1(\infty)] e^{-(e_n + e_p)t} + A_1(\infty)$$

The electron and hole generation rates now are no longer equal but are given by:

$$\begin{aligned} G_n &= e_n A_1(t) = e_n [A_1(0) - A_1(\infty)] e^{-(e_n + e_p)t} + e_n A_1(\infty) \\ G_p &= e_p A_0(t) = -e_p [A_1(0) - A_1(\infty)] e^{-(e_n + e_p)t} + e_p A_1(\infty) \end{aligned}$$



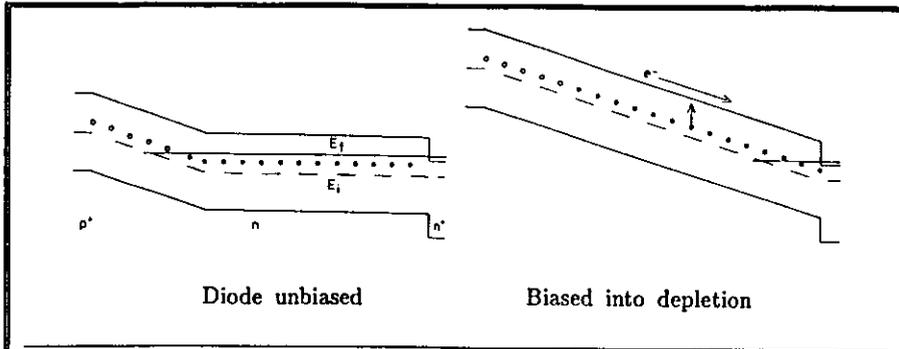


Figure 47: Deep defects with an energy level between the intrinsic and the Fermi level can be produced in a non-stationary state by pulsing a diode into depletion. Left: Idle state, filled dots denote traps having captured an electron. Right: When pulsed into depletion, occupied traps in the field zone will discharge, and the electrons from this process cause an electric current to flow.

If the traps are distributed uniformly over the diode volume the electrical current becomes:

$$\begin{aligned} \frac{I}{\text{volume}} &= \frac{1}{2}q(G_n + G_p) \\ &= \frac{1}{2}q(e_n A_1(\infty) + e_p A_0(\infty)) + \frac{1}{2}q(e_n - e_p)[A_1(0) - A_1(\infty)]e^{-(e_n + e_p)t} \\ &= \frac{I(\infty) + \Delta I(t)}{\text{volume}} \end{aligned}$$

Besides the steady-state dark current one observes a current transient. If the trap level is at least a few  $kT$  above  $E_i$  then  $e_n \gg e_p$  and the current transient simplifies to:

$$\frac{\Delta I(t)}{\text{volume}} = \frac{1}{2}q e_n A_1(0) e^{-e_n t}$$

The transient embodies two pieces of information: the electron emission rate  $e_n$  and the total number of discharging traps,  $A_1$ , from the integral of the transient over the time. This observation will, in the next section, lead us to an experimental technique to study trap discharging based on recording the concomitant transient currents.

We should on this occasion, however, warn the reader that the above description of the non-stationary process may well be oversimplified when compared to a real-life situation. The underlying, and indeed restrictive, assumption is that the carrier emission rates are the same for all traps involved.

Extending the model to the case where two different types of traps with different energy levels are present is fairly simple. The steady-state dark current (originating from the field zone) takes two distributions, one from each kind of trap. Similarly the transient part of the current will be the sum of two exponentials in time.

But even if only one kind of trap is being present the current transients may be more complex functions of time. If the carrier emission rates depend on an extrinsic parameter whose value

varies over the field zone (eg the electric field), then the current transient is a superposition of infinitely many exponentials and is itself not exponential.

Moreover, if the traps are produced in close proximity to each other (in clusters) the individual electron wave functions may partly overlap thereby influencing each others energy levels. The emission rates must then depend on the concentration of charged traps in a cluster and the above differential equation will become nonlinear. Its solutions will in general not be simple exponentials.

In the course of analysing current transients originating from non-stationary carrier emission, we should therefore not be preoccupied with finding exponential transients but preserve an open mind and seek the most concise description of the obtained data.

## 6.2 Current-transient recording

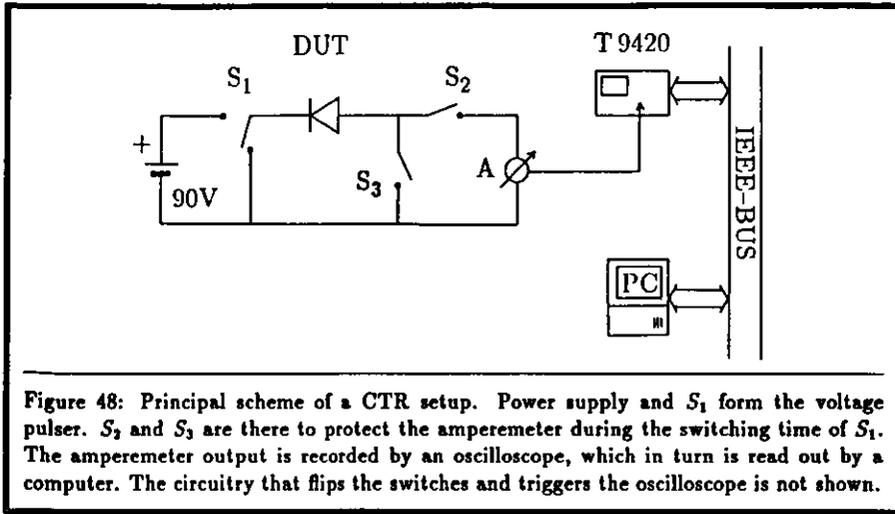
If by any means a defect in the field zone of a diode is prepared in a non-stationary charge state it will return to stationary conditions by electron or hole emission. The emitted carriers are collected by the electric field and cause a current to flow across the diode terminals. The induced current consists of the steady-state reverse current plus a current transient. The transient embodies information about the electron emission time and the total number of emitted carriers in a most straightforward manner. From the theory developed in the previous section, one expects, in the most simple case, an exponential transient. If more than one type of defect is involved, the transient may be a linear superposition of exponentials. The total charge contained in each component is proportional to the number of defects of each type that have discharged. By measuring the temperature dependence of the emission rates the activation energies are found. If the defect concentration is not too big, the position of the Fermi level can be calculated accurately, at any temperature. It is then possible to measure the occupancy of a defect level  $v$  the Fermi level. With some corrections applied for the degeneracies this yields the free energy. A combination of the enthalpy and free-energy data allows to estimate the entropy change associated with discharging the defect.

Up to now, no technique for defect characterization using the current transients associated with defect discharging has been described in literature. The following sections give the necessary details on the experimentation and the analysis as well as experimental results.

### 6.2.1 Sample preparation

Current-transient recording (to be abbreviated as CTR in the following) can be applied to study defects in any semiconductor material provided it is possible to manufacture a high-quality diode on it. The rectifying junction should be as abrupt as possible and strongly asymmetric. For the standard silicon detectors made from  $n$ -type material this calls for a  $p^+n$  junction manufactured preferably by ion implantation. The  $p^+$  concentration should exceed the  $n$ -type doping of the bulk by several orders of magnitude. This ensures that the depletion zone of the diode only extends into the bulk, where the defects of interest are located. Alternatively, a Schottky junction also is a workable solution.

When such a structure is reversely biased the field zone expands deeper into the bulk. CTR relies very much on the assumption that the width of the field zone remains constant once the applied bias voltage is steady. But discharging defects in the field zone change the space charge density, and the width of the field zone readjusts accordingly, which produces an additional current. The only way around this, is to make the reverse voltage step bigger than what is



needed to drive the diode into full depletion, ie, making the field zone hit the backside contact. Of course, the diode has to withstand this voltage.

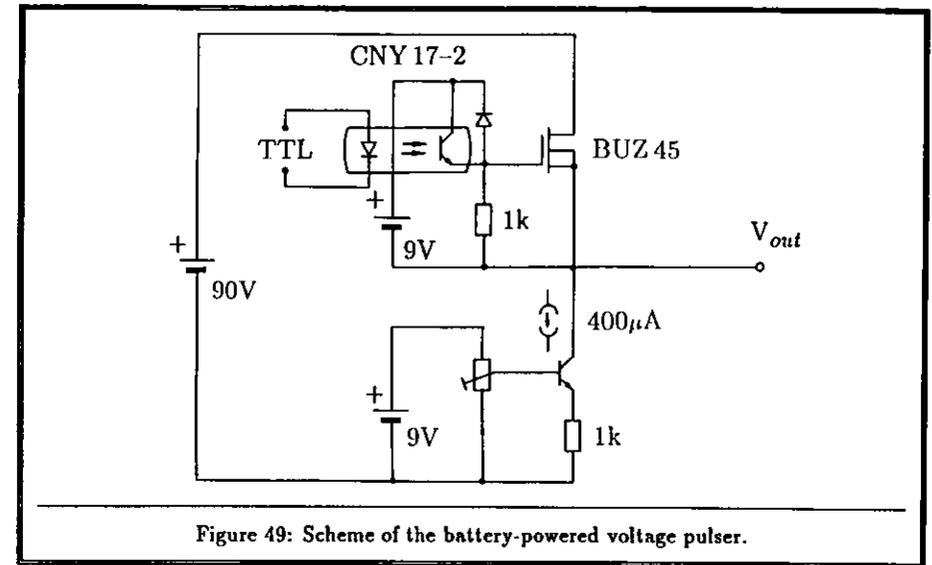
Some precaution may be necessary to avoid excessive minority carrier injection from the n-side contact; eg from surface defects. The best contact in this sense is a homotype low-high junction ( $nn^+$ ) manufactured by diffusion or ion implantation; cf chapter 3.

All diodes used in the experiments described here are  $p^+nn^+$  diodes. Their n-type bulk has a thickness of  $270\mu\text{m}$  while the highly doped contact regions on each side have a thickness of less than one micron. The surface areas are  $9.0\text{cm}^2$  and the volumes are  $0.24\text{cm}^3$ . They are fully depleted at  $\sim 50\text{V}$ , and the reverse voltage step was  $90\text{V}$  to ensure full depletion under all circumstances.

For a demonstration experiment, a simple version of a cooling system has been set up. The diode is mounted on a sample-holder made of aluminium to which it is thermally well-coupled. On the backside of the holder, adjacent to the diode, there is a standard DIN 4780 Pt-100 thermocouple. Its resistance is measured with a Keithley 190A multimeter using a four-wire technique (two current-leading wires and two sense wires). A few centimeters away, there is a resistive heater to warm the sample holder. The module is installed in a thermally insulating styrofoam box. The cooling is achieved by letting some liquid nitrogen evaporate inside the box. Beginning at  $-50^\circ\text{C}$  the warm-up rate is  $1.0\text{K}/\text{min}$  becoming less at higher temperatures. Comparing measurements (eg dark current) in a cool-down heat-up cycle, the accuracy of the temperature measurement was found to be better than  $0.5\text{K}$ . Temperatures above room temperature are achieved using the heater.

### 6.2.2 The CTR equipment

Figure 48 shows the general setup of a CTR experiment. In the idle state the power switch  $S_1$  is connected to ground, and the diode is discharged. When the device is triggered,  $S_1$  connects the diode to the power supply charging it abruptly to  $90\text{V}$ . Since it has a big capacitance, a large displacement current flows during the switch-on phase. This current is shunted to ground



via  $S_3$ . Once it has subsided  $S_3$  is opened and  $S_2$  in turn is closed, connecting the diode to a sensitive but fast amperemeter. Any current from discharging defects is now converted into a voltage. This voltage is recorded by a digital oscilloscope which was triggered when  $S_1$  was flipped. The CTR equipment is then switched back to idle while the computer (an ATARI 1024 ST) reads the oscilloscope. Once the computer has finished, the setup is ready for the next shot.

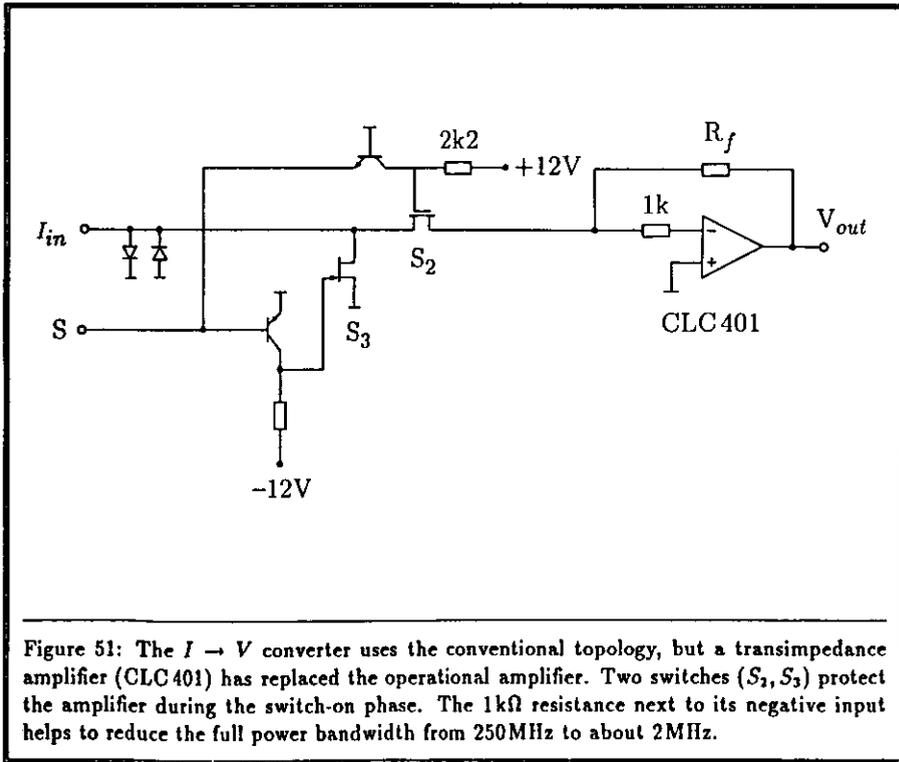
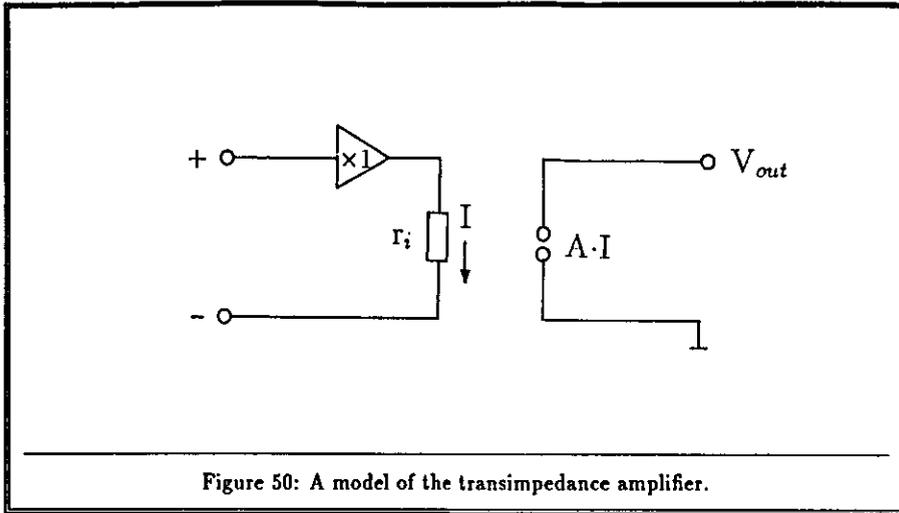
### The voltage pulser

The power supply is a  $90\text{V}$  battery made of ten  $9\text{V}$  Mono cells connected in series. It features very low high-frequency noise, and there is no low-frequency ripple from the mains. The output resistance is  $15\Omega$ , which is small enough.

A Siemens power MOS-FET (BUZ45,  $r_{DS,on} < 1\Omega$ ) in connection with an optocoupler provides a TTL-driven power switch. Its rise time (10% to 90%) is  $200\text{ns}$ , and the switch-on delay is  $700\text{ns}$ . In the off-state there is a leakage current of  $50\mu\text{A}$  flowing through the MOS-FET. This is sunk by a  $400\mu\text{A}$  current source. In the off-state the voltage supplied to the diode is less than  $50\text{mV}$ ; in the on-state it is greater than  $90\text{V}$ . Figure 49 shows the scheme.

### The amperemeter

At the heart of the CTR system there is a high-performance current-to-voltage converter. It is based on an integrated transimpedance amplifier, a new type of amplifier which has been commercially available in Europe since 1989. The operating principle is illustrated in figure 50. Of its two inputs one (+) has a high input impedance (about  $200\text{k}\Omega$ ) and is followed by a  $\times 1$  amplifier acting as an impedance converter. The other input has an impedance of only  $20\Omega$ . The current flowing out of the negative input is given by  $(V_+ - V_-)/r_i$ . A current-controlled



voltage source converts the input current into an output voltage with a high transimpedance  $A$  of nearly  $1\text{M}\Omega$ . A current-to-voltage converter using this device has been built using the standard topology, cf figure 51. The transimpedance of the converter results from paralleling  $A$  and the feedback resistance.

The salient feature of the new op-amp exploited in this application is that up to  $50\text{MHz}$  it produces no phase shift ( $< 0.2^\circ$ ). So, with the  $90^\circ$  phase shift of the RC low-pass formed by the feedback resistance and the diode capacitance the device still has a phase margin of  $90^\circ$ . It is thus stable with any capacitive load at its inputs.

A conventional operational amplifier has a rather different behaviour. At frequencies beyond a few hundred Hertz its phase shift is already  $90^\circ$ , increasing further towards the unity-gain frequency. An  $I \rightarrow V$  converter built with such an op-amp would inevitably oscillate when connected to a capacitive input load.

The open-loop input impedance of the CLC 401 is always  $20\Omega$  regardless of the frequency. Introducing a feedback reduces the impedance even further. The op-amp has been designed to drive  $50\Omega$  lines, so it can be connected to the oscilloscope via a standard  $50\Omega$  coaxial cable without problems. The extremely high full power bandwidth of  $250\text{MHz}$  had to be reduced by external means to avoid excessive noise. The details of the circuit are found in figure 51. The converter has been equipped with different feedback resistances to cover the range of sensitivities needed. The side table summarizes its performance in all three ranges.

$R_f =$	$1\text{k}\Omega$	$10\text{k}\Omega$	$100\text{k}\Omega$
transimpedance	$1\text{k}\Omega$	$9.5\text{k}\Omega$	$70\text{k}\Omega$
$r_i$ @ $1\text{MHz}$	$13\Omega$	$50\Omega$	$310\Omega$
settling time $< 1\%$	$2\mu\text{s}$	$4\mu\text{s}$	$6\mu\text{s}$
nonlinearity			
$0 \dots 4V_{out}$	$< 0.1\%$	$< 0.1\%$	$< 1\%$
rms noise*	$20\text{nA}$	$110\text{nA}$	$1.9\mu\text{A}$

Performance of the current-to-voltage converter.  
\*) rms equivalent of noise-current input measured with a detector diode ( $C=350\text{pF}$ ) connected to the ammeter input.

### Recording device

The digital oscilloscope, a LeCroy T9420, has been hooked up to the automatic test stand described in chapter 4 of this work. It is read out by the ATARI, and the data are stored on a hard disk for later analysis.

It features two channels, each equipped with a  $100\text{MSPS}$  8-bit flash ADC. Up to  $50,000$  points are stored per trace. For the limited computing power and storage facilities of the ATARI these are by far too many. Instead only a thousand points per trace are read; but one is free to choose which points to read. The oscilloscope is able to display parts of the signal even before the trigger arrived. This pre-trigger time was set to  $10\%$  of the full time scale. Reading the first  $100$  points of the displayed trace allows to measure the baseline before the CTR device was triggered. The time base was set such that on the last  $20\%$  of the trace the oscilloscope reading changed by less than the r.m.s. amplitude of the noise or one ADC count, whichever was smallest. Reading the last  $100$  points of the displayed trace allows to measure the baseline after the transient is gone. The remaining  $800$  points cover the transient. Beginning at the trigger, they are distributed over the entire time span with exponentially increasing increments. This way one catches the short as well as the fast signal components with about the same amount of information.

### 6.2.3 Raw data analysis

Analysis of the transients starts with determining a model function that fits the observation at all temperatures. The decision of whether or not the current transients can be ascribed to the action of one type of crystal defect only has to be made right at the beginning of the analysis. Once this is settled, the exact values of the parameters being employed in the model are found by a least-squares fitting procedure. This has to take into account the finite resolution offered by an 8-bit ADC.

Next, one has to remember that the diode always has a field zone of finite width, even when it is unbiased. Therefore, the traps, the discharging of which is observed, are not spread out over the whole diode volume but only over a part of it. Since the charge a freed electron induces on the diode terminals depends on the location from where it was emitted one has to apply a correction for this.

#### Least-squares fit

Adjusting the model parameters to fit the individual transients is done by a least-squares fit. It shall be treated here briefly because the error analysis is most conveniently based on an approximation of the  $\chi^2$  function at its minimum.

Let the model function used to describe the transients be

$$f(t) = f(p_1, p_2, \dots, p_K; t)$$

depending explicitly on the  $K$  parameters  $p_k$ . One seeks to vary them in such a way as to minimize the function

$$F(p_1, p_2, \dots, p_K) = \sum_{n=1}^N (f(t_n) - y_n)^2$$

in which the sum extends over all data points. The least-squares function  $F$  can be expanded into a Taylor series around a starting point.

$$F = F_0 + \epsilon^T G + \frac{1}{2} \epsilon^T H \epsilon + \dots$$

The vector  $\epsilon$  is the vector of the relative change of the parameters:  $\epsilon_k = (\delta p_k)/p_k$ .  $G$  is the gradient and  $H$  the Hesse matrix. Denoting  $\partial_k = \partial/\partial p_k$  for short we have:

$$\begin{aligned} G_k &= p_k \partial_k F \\ &= 2p_k \sum_n (f_n - y_n) \partial_k f \end{aligned}$$

$$\begin{aligned} H_{kl} &= \partial_k \partial_l F \\ &= 2p_k p_l \left[ \sum_n \partial_k f \partial_l f_n + \sum_n (f_n - y_n) \partial_{kl}^2 f_n \right] \end{aligned}$$

In the calculation of  $H_{kl}$  the second sum is neglected since it should on average vanish at the minimum of  $F$ .

Newton's method of finding the extremum iteratively is:

$$\begin{aligned} \epsilon^{(k+1)} &= \epsilon^{(k)} - H^{-1} G \\ p_k^{(k+1)} &= p_k^{(k)} (1 + \epsilon_k^{(k+1)}) \end{aligned}$$

With the new parameter vector, the new gradient is calculated, and the Hessian is updated. There exists an a-priori estimate on the difference of the actual value of  $F$  to its minimal value.

$$\delta F = G^T (H^{-1})^T G = (H^{-1} G)^T G$$

The algorithm converges well, but can, at the beginning, be sped up considerably by performing a one-dimensional search for the minimum along the direction  $H^{-1}G$ .

To make this fit procedure safe and secure, two refinements are needed. First, the stored data are not from a continuous point set but from the discrete set of the 256 counts the flash ADC's offer. The theoretical curve  $f(t)$  must therefore be projected onto this grid prior to computing the least squares sum.

The second refinement is a technicality. Computing a Newton correction calls for an inversion of the Hesse matrix. The Hessian may, however, be ill-conditioned. This means to say that  $\text{cond}(H)$ , the ratio of the biggest eigenvalue to the smallest, may be very big. But this is also the error amplification to be expected if the matrix is explicitly inverted by some one-step algorithm, eg Gaussian elimination. For this reason, the numerically risky business of explicit matrix inversion has been avoided altogether.

Instead, using an algorithm of Zurmühl & Falk, [Zu 88], the Hessian is iteratively decomposed into a triple product involving its modal matrix ( $X$ ), whose columns are the normalized eigenvectors, and the diagonal matrix ( $D$ ) consisting of the eigenvalues:

$$H = X D X^T \quad H^{-1} = X D^{-1} X^T$$

All its eigenvalues are distinct, and the eigenvectors are mutually orthogonal. The matrix product  $X^T X$  thus equals the unit matrix, and the above formula for  $H^{-1}$  can be verified by explicitly computing the products  $H H^{-1}$  and  $H^{-1} H$ .

The Newton step

$$H^{-1} G = X D^{-1} X^T G$$

is then computed from right to left without numerical problems.

Knowing the eigenvectors and eigenvalues is of further advantage as we shall see next.

#### Error analysis

If nothing was known about the interdependence of the fit parameters, determining the errors of the found parameters would be an ugly business. You would have to choose a certain parameter, say number one, vary it by some amount and then keep it fixed. Now, do a new fit allowing the other parameters to vary to either minimize the least squares sum or to maximize the likelihood. This way one computes the likelihood as a function of parameter number one under the condition that the other parameters are always readjusted to maximize the likelihood. The error interval of the parameters would be found from demanding that the likelihood falls to a certain percentage of its maximum value. This procedure would have to be repeated for every parameter and is clearly extremely time consuming, [Ea 82], [APS 92].

Matters simplify very much if the correlations among the fit parameters are known. And indeed the Hessian provides us with exactly that.

Using the definitions from above and letting  $\sigma_v$  be the r.m.s. noise of the data, for  $K$  fit parameters there holds:

$$\frac{\epsilon^T H \epsilon}{\sigma_v^2} = \left( \begin{array}{l} \text{asymptotically} \\ \chi^2(K) \text{ distributed} \end{array} \right)$$

and, besides an uninteresting constant prefactor, the likelihood function behaves around its maximum like

$$\mathcal{L} \propto \exp\left(-\frac{\epsilon^T H \epsilon}{2\sigma_v^2}\right)$$

ie like a  $K$ -dimensional Gaussian distribution, [Ea 82].

The errors along the eigendirections of the Hessian are independent of each other. The 'n-sigma' error along eigendirection number  $k$  with the eigenvalue  $d_k$  is:

$$\begin{aligned} \frac{\delta_k^2 d_k}{\sigma_v^2} &= n^2 \\ \delta_k &= \frac{n\sigma_v}{\sqrt{d_k}} \end{aligned}$$

The errors of the fit parameters then each take independent contributions from all eigendirections. Total errors are computed from summing the individual contributions quadratically:

$$\begin{aligned} X &= \text{column matrix of the eigenvectors} \\ \delta_k &= \text{error along eigendirection } k \\ \epsilon_k &= \text{error of the parameter } k \\ &= \sqrt{\sum_i (\delta_i X_{ki})^2} \end{aligned}$$

And this way of calculating the errors of the fit parameters is certainly preferable.

If the data were from a continuous point set the error analysis would be done with. But the binning of the data complicates things.

If the electronic noise is much larger than the ADC bin size then the binning is of no importance, and the choice for  $\sigma_v$  is the r.m.s. value of the noise. In the experiments described here, this situation was encountered at low temperatures where the decay times are long and the currents small. The other extreme was encountered at room temperature and above; here, the noise was smaller than the bin size, and  $\sigma_v$  was set to  $\Delta I$ . In general, the least-squares sum is a small multiple of the bin size squared, ie  $(\Delta I)^2$ .

## 6.3 Experimental findings

### 6.3.1 Raw data analysis

Earlier in this chapter, in section 6.1.5, it has been pointed out that there can be no a-priori knowledge on the 'correct' mathematical model that should be used to parametrize the data. The transients might be simple exponentials originating from one type of trap only. There may be two or more kinds of traps present and discharging, producing transients which are a superposition of simpler curves, possibly but not necessarily of exponential type. And, last not least, even the transients caused by a single kind of trap need not follow an exponential law. Not knowing what to expect, we must judge from the data.

A first inspection reveals that the time it takes for the currents to subside varies by several orders of magnitude as the temperature rises from the lowest value to its highest. Naturally, the question arises if there are common characteristics. To find out, turn to figure 52 in which four transients, obtained from one diode at very different temperatures, are plotted. In a process, the details of which are of no importance for the moment, they have been normalized. First, the dc-offsets, consisting of amplifier offset and steady-state dark current, have been subtracted. The individual scales of time and amplitude have then been multiplied by suitable numbers, and the four transients have then been superimposed to create this picture.

The aim of this section is to show from the raw data that all transients are of the same shape. To make this clear to see, figure 52 has been split in two parts. Alternatively the horizontal (time) or the vertical (current) axis has been divided into logarithmically equivalent increments. The whole measurement range is fairly well displayed in this manner.

As can be seen, the main portions of the individual transients fall onto the same general curve. The differences chiefly result from the limitations of the experiment. At very low temperatures, the charge carrier emission from the traps is slow, and the resultant electric current is small, but prevails for a long time. The electronic noise from the current-to-voltage converter then dominates the transient at big times, which is especially prominent when the current amplitude is drawn logarithmically. On the other hand, the transient is gone after a very short time at high temperatures. The oscillations that are visible in that case, at  $T = +24^\circ\text{C}$ , originate from the converter, whose ringing after switch-on has not yet subsided. At the very beginning, the curves start with a horizontal line. This happens because of the overflow the ADC experiences when the currents are too big.

Except for these artefacts, the transients are clearly all of the same shape, regardless of the temperature. It therefore makes sense to seek a suitable function of time and parametrize the individual transients with three quantities: an underlying dc-current and two normalization factors, one for the current amplitude and one for the time scale. The model pursued is of the following form:

$$I(t) = q\lambda f(\lambda t) + \text{offset}$$

The parameter  $\lambda$  is the quantity scaling the time axis, and the amplitude factor has been written as  $q \cdot \lambda$ . The, as yet unknown, function  $f(\lambda t)$  contains the information about the shape of the current transients. The total charge contained in it is

$$\begin{aligned} Q &= q \int_0^\infty \lambda f(\lambda t) dt \\ &= q \int_0^\infty f(x) dx \end{aligned}$$

with  $x = \lambda t$ .

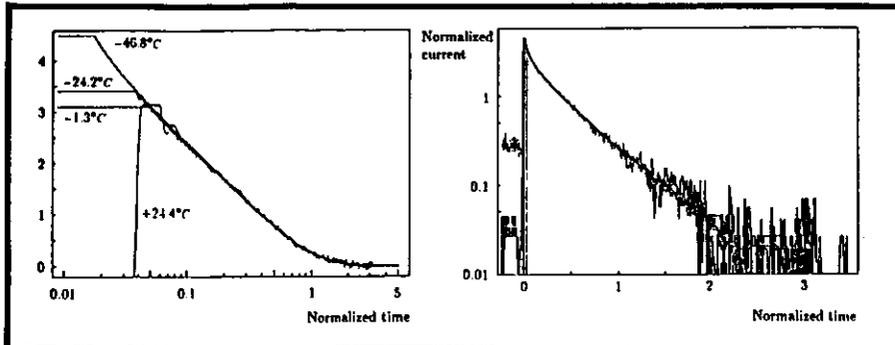


Figure 52: Four transients obtained from diode H154 at different temperatures. The model  $I(t) = q\lambda f(\lambda t) + \text{offset}$  has been fitted to the data. The time scales then were multiplied by  $\lambda$  and after subtracting the offset the currents were divided by  $q\lambda$ . The result of this normalization procedure is shown for four transients obtained over a temperature span of 70°C. On the left, the amplitude is shown against a logarithmic time scale. The transients are proportional to  $-\log(x)$  at  $x \ll 1$ . On the right, the amplitude is shown on a logarithmic scale against the time. At  $x > 1$  the transients tend to an exponential. Most important, it is clear to see that the shape of the transients does not change with the temperature.

We are left with the problem to determine  $f(\lambda t)$ . No attempt was made to find a theoretically justifiable function because there is virtually no literature on non-exponential carrier emission except for cases which do not seem to apply here (namely alloy broadening in III-V semiconductors and field-enhanced emission from a Coulombic defect, see later in this section). Instead, the goal was achieved by graphical inspection and guess work.

Primarily, the benefit in finding  $f(x)$  is not so much to be seen in its theoretical value, but in that it allows us to determine how the decay rate ( $\lambda$ ) and the charge ( $Q$ ) vary as a function of the sample temperature. Once a function has been adopted this can be done conveniently by least-squares fits.

From figure 52 it can be seen that the behaviour of  $f(x)$  at big values of  $x$  is exponential; ie,  $f(x) \sim \exp(-ax)$ . At small values of  $x$  we find  $f(x) \sim -\log(x)$ . One function (of perhaps many others) that behaves in this manner and fits the data well is  $f(x) = xK_1(x)K_0(x)$ . The  $K_0(x)$  and  $K_1(x)$  are Bessel functions of the purely imaginary argument, namely the MacDonald's function of order zero and one. The next box gives an account on their basic properties. Many further details can be found in [Wa40].

#### The fit procedure:

If the fitting is done by minimizing the simple least-squares function the decay rates are found with small statistical errors (about one percent) but the charge contained in the transients varies by as much as 10%. This can be attributed to the fact that the about 800 time points, at which the momentary current is measured, are not uniformly spaced. At the beginning, the time increment in between two points is as small as the scope setting allows. The increments

The basic properties of  $f(x) = xK_1(x)K_0(x)$  and its infinite integral.

$$K_0(x) = -\log(x/2) + \gamma + \mathcal{O}(x^2) \text{ for } x \rightarrow 0$$

$$K_0(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x} \left[ 1 + \mathcal{O}\left(\frac{1}{x}\right) \right] \text{ for } x \rightarrow \infty$$

Euler's constant is  $\gamma \approx 0.577 216$ .

$$xK_1(x) = 1 + \mathcal{O}(x^2) \text{ for } x \rightarrow 0$$

$$xK_1(x) \sim x\sqrt{\frac{\pi}{2x}} e^{-x} \left[ 1 + \mathcal{O}\left(\frac{1}{x}\right) \right] \text{ for } x \rightarrow \infty$$

The limiting expressions for  $f(x)$  are then:

$$f(x) = xK_1(x)K_0(x) = -\log(x/2) + \gamma + \mathcal{O}(x^2) \text{ for } x \rightarrow 0$$

$$f(x) = xK_1(x)K_0(x) \sim \frac{\pi}{2} e^{-2x} \left[ 1 + \mathcal{O}\left(\frac{1}{x}\right) \right] \text{ for } x \rightarrow \infty$$

The derivatives of  $K_0$  and  $K_1$  are

$$K_0'(x) = -K_1(x)$$

$$K_1'(x) = -\frac{K_1(x)}{x} - K_0(x)$$

The infinite integral is found from the integral representation of the product of any two  $K_\nu(x)K_\mu(x)$  by reversing the sequence of integration, which is allowed since the involved integrals converge absolutely.

$$\int_0^\infty xK_1(x)K_0(x)dx = \int_0^\infty \int_0^\infty 2x\cosh t \cdot K_1(2x\cosh t) dt dx$$

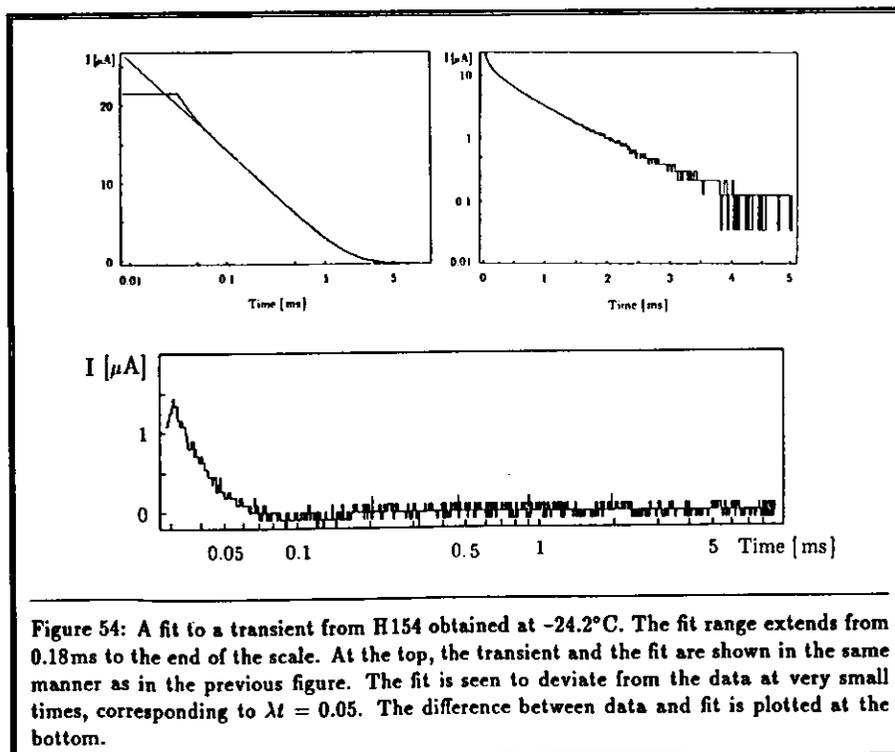
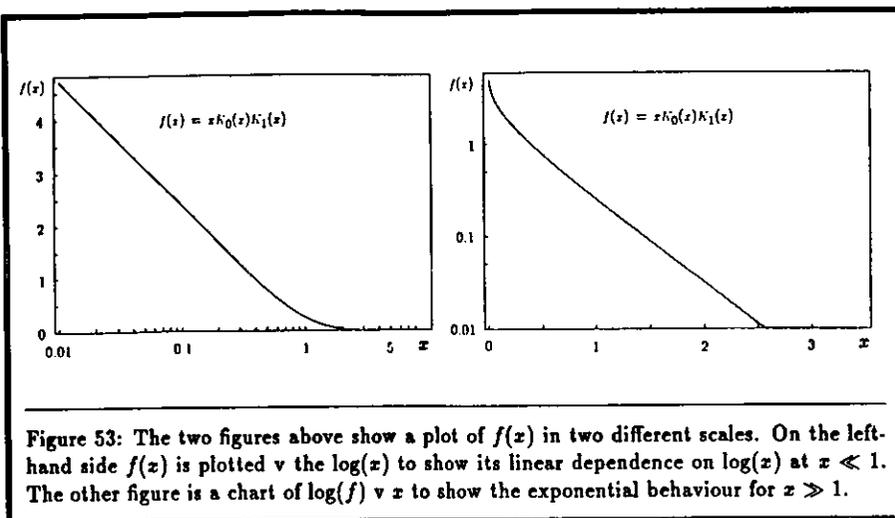
$$= \int_0^\infty uK_1(u)du \int_0^\infty \frac{dt}{\cosh t}$$

$$= \frac{\pi}{2} \cdot \frac{\pi}{4} = \frac{\pi^2}{8}$$

The integral representation of the product in the first line has been taken from [Wa40, paragraph 13.72]. The second line is arrived at by making the substitution

$$u = 2x\cosh t \quad dx = \frac{du}{2\cosh t}$$

The infinite integral over  $uK_1(u)$  is discussed in [Wa40, paragraph 13.21].



increases exponentially towards the end of the transient. There they are about 200 times bigger than at the beginning. In an unweighted least-squares fit, the fit function is 'drawn' through all available points in such a way that the average deviation of the function from the data is the same everywhere in the fit range. But a fixed deviation produces an error in the charge (ie the integral of the function) that is 200 times bigger at the end than it is at the beginning. As a fix, a weighting was introduced. Besides this, the fitting procedure also has to take into account the finite resolution offered by the eight-bit ADC. The function minimized in all cases was:

$$F = \frac{1}{t_{max} - t_{min}} \sum_i (\text{grid}(I(t_i) - y_i))^2 \cdot \Delta t_i$$

$$\text{grid}(x) := \Delta I \cdot \text{ceil}(x/\Delta I)$$

$$\text{ceil}(r) := \text{smallest integer greater or equal to } r$$

$$\Delta I := \text{current step corresponding to one ADC-count of the oscilloscope}$$

$$I(t_i) = q\lambda x_i K_1(x_i) K_0(x_i) + \text{offset}; \quad x_i := \lambda t_i$$

$$\Delta t_i = t_{i+1} - t_i$$

The summation in the first line is to be understood to extend over all data points in the fit range. The action of the function  $\text{grid}(x)$  is to project the difference between the calculated current  $I(t_i)$  and the measured current  $y_i$  on to the 'grid' of the 256 different current values that are provided by the eight-bit ADC of the oscilloscope.

The fit range for each fit was chosen according to the following rules. Its upper end always coincided with the end of the recorded transient, which occurred no earlier than after three times the inverse of  $\lambda$ . The lower end depended on the setting of the vertical scale of the oscilloscope and the performance of the current amplifier. The low end of the fit range was either

- the smallest time after which oscillations from the amplifier were smaller than five ADC counts,
- the smallest time after which the ADC of the oscilloscope was no longer in overflow due to the high current at the beginning,
- the smallest time after which the product  $\lambda t$  was bigger than 0.1,

whichever time was the biggest.

The reasons for the first two conditions are obvious. The third had to be introduced because at the smallest times the fit function deviates from the data in an unsystematic manner: In the neutron-irradiated diode (H137) the transients fall less fast at very small times, in the other diodes they fall faster. For most of the data sets the first two conditions were more stringent, and it was not possible to investigate the deviations with the necessary rigour. The systematic error in the charge introduced by neglecting the deviations is less than five percent. It therefore seems safe to disregard the behaviour of the transients at these very small times.

The modified least-squares function  $F$  is a weighted average of the square of the deviations of the fit from the data. At very low temperatures, below  $-30^\circ\text{C}$ , the electron emission process is slow, and the transient currents are very small. In these cases the electronic noise was much bigger than one ADC count. The square root of  $F$  then equals the r.m.s. value of the electronic noise as measured with a wideband r.m.s. voltmeter. At higher temperatures, where

the electron emission is fast and the currents big, the electronic noise eventually is smaller than the equivalent of one ADC count. In those cases the value of  $F$  ranged between  $0.25(\Delta I)^2$  and  $1.0(\Delta I)^2$  centering around  $0.4(\Delta I)^2$ .

As an example, a transient taken at  $-24.2^\circ\text{C}$  and the fit to it are shown in figure 54. The fit range extended from  $t = 0.18\text{ms}$  to the end of the recorded trace. The velocity parameter  $\lambda$  was found to be  $\lambda = 1.8\text{ms}$ . In the lower part of the figure, the deviation of the fit from the data is shown. Over the whole fit range, the deviation is zero or plus/minus one ADC count, with rare excursions. Fit and data diverge systematically only at times smaller than  $0.07\text{ms}$  which corresponds to  $\lambda t < 0.039$ .

### Results on the decay rate

In the theoretical section of this chapter it has been argued that the temperature dependence of the electron emission should obey an Arrhenius law. Written in a compact form with all expected temperature dependences explicitly stated it is

$$\tau := \frac{1}{e_n} = \tau_0 \left( \frac{T}{300\text{K}} \right)^{-2} \exp\left( \frac{H_{CT}}{kT} \right)$$

$$\log(\tau \cdot T^2) = \log\left( \frac{T^2}{e_n} \right) = \log((300\text{K})^2 \tau_0) + H_{CT} \cdot \frac{1}{kT}$$

in which  $H_{CT}$  is the enthalpy needed to lift the electron from the trap to the conduction band. This enthalpy is often called the activation energy. The second representation shows how to 'rectify' an Arrhenius law. Plot  $\log(T^2/e_n) \nu 1/kT$  and obtain a straight line. This is called an Arrhenius plot.

The observed transient are not of an exponential type, and it may seem questionable whether the parameter  $\lambda (= 1/\tau)$  should obey an Arrhenius law. Yet, it is a general expectation from physical chemistry that for any type of reaction, whether proceeding exponentially in time or not, the velocity parameter should indeed obey such a law. At least this is true, if during the reaction an energy barrier has to be overcome. The electron emission from a trap is no exception in this respect. The temperature dependence of the reaction velocity is chiefly governed by the exponential dependence on the inverse temperature. The structure of the prefactor is less clear, and in semiconductor physics there is only a comparatively small body of literature concerned with it; cf [Bö 87] for an overview. To stay in line with the theory of section 6.1.5 and common usage, the prefactor shall be assumed to vary as  $T^2$ .

We have denoted by  $e_n$  the electron emission rate of an exponential process. To avoid confusion, we denote the velocity parameter of the transients by  $\lambda$  and its inverse by  $\tau$ .

Now turn to fig 55; it is an Arrhenius plot of  $\tau$ . All data obtained from the diodes H137, H154, H155 and H156 have been pooled together in this figure. The straight line running through the points is a linear least-squares fit.

Since the asymptotic behaviour of the current is  $I \propto \exp(-2\lambda t)$  we state the result as follows:

$$\frac{1}{2\lambda} = \frac{\tau}{2} = 1.81\text{ps} \left( \frac{T}{300\text{K}} \right)^{-2} \exp\left( \frac{0.422\text{eV}}{kT} \right)$$

The lower part of fig. 55 shows the relative deviations of the measured  $\lambda$ 's from the above Arrhenius law. It is seen that the deviations are unsystematic except, perhaps, at temperatures below  $-50^\circ\text{C}$ , where the accuracy of the measurement is, however, only modest. Above  $-50^\circ\text{C}$  the r.m.s. deviation is about 5%.

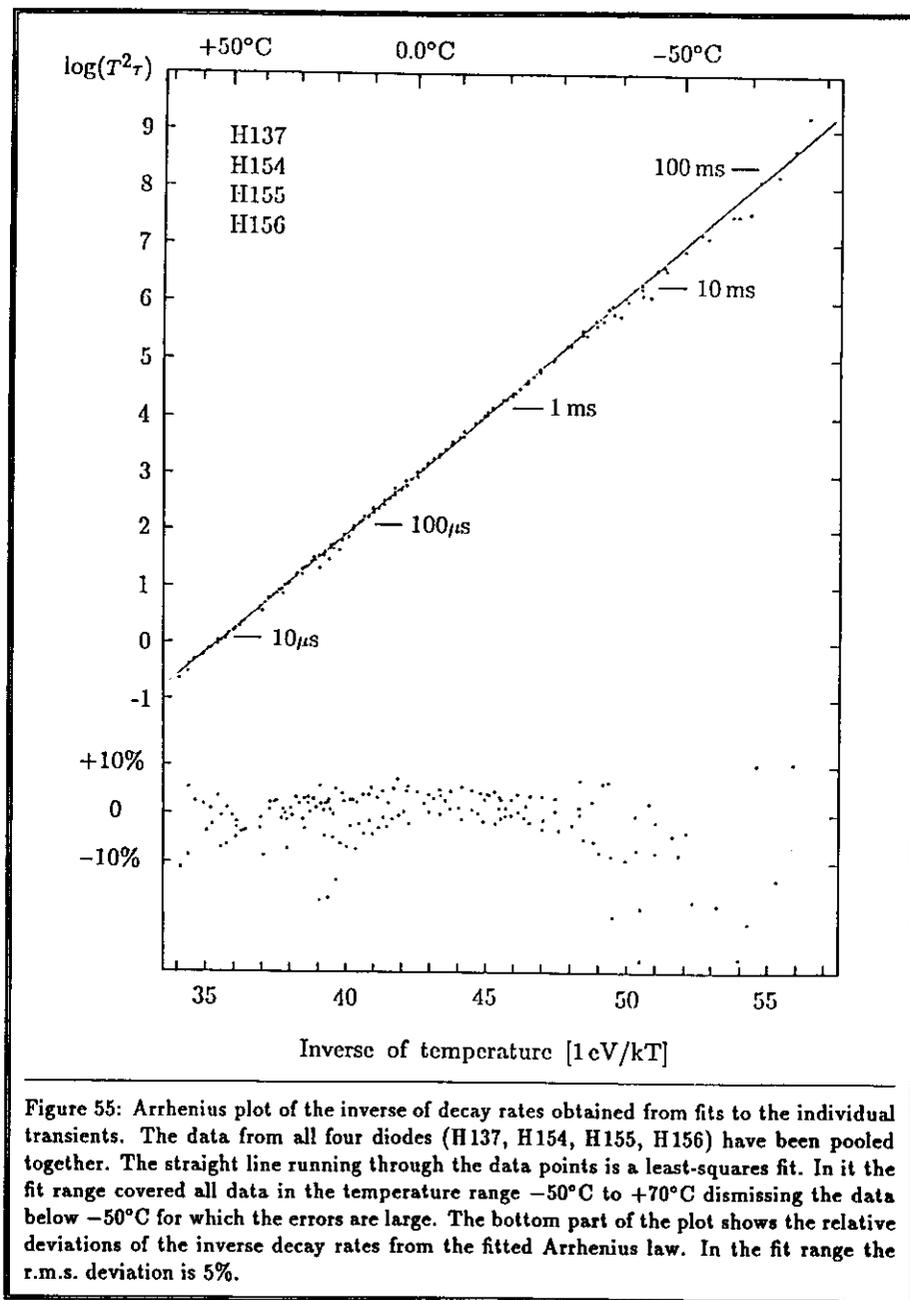


Figure 55: Arrhenius plot of the inverse of decay rates obtained from fits to the individual transients. The data from all four diodes (H137, H154, H155, H156) have been pooled together. The straight line running through the data points is a least-squares fit. In it the fit range covered all data in the temperature range  $-50^\circ\text{C}$  to  $+70^\circ\text{C}$  dismissing the data below  $-50^\circ\text{C}$  for which the errors are large. The bottom part of the plot shows the relative deviations of the inverse decay rates from the fitted Arrhenius law. In the fit range the r.m.s. deviation is 5%.

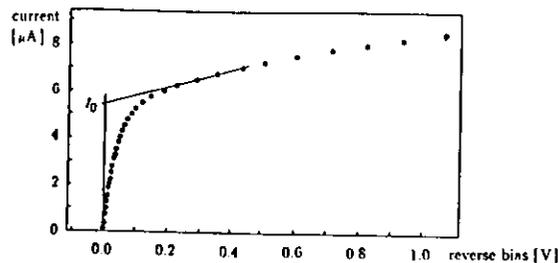


Figure 56: Current-voltage characteristic of the diode H154 (16kGy) at small voltages. Extrapolating to zero the dark current measured at voltages above 0.2V yields  $I_0$  from which the thickness of the zero-bias depletion zone is inferred.

Summing up, we can state:

- To an accuracy of 5% the non-exponential electron emission obeys an Arrhenius law. The temperature range covered extends from  $-50^\circ\text{C}$  to  $+70^\circ\text{C}$  over which  $\tau$  reduces from 30ms to less than 10 $\mu\text{s}$ .
- The activation energy is  $(0.4216 \pm 0.0007)\text{eV}$  (r.m.s. error).
- The prefactor  $\tau_0/2$  is  $(1.81 \pm 0.05) \cdot 10^{-12}\text{s}$ .

### The charge v temperature

The charge contained in each transient is plotted in figure 57 as a function of temperature for all four diodes, H137, H154, H155, and H156. From the measured values, one can calculate the concentration of traps that have emitted an electron. However, this is a little more involved than just dividing the charge by the diode volume.

To begin with, not all of the diode volume takes part. Even with no voltage applied there is a depletion zone of width ( $w_0$ ) at the  $p^+n$  junction. In this region the charge state of any trap remains unaltered. It thus constitutes passive volume.

Secondly, the amount of charge induced by an emitted electron is proportional to the distance it has travelled, (cf chapter 3). Let the origin of the  $x$ -axis be at the  $p^+n$  junction, and let the position  $x = d$  coincide with the  $nn^+$  junction. The thickness of the diode is denoted by  $d$ , its cross-sectional area by  $A$ ; the elementary charge is  $e$ . The relation between the measured charge ( $Q$ ) and the concentration ( $C$ ) of traps having emitted an electron is (cf chapter 3)

$$\begin{aligned} Q/e &= A \int_{w_0}^d \frac{d-x}{d} dx \cdot C \\ &= \frac{A(d-w_0)^2}{2d} \cdot C \end{aligned}$$

It has been assumed in this derivation that the traps are homogeneously distributed throughout the diode volume.

We are left with the problem to determine  $w_0$ . In the previous chapter, the current-voltage characteristics was used to determine the space charge profile after irradiation. The same

measurement can be used to find  $w_0$ . The underlying assumption is that the dark current is proportional to the depleted volume; ie the trap concentration is assumed not to vary with the depth into the diode. Fig. 56 shows a magnified part of such an I-V characteristic at small applied voltages. At voltages below the thermal voltage  $U_T = kT/e \sim 25\text{mV}$  the current rises linearly. The reason is that the balance between the drift current and the diffusion current in the junction is gradually overcome by the applied bias. Once it is bigger than a few times  $U_T$ , all charge carriers that are thermally generated in the space charge region are swept out by the electric field. Extrapolating the dark current measured over the range 0.2V...0.5V down to zero volts gives  $I_0$ , which is the contribution of the zero-bias depletion region of width  $w_0$ . Let  $I_d$  be the volume-generated dark current corresponding to the fully depleted diode (of width  $d$ ), and obtain  $w_0$ :

$$w_0 = \frac{I_0}{I_d} \cdot d$$

The width of the fully depleted diode is found from a capacitance measurement. Its value is, of course, the same before and after irradiation.

Turn again to fig. 57. The left-hand scale shows the measured transient charges; the right-hand scale shows the calculated concentrations of discharging traps. A remark concerning the plot of H155 is in order. This CTR run has been made with a prototype of this system which disallowed measurements below  $-20^\circ\text{C}$ , and a different pulser had been used to drive the diode into depletion. The switching levels were 1.4V in the off-state and 50V in the on-state. In the H155, the depletion width at 1.4V is already  $\sim 95\mu\text{m}$  (inferred from dark current measurements) and the recorded charge is therefore comparatively small. The other three measurements have been made using the improved system, ie with switching from 0.05V to 90V.

The observed concentrations of discharging traps ( $N_{dis}$ ) do not necessarily coincide with the total trap concentration. As can be seen, they are a function of temperature. At high temperatures, when the Fermi level tends to the bandgap centre,  $N_{dis}$  is small, as expected. At low-enough temperatures, the Fermi level rises above the acceptor level, and eventually all of them are charged. For example, we conclude from the low-temperature limit of  $N_{dis}$  that the trap concentration ( $N_T$ ) in diode H137 is  $N_T = 0.72 \cdot 10^{12}\text{cm}^{-3}$ .

As the dose increases (diode H156 to H154) the low-temperature limit of  $N_{dis}$  does not increase beyond the doping concentration of these diodes, though the trap concentration should certainly increase with the dose. Instead, in the diodes H155 and H154,  $N_{dis}$  approaches a value that is exactly the same as the doping concentration before the irradiation.

We cannot overstate the importance of this observation: It discloses the nature of the defect. If it was a donor, being neutral with its electron attached, there would be no reason why the doping should set an upper limit to the number of occupied defects. If it is an electron acceptor, the limiting behaviour has a most natural explanation: The defects cannot trap more electrons than are provided by the dopants.

Moreover, we can infer the charge state of the acceptor from which the electrons are emitted. If we assume that the doping is about the same before and after irradiation, then a look at H137 provides the answer. There  $N_{dis}$  is  $0.72 \cdot 10^{12}\text{cm}^{-3}$  while  $N_D$  is  $1.0 \cdot 10^{12}\text{cm}^{-3}$ . The acceptor level that emits the electrons can only be in the  $-1$  charge state. Any higher charge state is forbidden by the demand for quasineutrality in the base of the diode.

We are thus certain that the process observed is  $A^- \rightarrow A^0 + e^-$ , where  $A$  denotes the hitherto unknown acceptor-like trap. In continuation, the measurement of the trap occupancy tells us that the dopant concentration was not changed by the irradiation, at least to a precision

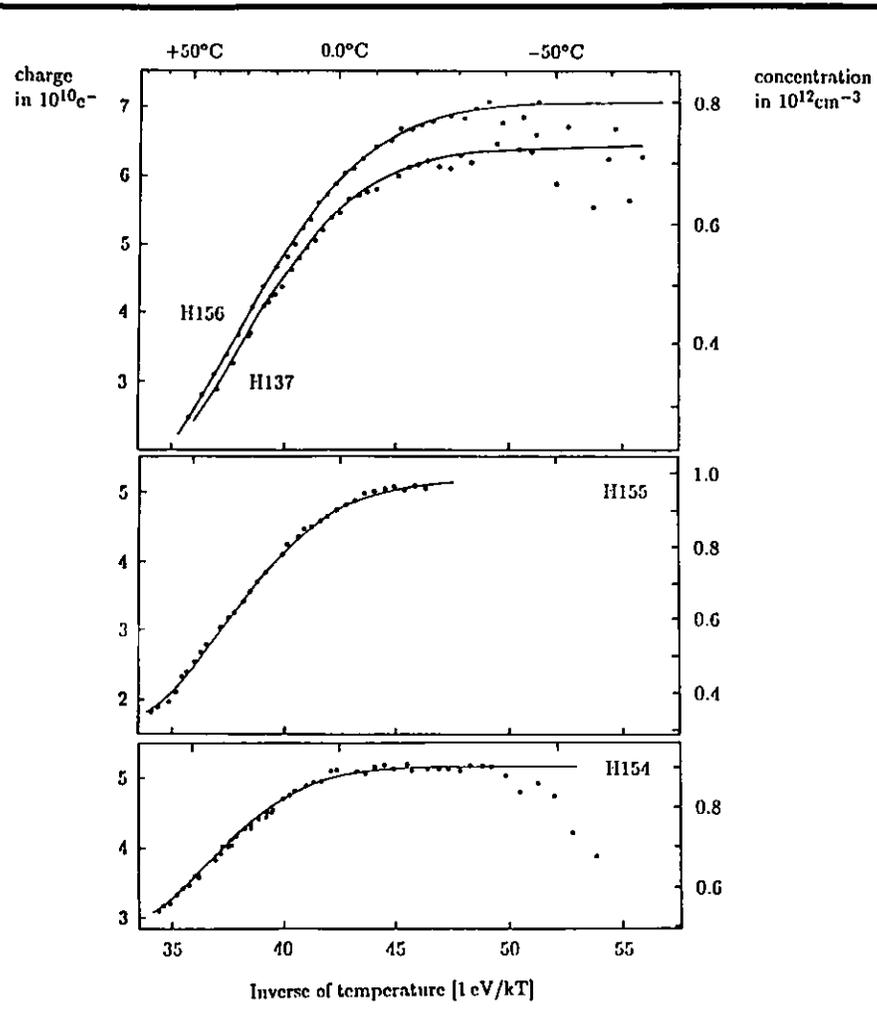


Figure 57: The charge ( $Q$ ) contained in the current transients as a function of the inverse temperature ( $1/kT$ ). The charge was obtained from fitting  $I(t) = q\lambda f(\lambda t)$  and using  $Q = \pi^2/8 \cdot q$ .

The right-hand scales show the concentrations calculated from the total charge (see text). The smooth lines running through the data points are the results from least-squares fits using the theory of trap occupancy developed earlier in this chapter. The presence of an electron acceptor with an emission enthalpy of 0.422eV was assumed. Three parameters were allowed to vary in the fit: the doping concentration, the acceptor concentration, and the ionization entropy. The r.m.s. deviation between fit and data is less than  $0.01 \cdot 10^{12} \text{cm}^{-3}$ .

diode	dose kGy	entropy k	$N_T$ [ $10^{12} \text{cm}^{-3}$ ]	$N_D$ [ $10^{12} \text{cm}^{-3}$ ]	$N_{D, \text{before}}$ [ $10^{12} \text{cm}^{-3}$ ]
H156	5.4	-1.08(7)	0.797(4)	1.05(5)	1.00(5)
H155	10	-1.05(8)	1.69(11)	0.987(5)	0.95(5)
H154	16	-1.72(20)	1.91(25)	0.909(3)	0.93(3)
H137	*	-1.03(11)	0.720(4)	1.10(6)	na

\* ) H137 was irradiated with 14 MeV neutrons to a fluence of  $0.72 \cdot 10^{13} \text{cm}^{-2}$ . na) not available. Numbers in parentheses give statistical r.m.s. errors in multiples of the smallest quoted decimal unit. The first column entry is the received dose in kilo-Gray. There follow the entropy in multiples of the Boltzmann constant, the total trap concentration, the doping concentration after irradiation (from the fits) and before (from C-V characteristics).

Table 1: Summary of the results obtained from comparing the concentration of discharging traps measured at various temperatures to the prediction from theory. It was assumed that the trap can be charged from neutral to negative and that the reaction enthalpy for electron emission is 0.422eV. The entropy, the total trap concentration, and the doping concentration were the free parameters of the fit. The average discrepancy between fit and data is somewhat less than  $0.01 \cdot 10^{12} \text{cm}^{-3}$ .

of a few percent of the concentration of the main radiation-induced effect. Very obviously, this is inconsistent with findings from Lindström et al., [Li 90], and Borchì et al., [Bo 91], who claim that the phosphorus-vacancy (PV) complex constitutes a major part of the radiation damage caused by neutrons in weakly n-doped material. But these authors are also at variance with a large body of literature where the PV complex is reported to be produced in substantial concentrations only if  $N_D \sim 10^{16} \text{cm}^{-3}$  or larger; cf [Zi 89] ( $N_D = 3.9 \cdot 10^{13} \text{cm}^{-3}$ ), [Wa 90] ( $N_D = 6.9 \cdot 10^{13} \text{cm}^{-3}$ ), [Ev 76] ( $N_D = 1.6 \cdot 10^{14} \text{cm}^{-3}$ ), [Br 76] [Br 82] ( $N_D = 8.0 \cdot 10^{13} \text{cm}^{-3}$ ), [Sv 87] ( $N_D = 1.2 \cdot 10^{15} \text{cm}^{-3}$ ) and others. In all these cases the  $VV^=$  and  $VV^-$  are seen in equal concentrations, but no PV are being detected. The PV production becomes dominant only if  $N_D > 10^{16} \text{cm}^{-3}$ , see eg [Mo 82], where  $N_D = 3 \cdot 10^{16} \text{cm}^{-3}$ , and [Aw 89], where  $N_D$  goes up to  $\sim 10^{17} \text{cm}^{-3}$ .

Armed with the knowledge that the deep defect acts like a simple acceptor, we now proceed to calculate how its occupancy should vary with the temperature. The filling pulse duration was 0.1s at all but the lowest temperatures, where it was set to 1s. The base of the diode can therefore be assumed to be in thermal equilibrium. We use the theory and notation of section 6.1.3. The ratio of charged to neutral acceptors is  $A_1/A_0$ , the condition of thermal equilibrium fixes the product  $pn$ , and these two relations are to be inserted into the condition of charge neutrality:

$$\begin{aligned}
 \frac{A_1}{A_0} &= g \frac{n}{N_C} \exp\left(\frac{E_{CT}}{kT}\right) \\
 &= \frac{n}{N_C} \cdot g \exp(-S_{CT}/k) \exp\left(\frac{H_{CT}}{kT}\right) \\
 p &= \frac{n_i^2}{n} \\
 N_D - A_1 - n + p &= 0
 \end{aligned}$$

At any fixed temperature this produces a cubic equation in  $n$  which was solved numerically. For the purpose of fitting the observed temperature dependence of the trap occupancy, three quantities were treated as unknowns: the dopant concentration  $N_D$ , the total defect concentration  $A$ , and the entropy  $S_{CT}$ . The dopant concentration  $N_D$  was treated as unknown because it might have changed as a result of the irradiation. The other two parameters are truly unknown. The degeneracy factor  $g$  has been fixed to  $g = 2$ . The reaction enthalpy  $H_{CT}$  has already been determined very accurately from the Arrhenius plot of figure 55 in which the data from all diodes could be combined. It has therefore been treated as a known parameter and was not allowed to vary in the fits.

The fits for each diode were done by calculating the concentration of charged acceptors at each temperature as a function of the three fit parameters and minimizing the square of the deviations between theory and data. The smooth curves running through the data points in figure 57 are the fits obtained this way. In all cases the r.m.s. difference between calculated and measured concentrations was  $\sim 0.01 \cdot 10^{12} \text{cm}^{-3}$ , reflecting the accuracy of the measurement. The table 1 is a compilation of the fit results. We note that

- to within 5%,  $N_D$ , as obtained from the fit, equals the dopant concentration prior to irradiation,
- the entropy is  $S_{CT} = -1.05k$  with an r.m.s. error of 5%.

We conclude that the doping concentration is not altered by the irradiation. The sizeable reduction of the space charge density in the depletion zone which has been reported in chapter 5 cannot be due to dopant removal, and we will come back to this point later.

In the diodes H155 and H156 the trap concentration is proportional to the received dose. In the most heavily irradiated diode (H154) the trap concentration comes out too low by 20%. Moreover, the entropy obtained here differs a great deal from the value found in the other three diodes. For the time being, it remains unclear what the cause for this discrepancy might be.

The outstanding new feature is that the entropy could be measured quite directly. Its magnitude influences the free energy. And this is the quantity the fits are sensitive to.

The determination of  $S_{CT}$  depends only on the assumption that the measured activation energy is indeed the reaction enthalpy  $H_{CT}$  and does not contain a contribution from an energy barrier in the capture cross-section. If this is fulfilled, then the reaction entropy  $S_{CT}$  has been determined to an accuracy of 5%. Measuring the trap occupancy as a function of temperature obviously provides a way to determine the entropy factor very precisely.

In the theoretical section of this chapter it has been argued that the change in entropy that results from lifting an electron from the valence band to the neutral acceptor ( $S_{TV}$ ) and from the acceptor to the conduction band ( $S_{CT}$ ) should be:

$$S_{TV} = S_I + S_h \quad S_{CT} = -S_I + S_e$$

$S_h$  and  $S_e$  are the contributions from creating a free carrier, and  $S_I$  is the ionization entropy brought about by the alteration of the electronic and atomic configuration when the acceptor is charged. With these definitions and the measured values we find:

$$S_e = 0.6k, S_{CT} = 1.05k \implies S_I = +1.65k$$

The entropy of ionization is positive which means the lattice is weakened by the bound electron.

In terms of free energy the acceptor level does not remain at a constant depth below the conduction band edge, but falls slowly with temperature:

$$\begin{aligned} E_{CT} &= H_{CT} - T \cdot S_{CT} \\ &= 0.422\text{eV} + 1.05kT \end{aligned}$$

To conclude this section on the experimental findings we note the following:

- Four weakly  $n$ -doped diodes have been exposed to either 14 MeV neutrons ( $0.72 \cdot 10^{13} \text{cm}^{-2}$ , H137) or photon-induced particle cascades developed in lead as the absorber material (5.4kGy, 10kGy, 16kGy, for H154, H155, H156, respectively).
- the produced deep-defect concentration is proportional to the received dose.
- The defect acts as a simple acceptor.  
The free energy for electron emission is  $E_{CT} = 0.422\text{eV} + 1.05 \cdot kT$ .
- The concentration of dopant atoms is not altered by the irradiation.

### 6.3.2 Chemical identity of the deep defect

The CTR measurements have already provided us with detailed knowledge about the defect. It has (at least) one acceptor level, whose thermodynamic characteristics could be determined very precisely. The absolute value of the electron emission rate together with its activation energy are the most useful quantities when it comes to identifying the defect. Put together they are like a spectroscopic fingerprint.

There are two acceptor-like defects that under the circumstances deserve attention: the phosphorus-vacancy ( $PV$ ) and the divacancy ( $VV$ ). The  $PV$  has an acceptor level close to the one observed in this work, but it is unlikely that it has been produced in substantial concentrations. First of all, consider the production mechanism. Simple vacancies are produced by the penetrating radiation. These then diffuse through the crystal to combine with the phosphorus and form  $PV$ 's. But the vacancies also combine with oxygen to form stable, immobile defects, and oxygen is in our weakly doped diodes 10,000 times more abundant. As long as the oxygen is not used up, no  $PV$ 's can probably be produced. The second nail in the coffin is that in the experiments described here no change of the doping concentration could be detected. We are therefore left with the divacancy as the most promising candidate.

The divacancy has been studied extensively by EPR, DLTS, and by photoabsorption techniques. Some of its characteristics have been in dispute but it is now known for certain that depending on the Fermi level it occurs in four charge states.

There is a donor level (+/0) at  $E_V + 0.2\text{eV}$ , an acceptor level (0/-) at  $E_C - 0.42\text{eV}$ , and a double acceptor level (-/=) at  $E_C - 0.23\text{eV}$ . Both, the donor and the double acceptor level are outside the range accessible in this study. The acceptor level (0/-) seems to have the right properties. The charge state (-1), the activation energy of 0.42eV, and its proximity to the conduction band match the properties of the defect found in this work.

We should therefore want to compare our results with those obtained for the divacancy when DLTS, an otherwise very successful method, is used. However, the electron emission was observed to be non-exponential, and the analysis of DLTS spectra, on the other hand, relies on the assumption of an exponential-like emission. As a consequence, the actual DLTS response for the divacancy may differ somewhat from the expected; cf sect 6.3.4.

Author	activation energy [eV]	prefactor [ $10^{-12}$ s]
This work	0.4216(7)	1.81(5)
Wang et al., [Wa 90]	0.43	1.6
M.Hüppi, [Hü 90]	0.42(2)	1.7
Brotherton et al., [Br 82]	0.413	1.6(2)
C.T.Sah et al., [Sa 85]	0.425(4)	2.9(11)
B.G.Svensson et al., [Sv 82]	0.424	1.8

Prefactors are cited as given by the authors. If the emission had been non-exponential in the way observed in this work the prefactors should be divided by 1.23 to compare them to the 1.81ps found in this work. Numbers in parentheses are errors in multiples of the smallest quoted decimal unit.

Table 2: Electron emission parameters for the -1 charge state of the divacancy as observed by various groups.

It turns out that the activation energy is correctly determined, but that the emission rate from DLTS is consistently 23% higher than  $2\lambda$ —the  $\lambda$  of the function  $f(\lambda t) = (\lambda t)K_0(\lambda t)K_1(\lambda t)$ . All in all, a comparison with DLTS data is fairly simple. A compilation of DLTS results on the divacancy acceptor level is found in table 2. Cited are references where the measurements have been made on n-type phosphorus-doped material in which  $N_D$  did not exceed  $\sim 10^{18}\text{cm}^{-3}$ . Otherwise, the signal from the divacancy and the phosphorus-vacancy trap interfere.

Usually, a 'typical' DLTS spectrum, obtained at a specified rate window (see below), is reproduced in the above references. The rate windows ranged from  $223\text{s}^{-1}$  to  $3\text{s}^{-1}$ . The cited temperatures at which the DLTS peak occurs are within  $\pm 1\text{K}$  of the ones that can be calculated from the parameters found in this work. There is only one exception to this. In all papers of B.G.Svensson et al. the peaks consistently occur at a temperature 8K higher than calculated, which also puts them at variance with the other authors.

To within a fine error margin of only 2% the activation energy found in this work matches that of the electron emission from the  $VV^-$ . The spread in the prefactors is bigger (20%) but they also coincide within errors. Putting together that the charge state (-1) and the position of the trap level (in the upper half of the bandgap) is correct with the perfect agreement in the electron emission rates, it seems safe to conclude that the trap in question is indeed the divacancy and that the observed electron emission stems from the  $VV^- \rightarrow VV^0 + e^-$  transition.

As a cross check, we should also compare the measured trap concentration to the received dose. The diodes were exposed to electromagnetic showers at depths of 6,8, and 10 radiation lengths (cf chapter 5). The showers originated from bremsstrahlung gammas produced by stopping 2GeV electrons. The diodes were thus located well behind the shower maximum (cf chapter 2) and the energy of the penetrating particles was mainly in the multi-MeV range.

The production rates of divacancies by electrons of various energies are quite well known:

- at 1MeV it is  $\eta = 0.002\text{cm}^{-1}$ , [Ev 76] [Co 65] [Br 82];
- at 2MeV it is  $\eta = 0.0167\text{cm}^{-1}$ , [Sv 87];
- at 12MeV it is  $\eta = 0.043\text{cm}^{-1}$ , [Br 82].

It rises steeply at the threshold and its value at 12MeV can be regarded as the limit value. The trap concentration to be expected after irradiation is the product of  $\eta$  and the particle fluence.

The energy of the particles in a shower is of course not unique, but we shall use  $\eta = 0.033\text{cm}^{-1}$  for this rough estimate. To produce a trap concentration of  $10^{12}\text{cm}^{-3}$  at this  $\eta$  a fluence of  $3 \cdot 10^{13}\text{cm}^{-2}$  is required. In silicon this fluence of minimal ionizing particles deposits a dose of

$$D = 1.66\text{MeVg}^{-1}\text{cm}^2 \times 3 \cdot 10^{13}\text{cm}^{-2} \\ = 8\text{kGy}$$

This actually matches the dose the diodes have received; 5.4kGy to 16kGy.

The  $VV^-$ -production rate by 14MeV neutrons has not been measured previously, and we can now provide this number from the results obtained from H137. It received a fluence of  $0.72 \cdot 10^{12}\text{cm}^{-2}$  and showed  $0.72 \cdot 10^{13}\text{cm}^{-3}$  of divacancies giving  $\eta = 1.0\text{cm}^{-1}$ . With a silicon solid state density of  $4.8 \cdot 10^{22}\text{atoms/cm}^3$  this gives an effective production cross-section of 21b. At 14MeV the total interaction cross-section is 2.86b, which means that per hit 7.3 divacancies are produced.

### Comment on non-exponential transients

For an isolated trap the probability to emit the bound charge carrier in a fixed time interval is a constant, independent of its history. As a result, trap discharging has a time dependence that is purely exponential. The observed transient is a superposition of all individual emission events. It may now be that the environment in which the traps are embedded influences the emission rate. If this is not the same for all traps the transient current cannot be exponential.

Two such influences have been discussed in literature and are by now well established. One is alloy broadening in III-V semiconductors. In  $GaAs$ , for example, a certain fraction ( $x$ ) of the  $Ga$  atoms may have been replaced by  $Al$  to produce  $Ga_xAl_{1-x}As$ . The energy levels of a trap in this alloy semiconductor may depend on the chemical environment it is in. The exact position of the levels may depend on whether its next (and second-next) neighbours are the big heavy  $Ga$  atoms or the small and light  $Al$ .

The other effect is field-enhanced emission. It is observed when a carrier is only loosely bound to its trap and orbits it at a great distance. In the presence of a strong electric field, the energy barrier for carrier emission is lowered because of the potential energy the carrier gains in the field. The reduction in barrier height depends on the magnitude of the electric field, and in a reverse biased diode this is not constant, which makes the resulting current transients a superposition of exponentials with very different emission rates. An example of a shallow trap for which the Frenkel-Poole effect has been studied in some detail is interstitial boron in silicon, [Ha 87].

Silicon is an elemental semiconductor and alloy broadening does certainly not occur. Field-enhancement of the carrier emission rate is negligible for the divacancy. A very detailed analysis of EPR spectra, [Si 90], revealed that its central cell, which is  $\sim 5\text{\AA}$  in diameter, contains 80% of the charge of the bound electron. The rest is spread out along the zig-zag path of silicon atoms pointed to by the long axis of the divacancy. The density of the outspread charge falls exponentially with the distance, measured along this direction, with a decay length of  $3.4\text{\AA}$ . For the electric fields encountered in our diodes ( $\leq 5\text{kV/cm}$ ) and a charge extension of  $10\text{\AA}$  one calculates a barrier lowering of 0.5meV. At room temperature this translates into an increase of the emission rate of only 2%, [La 83].

With the  $VV^- \rightarrow VV^0 + e^-$  transition the limiting normalized exponential-like emission rate is  $e_n = -f'(x)/f(x) \rightarrow 2$  for  $x > 1$ . At the low end of the time scale, at  $x = 0.1$ , the rate is bigger by a factor of ten. Thus field-enhanced emission is also ruled out.

What remains is a possible dependence of the  $VV^-$  level on lattice strain in its immediate surroundings. The diode H137 has been exposed to 14MeV neutrons. The silicon atoms displaced by those have an average kinetic energy of 0.7MeV. At the beginning of the slow-down they lose energy mainly by ionization and few isolated displacements. At the end of their track the displacement events are more violent, creating densely disordered regions where consequently the mechanical strain in the lattice is large.

The diodes H154, H155, and H156 have been irradiated with electromagnetic showers which developed in lead as the absorber. The cascades of particles and photons produce a considerable fluence of photon-neutrons as well. And the radiation damage takes two contributions (comparable in size)—one from the electrons and positrons and the other from the neutrons; the latter with the above described implication concerning lattice strain.

That the divacancy levels do respond to lattice strain has been investigated in great detail by a group around B.G.Svensson. They started from the microscopic model proposed by Corbett et al., [Co61]. In the neutral divacancy a long electron bond is formed between two silicon atoms located at the opposite ends of the defect. The other four atoms with dangling bonds form short bonds with their next neighbour. Because of crystal symmetry there are three equivalent bonding configurations (three equivalent directions for the long bond). The activation energy for switching from one configuration to another has been determined by EPR studies to be  $\sim 60$ meV, [Co65]. At temperatures well below 80K this bond-switching stops.

When the divacancy is loaded with one or more surplus electrons symmetry-lowering (Jahn-Teller) distortions of the lattice set in. Energy levels take contributions from the electronic binding energy and the Jahn-Teller distortions. At room temperature, bond-switching occurs at a very high rate, and there may simply not be enough time in between for the defect to relax into its Jahn-Teller ground state. The existence of such an effect in the divacancy was demonstrated by showing that at low temperatures ( $< 50$ K) the  $VV^-$  has an excited state, 0.34eV above the ground state, that is gone above 80K when bond-switching fully sets in, [Sv88].

In a different approach, the two acceptor levels of the divacancy have been studied using DLTS. The study begins with electron-irradiated material in which only isolated defects and no damage clusters were produced. In a series of experiments the electron dose and energy were varied. This was combined with a measurement of the trap concentration profiles, and a perfect one-to-one correspondence between the level at  $E_C - 0.42$ eV and at  $E_C - 0.23$ eV was found. They were attributed to the  $-1$  and the  $-2$  charge state of the divacancy, [Sv87].

The same procedure was repeated irradiating diodes with increasingly heavy ions from  $^1H$  to  $^{127}I$ , [Sv91]. It was found that close to the entrance point, where the ions lose their energy predominantly by ionization and low-energy displacements, the one-to-one correspondence was still valid. But on the far side, where the ion tracks end in damage clusters, the correspondence broke down. The concentration of the double acceptor niveau reduced compared to the simple acceptor niveau by as much as a factor of 20 in the case of iodine implantation.

The authors ascribe this to the fact that the  $VV^=$  level is a motionally averaged state—brought about by bond-switching between its various configurations. Increased lattice strain in the disordered end region would remove the perfect equivalence between the three bonding configurations and might in its extreme even inhibit bond switching. The Jahn-Teller distortion that can now occur would change energy levels. Jahn-Teller distortions minimize the total

energy, ie the elastic and the electronic energy. If the gain in elastic energy is big enough the electronic level can even be shifted up into the conduction band. In a DLTS experiment it would then vanish, which is what seems to happen to the  $VV^=$  level. Note that the  $VV^=$  has no unpaired electron, and is not observable using electron paramagnetic resonance (EPR). Thus there is no information whether it exists at temperatures of  $\sim 50$ K when bond-switching is frozen in and and Jahn-Teller distortions set in to full effect.

The authors also report a broadening of the DLTS peak associated with the electron emission from the  $VV^-$ . The full width at half maximum is 21K in the electron-irradiated material and increases to 29K for iodine irradiation. They do not explicitly report on non-exponential transients but the broadening of the DLTS peaks can have no other reason.

In the irradiations done for this study, diodes have been exposed to fast neutrons and electromagnetic showers. In both cases a considerable fraction of the damage is eventually caused by energetic silicon recoils. It may therefore be that the non-exponential current transients are caused by a smearing of the electronic level of the  $VV^-$ , which is induced by varying degrees of lattice-strain.

The observed current transients would be a superposition of the form

$$I(t) = \int_0^{\infty} g(\lambda)e^{-\lambda t} d\lambda$$

where  $g(\lambda)$  would be proportional to the density of traps that have an emission constant  $\lambda$ . One may regard the above equation as a Laplace transform of  $g(\lambda)$  and might be tempted to invert it since  $I(t)$  is a known function. However, such an attempt is frustrated by the fact that there is an infinite set of functions whose Laplace transform is constant and equal to zero. In short, if the Laplace transformation of a function is known only on the real axis (or any other for that matter) its inverse cannot be uniquely determined, [Do58].

All we can do is estimate (guess in a mathematicians opinion) the width of the energy level. At big times we have  $I(t) \propto \exp(-2\lambda t)$ . At the smallest time where the model function  $f(\lambda t)$  still fits well  $I(t)$  we have  $\lambda t = 0.1$ . An exponential tangent to  $I(t)$  at this point is  $\propto \exp(-23\lambda t)$ , ie the emission constant is roughly ten times bigger. This translates into a level width of 2.4kT or 60meV at room temperature.

### 6.3.3 The midband defect

During the previous discussion we have left aside the following facts.

- The dark current of the irradiated diodes has an activation energy of 0.65eV. This means there is defect level at about the centre of the bandgap—a mid-band level.
- The net doping in the space charge region is reduced as a result of the irradiation meaning that at least part of the mid-band traps are charged negatively there. The negative space charge density introduced into the depletion zone is about one third of the divacancy concentration in the diodes exposed to electromagnetic cascades and two thirds in the neutron irradiated diode.
- On the other hand, the measured concentration of  $VV^-$  approaches, at low temperatures and high doses, the doping concentration that was measured before the irradiation. This leaves no place for a negatively charged trap in the base of the diode besides the  $VV^-$ .

diode	dose	$j_{dark}$	$\Delta N_D$	$G$	$A$	$[VV]$
	kGy	$\mu A cm^{-3}$	$[10^{12} cm^{-3}]$	$e^{-} \cdot h^{+} / s$	eV	$[10^{12} cm^{-3}]$
H 156	5.4	32.5	0.27(2)	750(50)	0.65(1)	0.797(4)
H 155	10	61.3	0.47(2)	820(40)	0.65(1)	1.69(11)
H 154	16	101	0.74(5)	850(50)	0.66(1)	1.91(25)
H 137	*	57.2	0.45(7)	850(100)	0.65(1)	0.720(4)

\* H137 received  $0.72 \cdot 10^{13} cm^{-3}$  of 14MeV neutrons. The entries are the received dose, the dark current per volume at 300K, the apparent reduction of doping ( $\Delta N_D$ ) in the space charge region (quoted from chapter 5). From the ratio of latter two numbers the electron-hole pair generation rate per unit of induced space charge is calculated; rate  $G$  in pairs per second at 300K. There follows the activation energy of the dark current measured between 280K and 300K. Last entry is the concentration of divacancies.

Table 3: The action of the mid-band trap, namely to introduce dark current and negative space charge are summarized in this table.

- If the mid-band trap remains neutral in the base, for whatever reason, one should be able to observe it charging up in the depletion zone. However, immediately after the current transient from the  $VV^{-}$  discharge is gone the dark current has its steady state value, to an accuracy of 1%.

It is certain that the dark current is not produced by the divacancy. Its level closest to the band centre is 0.42eV below  $E_C$  and  $1.20eV - 0.42eV = 0.78eV$  above  $E_V$ . In terms of free energy the  $VV^{-}$  level is located 100meV ( $\approx 4kT$  at room temperature) above the mid-band position. In the depletion zone the  $VV$  are all neutral and the activation energy of the dark current produced by them is 0.78eV—and not 0.65eV as observed. We are therefore sure that neither the dark current nor the negative space charge introduced into the depletion region are due to the divacancy. But the radiation damage we are chiefly interested in is brought about by another defect with an energy level almost at the centre of the bandgap. Table 3 gives a summary of its characteristics obtained so far. We see that the electron-hole pair creation rate per trap as well as the activation energies are identical in all four diodes which leads us to the conclusion that it is the same kind of trap in each case.

While a mid-band level is clearly present it has never shown up in DLTS studies. In no DLTS study known to the author whether done on  $n$ - or  $p$ -type material, whether irradiated with electrons, protons, neutrons or heavy ions has there ever been reported an electronic level at  $E_C - 0.65eV$  or  $E_V + 0.65eV$ . There is only one exception to this due to S.K.Bains et al., [Ba 86]. Boron-doped silicon ( $1 - 3 \Omega cm$ ) was irradiated with electrons of 1.5MeV at 20K and then warmed up to room temperature. An energy level at  $E_C - 0.67eV$  has been found, but the carrier emission rate is too low by a factor of four compared to that of the mid-band trap seen in this work.

It is only after the irradiated diodes have been heated to above 300°C for several hours that energy levels next to the band centre are commonly reported; [Aw 88], [Hü 90], [Ja 85]. But the findings do not reproduce well in between the authors, and in all cases are the emission rates of the produced traps too small by factors of 4 to 100.

It thus seems that DLTS is insensitive to the mid-band trap. From what has been observed using CTR this is quite understandable. Immediately after the  $VV^{-}$  transient is gone, the dark

current has its steady-state value; ie the one that one measures manually when using a power supply and an amperemeter. This has been verified to an accuracy of a few percent over the temperature range 10°C to 50°C. At the high end the agreement was good to 1%.

The implication is that the occupancy of the mid-band trap does not change after the  $VV^{-}$  discharge has subsided. And it probably does not change during the main portion of the  $VV^{-}$  transient. If it did it would affect the dark current because in general electron emission and hole emission rates from a level differ by orders of magnitude, which makes the momentary dark current a strong function of the charge state predominant at a particular time.

Since no change of the occupancy occurs after the  $VV^{-}$  has discharged, any change that might occur must occur faster than the  $VV^{-}$  discharging—and a change in occupancy should indeed occur. To see why, consider the case of the diode H154 irradiated to 16kGy of dose. There the apparent reduction of doping in the field zone amounts to 80%. Yet, the  $VV^{-}$  concentration approaches the doping concentration that was measured before the irradiation. The mid-band trap, for whatever reason, remains neutral in the base, but in the field zone contributes a space charge amounting to 80% of the doping. When the diode is pulsed into depletion the mid-band trap should be neutral at the beginning, gradually charging up by hole emission. During this process, both, the dark current and the space charge, would change. The latter would necessarily give rise to a DLTS peak with an activation energy at about the band centre which has never been reported.

And this is the conundrum. How can a trap that very obviously does introduce a negative space charge into the field zone be uncharged in the base of the diode where there are plenty of electrons around to be captured? To make it even more weird, when the diode is pulsed into depletion the mid-band traps in the depletion zone immediately turn up in the correct steady-state charge state with no measurable time constant attributable to it.

For the time being, no sound solution of the puzzle can be given. But we can outline the direction along which to search. We begin by proposing that the mid-band level is associated with defect clusters.

We begin with an examination of the properties of the defect clusters that are produced when the recoiling silicon atom is finally stopped. At the beginning of its path it loses its energy mainly by ionization. Once it is slow enough ( $E < 5keV$ ) it will deposit the rest 'on the spot' by displacing lots of other silicon atoms. This process, originally suggested by Brinkmann, [Br 54], is thought to produce a small amorphous region. More recent theoretical, [Ce 87], and Monte Carlo, [Mu 82], calculations indicate that these clusters have a diameter of  $\sim 50\text{\AA}$  that the energy deposited to produce them is between 2keV and 8keV, and that per primary knock-on atom of 50keV, two or three clusters are produced. Transmission electron micrographic pictures of phosphorus-irradiated silicon (100keV,  $P^{+}$ ) show spots of 20Å in diameter where the electron transmissivity is notably different from the surrounding matrix, [Ho 80].

It is clear that the individual defects which make up a cluster cannot be viewed as independent when they are separated by only one or two lattice constants. Instead, the cluster should be regarded as one big defect that introduces a probably big number of electronic states into the bandgap.

In spite of this, the number of electrons or holes it can trap is probably limited to quite a few because of carrier-carrier repulsion. The self energy of a homogenous cloud of charge with a total charge  $Ze$  confined to a sphere of radius  $R$  is, [Ja 79]:

$$W = \frac{16}{15} \pi^2 R^5 \frac{(Z)^2 \cdot 1.224eV\text{\AA}}{\left(\frac{4}{3}\pi R^3\right)^2}$$

$$= \frac{3 Z^2 \cdot 1.224 \text{eV}\text{\AA}}{5 R} = \frac{0.73 \text{eV}\text{\AA} Z^2}{R}$$

The constant  $1.224 \text{eV}\text{\AA}$  is equal to  $e^2/(4\pi\epsilon)$ . At  $R = 10\text{\AA}$  (from the micrograph) the self energy of trapped electrons would be  $0.070 \text{eV}$ ,  $0.28 \text{eV}$ , and  $0.63 \text{eV}$  for one, two, or three trapped electrons. Depending on the cluster size and the details of its electronic states it seems that three or four charges are the maximum number that can possibly be trapped.

We proceed to make plausible that both, the dark current and the negative space charge introduced into the depletion zone, can be brought about by Brinkmann clusters.

Consider first the average charge a cluster has in the depletion zone of a diode. Since the activation energy of the dark current is  $0.65 \text{eV}$  and the clusters are charged negatively, we conclude that their first acceptor level is a little below the intrinsic level. Because of the high electron self interaction energy that there is when two electrons occupy the defect, the double acceptor level should be about  $200 \text{meV}$  above the intrinsic level and thus be unoccupied in the field zone. The stable charge state is therefore most likely the  $-1$  state.

In the diode H137 the concentration of negative charged traps is  $0.45 \cdot 10^{12} \text{cm}^{-3}$  after it has received a  $14 \text{MeV}$ -neutron fluence of  $0.72 \cdot 10^{12} \text{cm}^{-2}$ . The space charge density in the depleted diode was first measured four days after the irradiation and again two years later. To within an accuracy of  $0.01 \cdot 10^{12} \text{cm}^{-3}$  it had not changed. This matches the findings of Wunstorff et al., [Wu 91], who find that after  $14 \text{MeV}$ -neutron irradiation a part of the produced negative space charge density dissolves with an exponential decay constant of 9 hours. Once it is gone, no further changes are reported to occur.

We now know the introduction rate of clusters by  $14 \text{MeV}$  neutrons: It is, with an error of 10% from the neutron fluence determination and 20% from determining  $\Delta N_D$ ,

$$\eta = \frac{0.45 \cdot 10^{12} \text{cm}^{-3}}{0.72 \cdot 10^{12} \text{cm}^{-2}} = 0.63 \text{cm}^{-1}$$

This has to be compared with the introduction rate of primary recoils. At  $14 \text{MeV}$  the neutron interaction cross-section is  $2.86 \text{b}$ . The introduction rate of primaries becomes  $0.137 \text{cm}^{-1}$ . And the ratio of both rates gives the number of clusters per primary interaction: 4.6 cluster per hit. This is higher than the two to three clusters per  $50 \text{keV}$  primary, but given that for  $14 \text{MeV}$  neutrons the average primary recoil energy is  $720 \text{keV}$ , it does not seem unlikely that the number of clusters is somewhat bigger in this case.

Next we have to deal with the fact that the dark current decreased to 40% of its post irradiation value in a period of 100 days whereas the cluster concentration remained constant. A clue comes from annealing studies made by L.M.Howe et al., [Ho 80]. They observed by inspecting micrographs from a transmission electron microscope that the clusters introduced by phosphorus implantation show some annealing already at room temperature. Beginning with detailed annealing studies at  $350 \text{K}$  they found that the clusters do not suddenly dissolve at a given temperature or turn into other types of lattice defects but gradually shrink in size until they are gone (for the microscope).

The reduction of the dark current can then be understood to be the result of cluster shrinking. From the activation energy ( $0.65 \text{eV}$ ) and the absolute value of the electron-hole generation rate ( $800 \text{s}^{-1}$  at  $300 \text{K}$ ) we can estimate the electron capture cross-section. It is the electron cross-section because by claiming the clusters to be on average charged negatively we have implied that the electron emission is the slower process and thus determines the overall rate of thermal pair generation. If we ignore the degeneracy factor and the emission entropy we can

calculate a cross-section that is often called an emission cross-section, indicating the source it was obtained from.

$$\begin{aligned} e_n &= \sigma v_{th} N_C \exp\left(-\frac{H_{CT}}{kT}\right) = 800 \text{s}^{-1} \quad (\text{at } 300 \text{K}) \\ \sigma &= 2300 \text{\AA}^2 = \pi R^2 \\ R &= 27 \text{\AA} \end{aligned}$$

This huge value already implies that the defect is big. The capture radius  $R$  is bigger than what is seen under the microscope, but it is only a crude estimate anyhow. If the electron emission rate from a cluster is linked to its radius in the way indicated above, then a reduction of the dark current by a factor of 2.5 calls for a shrinking of the cluster to 63% of the diameter it had on creation. It probably shrinks by slowly evaporating simple vacancies, which are free to move, once they have escaped from the cluster.

Small clusters, the ones likely to be produced by silicon recoils, begin to anneal out at  $350 \text{K}$  and are mostly gone at  $500 \text{K}$ —in an isochronal anneal study: 10 minutes at a few temperatures between  $350 \text{K}$  and  $750 \text{K}$ , [Ho 80]. Lindström et al. have done a similar measurement: 60 minutes at several temperatures between  $300 \text{K}$  and  $500 \text{K}$ . After each step, they measured the dark current at  $300 \text{K}$ . They found that annealing sets in at  $370 \text{K}$  and is completed at  $500 \text{K}$ , [Li 89]. Similar results have been obtained in 1 min. isochronal anneals: Onset at  $\sim 400 \text{K}$  and completion at  $\sim 600 \text{K}$ , [Li 91].

To sum up, it seems that the amorphous regions at the end of the paths of energetic recoils can account for the dark current and the space charge introduced into the depletion zone. What remains to be understood is their mysterious behaviour in the neutral base of the diode where they obviously remain neutral though the Fermi level is well above the bandgap centre. It seems as if the negative charge was shielded by a loosely bound hole that is stripped off when the diode is pulsed into depletion. This way the negative core charge would appear without time delay.

### 6.3.4 Comparison with DLTS

DLTS has been invented in 1974 by D.V.Lang, [La 74]. The abbreviation stands for deep level transient spectroscopy. 'Deep level' refers to charge carrier traps whose ionization energy is at least a few times bigger than that of the shallow donors and acceptors that are used as dopants. The transient involved is the capacitance transient induced by carrier emission from the traps, and the method is spectroscopic in the sense that different types of traps will give rise to transients with different decay rates.

DLTS can be applied to junction and Schottky barrier diodes. The diode is held under reverse bias and must not be fully depleted. All electron acceptor levels above the bandgap centre are empty in the depletion zone of the diode. Now a 'filling pulse' is applied, ie the bias is reduced for a certain time. During this the depletion zone is smaller, and formerly empty traps can be filled with electrons. At the end of the filling pulse the diode is brought back to the initial bias. But its capacitance immediately after the filling pulse is different from what it was before. The reason is the difference in the space charge density caused by the now filled traps. As they discharge the capacitance difference reduces gradually to zero. It can be shown that if the trap concentration ( $N_T$ ) is very much smaller than the doping ( $N_D$ ) and all traps in

the depletion zone have been filled, the transient will trace the trap discharging, [La 83].

$$\frac{C(t) - C(\infty)}{C(\infty)} = \frac{N_T}{2N_D} e^{-\lambda t}$$

The DLTS apparatus is an analog device; it does not digitize and store the transient. In order to extract both the ratio  $N_T/N_D$  and the emission constant  $\lambda$ , a filtering operation is applied. The output signal of the apparatus is the difference of the capacitances measured at two preset times,  $t_1 < t_2$ , after the end of the filling pulse and continuously averaged over the repeating pulses:

$$\begin{aligned} S &= C(t_1) - C(t_2) \\ &= \frac{N_T}{2N_D} (e^{-\lambda t_1} - e^{-\lambda t_2}) \end{aligned}$$

The signal is proportional to the trap concentration and is a function of  $\lambda$ . If  $\lambda t_1 \gg 1$  or  $\lambda t_2 \ll 1$ , the signal is small. At some intermediate value of  $\lambda$  it will attain its maximum.

$$S(\lambda) = \text{maximum} \implies \hat{\lambda} = \frac{\log(t_2/t_1)}{t_2 - t_1}$$

The emission rate at which the maximum signal is expected, is the 'rate window' of the setup. The emission rates of the carriers from their traps change exponentially with the inverse of temperature. One therefore varies the sample temperature over a wide range while continuously recording the DLTS signal. From low to high temperatures the DLTS signal will exhibit a number of peaks each time the emission constant of a certain type of trap matches the preset rate window. Repeating the temperature scan with some other rate window allows to determine the activation energy of the traps.

It is also possible to measure the capture cross-section of a trap. A rate window is preset, and the temperature regulated in order to have the maximum signal height. Then the signal height is measured as a function of the filling-pulse width. One expects the trap filling to obey an exponential law. The filling-rate constant for an electron trap is  $\lambda_{fill} = \sigma v_{th} n$ . Here,  $\sigma$  is the capture cross-section,  $n$  is the concentration of the electrons during the filling pulse and  $v_{th} \approx 1.11 \cdot 10^7$  cm/s (at 300K) their thermal velocity. The procedure must be repeated at a number of temperatures in order to recognize the action of a possibly existing energy barrier.

In practice, this kind of measurement is limited to rather small capture cross-sections ( $\mathcal{O}(10^{-16} \text{cm}^2)$ ) because the filling rates become too high otherwise.

### Limitations of DLTS

DLTS is a well-proven and highly successful method to characterize deep defects. It knows only one major limitation: the trap concentration must be very small compared to the doping. If this is not the case, all three basic assumptions on which it relies are violated. These are:

- The width of the depletion zone is related to the device capacitance by  $w = \epsilon A/C$ .
- The change of the capacitance is small and proportional to the change of the depletion-zone charge density.
- The capacitance transient traces the carrier emission and is exponential.

Electric field distribution in the depletion zone in the stationary case (continuous lines) and immediately after the trap-filling pulse is gone (dashed line). In the region  $w < w_0$  the material is always depleted, so the space charge density there ( $\rho_0$ ) is time independent. Applying a backward bias moves the edge of the depletion zone to  $w(t)$ . In the region  $w_0 < w < w(t)$  the charge density  $\rho(t)$  changes with time until stationary conditions are reached and  $w = w(\infty)$ .

The change of the depletion capacitance after the filling pulse. The total voltage across the diode has been

$$V_0 = \frac{1}{2\epsilon} \rho_0 w_0^2$$

prior to the end of the filling pulse and will be

$$V_\infty = \frac{1}{2\epsilon} \rho_0 w_\infty^2$$

long after the pulse. Note that it has been assumed that the charge density ( $\rho$ ) is constant throughout the detector volume. The voltage step  $\delta V = V_\infty - V_0$  is constant in time and can be read off the above figure.

$$\begin{aligned} V(t) - V_0 &= -w_0 E'(w_0) - \frac{1}{2}(w(t) - w_0) E'(w_0) \\ &= -\frac{1}{2}(w_0 + w(t)) E'(w_0) \\ E'(w_0) &= -\frac{1}{\epsilon}(w(t) - w_0) \rho(t) \end{aligned}$$

The last relation is obtained by integrating  $\text{div } E = \rho/\epsilon$  from  $x = w(t)$  to  $x = w_0$ . From this the change of the charge density may be expressed in terms of widths or capacitances using  $w = \epsilon A/C$ :

$$\frac{\rho(t)}{\rho_0} = \frac{w_\infty^2 - w_0^2}{w(t)^2 - w_0^2} = \frac{C_0^2 - C_\infty^2}{C_0^2 - C(t)^2} \frac{C(t)^2}{C_\infty^2}$$

Only if  $\delta C = C(t) - C_\infty \ll C_\infty$  the linear relationship  $\delta\rho/\rho = 2\delta C/C$  is recovered.

To measure the capacitance of the diode a small (10mV r.m.s.) high frequency ac signal is applied to its terminals. In response to this, the edge of the depletion zone oscillates to and fro. Just by how much, depends on the immobile-charge density in the base, ie the dopant density minus the density of the trapped majority carriers. If the majority-carrier traps outnumber the dopants, the neutral base of the diode becomes nearly intrinsic. In that case the movement of the depletion zone boundary becomes virtually unrestricted. And the capacitance of the device then bears no relationship to the depletion zone width that there is when no ac measuring voltage is applied.

If the trap concentration is not small enough, the change in capacitance is not proportional to it; see box on next page. Since in DLTS one does not record the whole transient, there is no handle to determine the trap concentration in this case.

Besides the effect of a non-linear response of the capacitance to the trap discharging there is another more severe difficulty if  $N_T$  is not small enough. It has been pointed out by Landsberg, [La,87], Grimmeis, [Me 83], and others that in this case the recapturing of emitted carriers at the end of the depletion zone becomes important. As a result, the transients becomes non-exponential and inferred capture cross-sections can be off by orders of magnitude.

There are other minor limitations. One of them is important in the case of diodes made on high-resistivity material. It must not be overlooked that the Fermi level in such diodes can be quite close to the bandgap centre. If the rate window has been set too high, the DLTS peak should occur at a temperature at which the Fermi level passes or has passed the trap level in question. In that case, its occupancy becomes a strong function of temperature. This shifts and deforms the DLTS peak and renders concentration measurements useless.

In  $n$ -doped material the Fermi level lies at  $E_F = E_C - kT \log(N_D/N_C)$ . At 100K, 200K, 300K it is 0.134eV, 0.286eV, and 0.444eV below the conduction band if  $N_D = 1.0 \cdot 10^{12} \text{cm}^{-3}$ .

### DLTS and non-exponential emission from a trap

We have seen above that the DLTS signal is the difference of capacitances measured at two fixed times after the end of the filling pulse. What would be the response of DLTS to a non-exponential carrier emission from a trap, especially to the kind observed here?

The capacitance transient reflects the change in the space charge density. It is therefore proportional to the transient current integrated with respect to time. If the time dependence of the transient current is described by  $I(t) = q \cdot \lambda \cdot f(\lambda t)$  then the DLTS signal is proportional to:

$$\begin{aligned} S(\lambda) &= q \int_{t_1}^{\infty} \lambda f(\lambda t) dt - q \int_{t_2}^{\infty} \lambda f(\lambda t) dt \\ &= Q(\lambda t_1) - Q(\lambda t_2) \end{aligned}$$

where  $Q$  denotes the integral of  $I(t)$ . To find the  $\hat{\lambda}$  at which  $S(\lambda)$  attains its maximum, differentiate the last of the above equations with respect to  $\lambda$  and equate the result to zero:

$$\begin{aligned} \left. \frac{d}{d\lambda} \right|_{\lambda=\hat{\lambda}} S(\lambda) &= 0 \\ 0 &= \hat{\lambda} t_1 f(\hat{\lambda} t_1) - \hat{\lambda} t_2 f(\hat{\lambda} t_2) \end{aligned}$$

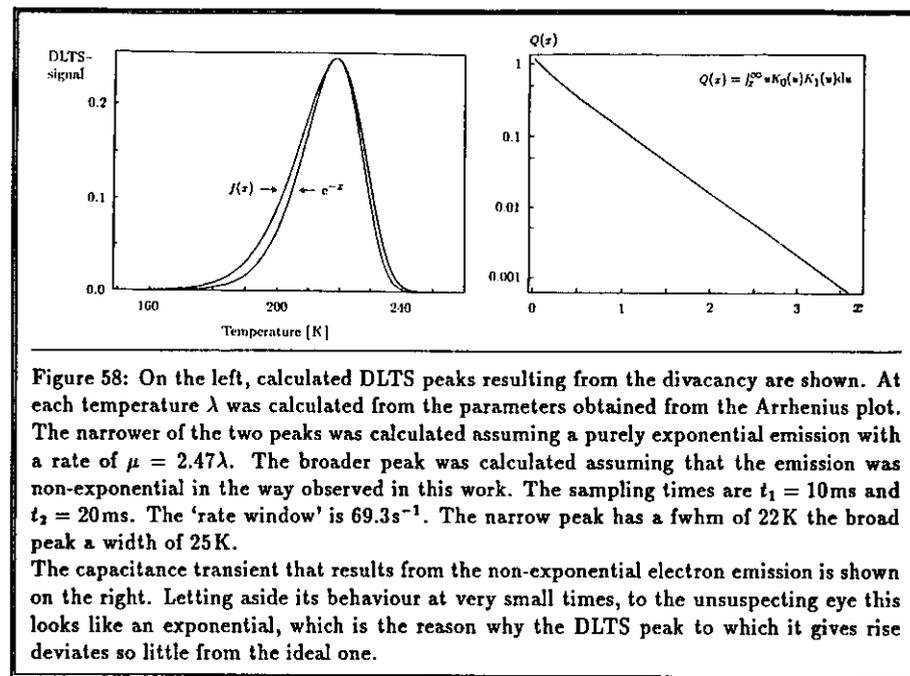


Figure 58: On the left, calculated DLTS peaks resulting from the divacancy are shown. At each temperature  $\lambda$  was calculated from the parameters obtained from the Arrhenius plot. The narrower of the two peaks was calculated assuming a purely exponential emission with a rate of  $\mu = 2.47\lambda$ . The broader peak was calculated assuming that the emission was non-exponential in the way observed in this work. The sampling times are  $t_1 = 10\text{ms}$  and  $t_2 = 20\text{ms}$ . The 'rate window' is  $69.3\text{s}^{-1}$ . The narrow peak has a fwhm of 22K the broad peak a width of 25K. The capacitance transient that results from the non-exponential electron emission is shown on the right. Letting aside its behaviour at very small times, to the unsuspecting eye this looks like an exponential, which is the reason why the DLTS peak to which it gives rise deviates so little from the ideal one.

And  $\hat{\lambda}$  depends on the function  $f(x)$  and the two times  $t_1, t_2$ . In the case of DLTS and exponential decays we had

$$\begin{aligned} I(t) &= q\lambda e^{-\lambda t} \\ \hat{\lambda} &= \frac{\log(t_2/t_1)}{t_2 - t_1} \end{aligned}$$

In our case  $I(t)$  is not exponential—and therefore the following result is surely surprising:

$$\begin{aligned} I(t) &= q\mu \cdot (\mu t) \cdot K_0(\mu t) \cdot K_1(\mu t) \\ \hat{\mu} &= \text{const} \cdot \frac{\log(t_2/t_1)}{t_2 - t_1} \end{aligned}$$

The constant in the last equation is not exactly a constant. As the ratio  $t_2/t_1$  varies from 1.0 to 100 it changes gradually from 0.406 to 0.403. By all practical means this is a constant, which we set to an average value of 0.405.

At big times,  $\mu t > 1$ , the current approaches its asymptotic behaviour  $I(t) \propto \exp(-2\mu t)$ . We should thus compare  $2\hat{\mu}$  to the DLTS value  $\hat{\lambda}$ :

$$\left( \begin{array}{l} 2\hat{\mu} \\ \text{(this work's limit} \\ \text{emission rate)} \end{array} \right) = 0.81 \cdot \left( \begin{array}{l} 0.810 \cdot \hat{\lambda} \\ \text{simple exponential} \\ \text{emission rate} \end{array} \right)$$

And this relationship is independent of the particular ratio  $t_2/t_1$  used in the DLTS measurement.

This especially implies that using DLTS only, it cannot be decided whether the transient is exponential or not. Thurber et al. have suggested a test on exponentiality, [Th 82]. The idea is to do DLTS runs with a set of different ratios  $t_2/t_1$  while  $t_1$  is always adjusted so as to keep constant the rate window. If the peak in question keeps showing up at the same temperature, it is concluded that the transient is exponential. We see now that a function that deviates considerably from an exponential can as well pass this test.

What is important for this work is that the electron emission from the  $VV^-$  is non-exponential in a very special way which escapes detection by DLTS and allows us to compare the findings of this work to those obtained with DLTS. The impact that the non-exponential transients have on a DLTS measurement is illustrated in figure 58.

---

## 7 Summary

The performance of the silicon-detector based hadron-electron separator of ZEUS has already been summarized at the end of chapter two, and this concluding section shall concentrate on the methodology of point defect characterization and the related physics insights. A compilation of the numerical results follows at the end.

### 7.1 Resumé

The motivation to study crystal defects has been the severe radiation damage that silicon detectors will experience in experiments at today's and tomorrow's particle accelerators. Precise knowledge of the identity, rate of formation, and the electric characteristics of the radiation-induced crystal defects is a necessary prerequisite for predicting the performance deterioration of silicon detectors in any experiment.

Except for structural information most of our knowledge about crystal defects comes from experiments in which the thermally activated emission of captured charge carriers is observed. This usually yields the emission rates and the activation energies but not the entropies which are needed to complete the thermodynamic description. Moreover, the methods provide no way to tell a donor from an acceptor nor do they allow insights into how the emission proceeds with time. All of them are applicable only if the defect concentration is small compared to the doping with shallow acceptors or donors.

To overcome these shortcomings a new technique has been developed and is presented in this work. The new method is based on recording in real time the transient electric current brought about by the emission of charge carriers from a defect and has therefore been named current-transient recording, or CTR for short.

It is well suited to work on low-doped, fully-depletable silicon detectors. With no bias applied only a small fraction of the detector volume is depleted of mobile carriers while the rest is not. The undepleted part is in thermal equilibrium and all electron states below the Fermi level, which is in the upper half of the bandgap, are filled. In the depletion zone only those electron states below the intrinsic level, which is at the bandgap centre, are filled. When the detector is pulsed into full depletion, electron states that have been between the Fermi level and the intrinsic level will discharge, causing an electric current to flow. It is sensed by a fast current-to-voltage converter whose output is routed to a digital oscilloscope where the transient is stored. The oscilloscope is read out by a computer and the transient is thus available for later analysis. The set of current transients gathered during a temperature scan forms the raw data of a CTR experiment.

Like the main body of experimental techniques, CTR observes thermally activated charge carrier emission, but it overcomes various limitations. It knows no high-concentration limit which makes it a suitable means to study high-level radiation damage in low-doped silicon detectors. It allows to measure the occupancy of a defect level as a function of temperature from which the defect type (donor or acceptor) and its ionization entropy emerge. Moreover, the emission process is followed and recorded in all detail.

CTR has been used successfully to study the radiation damage caused by fast neutrons and by electromagnetic showers developed in a lead-scintillator calorimeter.

It was found that the current transients did not follow an exponential law. It turned out that over a temperature range of more than 100°C the shape of the transients did not change, which led to the conclusion that there is only one type of defect active. The model function

to describe the invariant transient shape was found by trial and error, and for each individual transient an amplitude and a velocity parameter were found by least-squares fits.

As a function of temperature the velocity parameter followed an Arrhenius law over all the accessible temperature range from  $-70^{\circ}\text{C}$  to  $+70^{\circ}\text{C}$ . At short times the transients are strongly non-exponential, varying in proportion to the logarithm of time, but at big times they tend to an exponential function. The absolute values of the limiting (exponential) emission rate as well as the activation energy obtained from the Arrhenius law match very well the generally accepted findings for the first acceptor level of the divacancy.

That the defect is indeed an acceptor is proven by the fact that the measured concentration of discharging defects attains, at low temperatures and high irradiation doses, exactly the value of the  $n$ -type doping, but does not exceed it. For acceptor-like defects this is to be expected since it is impossible to have more electrons trapped at the acceptors than are provided by the dopants. If the defect was donor-like, being neutral with its electron attached, such a behaviour would be inexplicable. It also shows clearly that the dopant density was not altered as a result of the irradiation, ie there is no substantial production of phosphorus-vacancy complexes.

The temperature dependence of the defect occupancy could be described to an accuracy of 1% using the theory developed in this work. The reaction entropy of the electron emission was taken as unknown and adjusted to fit the prediction to the data.

The somewhat unexpected non-exponential behaviour of electron emission from the divacancies may have a simple reason. The actual damage following neutron irradiation (photoneutrons in the case of electron showers) is due to energetic silicon recoils, which produce isolated divacancies at the beginning of their path but a cluster of close-by defects at the very end. Depending on how close to such a cluster a divacancy is produced it experiences varying degrees of lattice strain to which the energy level of the charged divacancy is known to respond. The non-exponential emission characteristic could thus be brought about by the superposition of a continuum of exponential processes with different rates.

The macroscopic radiation damage could not be related to the divacancy. The dark-current is not proportional to the divacancy concentration and its activation energy also does not fit. There is a negative space charge introduced into the depletion zone of the diode, in proportion to the dark current, while the divacancy is neutral in the depletion zone. The activation energy of the dark current of  $0.65\text{eV}$  and the negative space charge clearly spell out the presence of a midband defect level located just below the bandgap centre. However, such a defect has never been observed in DLTS studies and there is a remarkable blindness of CTR with respect to it; all the transients are ascribable to the sole action of the divacancy. The presence of a midband defect following neutron irradiation has been known for three decades. In this light it is astonishing that its failure to show up in DLTS studies is only rarely commented in literature.

Since it is inconceivable that the defect clusters which are produced by silicon recoils should remain electrically inactive it is tempting to see them as the cause of the macroscopic damage. An estimate of the cluster introduction rate matches well the measured values for the negative space charge.

When comparing measurements made two days after the irradiation to those made 100 days later one finds that the amount of negative space charge has not changed while the dark current has reduced by a factor of 2.5, with no change of its activation energy. This needs not contradict the hypothesis since the rate of carrier emission from a defect is proportional to its capture cross-section, which is related to the size of the defect. After their formation defect clusters are known to shrink until stable conditions are arrived at. The apparent room temperature 'annealing' of the dark current may simply reflect the cluster shrinking.

## 7.2 Quantitative results

### The specimen

The diodes were of  $p^+nn^+$  type, had  $9.0\text{cm}^2$  cross-section area and a thickness of  $270\mu\text{m}$ . The  $n$ -type bulk doping was  $0.90 \dots 1.0 \cdot 10^{12}\text{cm}^{-3}$ .

One has been irradiated with  $14\text{MeV}$  neutrons from a deuterium-tritium source to a fluence of  $0.72 \cdot 10^{12}\text{cm}^{-2}$  within 90 minutes. Three diodes have been exposed to electromagnetic showers generated by  $2\text{GeV}$  bremsstrahlung gammas in a lead-scintillator calorimeter. They were positioned behind the shower maximum. Within five hours doses of  $5.4\text{kGy}$ ,  $10\text{kGy}$ ,  $16\text{kGy}$  as calculated with EGS were accumulated.

### Phenomenological findings

The dark current (at constant temperature) increased in proportion to the received fluence or dose. Over a period of  $\sim 100$  days the dark currents reduced to 40% of the immediate post irradiation value in the case of neutron irradiation and to 65% of the value measured four weeks after the irradiation in the case of exposure to electromagnetic showers. Measured two years after the exposure, the rates of introduction of dark current are at  $20^{\circ}\text{C}$

$$\begin{aligned} 14\text{MeV neutrons:} & \quad \frac{\Delta I}{\text{volume} \times \text{fluence}} = 4.6 \cdot 10^{-17}\text{A cm} \\ \text{e.m. showers:} & \quad \frac{\Delta I}{\text{volume} \times \text{dose}} = (3.3 \pm 0.3)\text{nA/cm}^3\text{Gy} \end{aligned}$$

The temperature dependence of the dark current obeys the same Arrhenius law in all cases:

$$I(T) \propto T^2 \exp\left(-\frac{A}{kT}\right) \quad \text{with} \quad A = (0.65 \pm 0.01)\text{eV}$$

The irradiation adds a negative space charge density ( $\rho$ ) in the depletion region of the diodes, which is proportional to the increase in dark current. At  $300\text{K}$ , the dark current per  $10^{12}$  of these defects is  $130\mu\text{A}$ , ie at each defect electron-hole pairs are being produced at a rate of  $820/\text{s}$ . The defect introduction rates are

$$\begin{aligned} 14\text{MeV neutrons:} & \quad \frac{\Delta \rho}{\text{fluence}} = 0.63\text{cm}^{-1} \\ \text{e.m. showers:} & \quad \frac{\Delta \rho}{\text{dose}} = (4.8 \pm 0.2) \cdot 10^7\text{cm}^{-3}\text{Gy}^{-1} \end{aligned}$$

### Characteristics of the divacancy

In all four diodes, the trap responsible for the measured transients has been identified as the divacancy. The current transients that are brought about by electron emission from the singly-negative charge state were found to be non-exponential at small times, but tending to an exponential at big times. The evolution in time can, at any temperature, be described by a single function  $f(\lambda t)$ , which for big times is asymptotically equal to  $\exp(-2\lambda t)$ . Of the divacancy, the following quantities could be determined:

- As a function of temperature,  $\lambda$  obeys an Arrhenius law of the form:

$$\frac{1}{\lambda} = \tau_0 \left( \frac{300}{kT} \right)^2 \exp \left( -\frac{A}{kT} \right)$$

The activation energy is  $A = (0.4216 \pm 0.0007) \text{eV}$

The prefactor is  $\tau_0 = (3.6 \pm 0.1) \cdot 10^{-12} \text{s}$

- The reaction entropy for electron emission is  $S_{CT} = (-1.05 \pm 0.05) \text{k}$ .  
 $S_{CT}$  was found from the temperature dependence of the occupancy of the (0/-) level.
- The production rate of divacancies in weakly n-doped silicon was found to be:

$$\begin{aligned} \eta &= 1.0 \text{cm}^{-1} && \text{for 14 MeV neutrons} \\ \eta &= 1.6 \cdot 10^6 \text{cm}^{-3} \text{Gy}^{-1} && \text{for e.m. showers developed in lead} \end{aligned}$$

## A Symbols and their Meanings

### Energy Levels

#### Free Energies

Free energies are denoted by a capital  $E$ . When divided by  $kT$ , they are denoted by  $\xi$ .

$E_C$  conduction band energy

$E_V$  valence band energy

$E_C - E_V = E_{CV}$  free energy of the bandgap, 1.12 eV @ 300 K

$E_i$  intrinsic level

$E_F$  Fermi level

$E_T$  trap energy level

#### Enthalpies

Reaction enthalpies are denoted by a capital  $H$ .

$H_C - H_V = H_{CV}$  bandgap enthalpy, 1.201 eV @ 300 K

#### Entropies

Reaction entropies are denoted by a capital  $S$ .

$S_C - S_V = S_{CV}$  bandgap entropy, 2.96 k @ 300 K

### Distances

$d$  diode thickness

$w$  field zone thickness

### Voltages

$V_{bi}$  built-in voltage, diffusion voltage

$V_d$  depletion voltage

$V_o$  applied voltage

$V_s$  overbias (excess) voltage

$V_t$  total voltage drop across the  $pn$  junction

$V_{op}$  operating voltage

$V_T = kT/e$  thermal voltage

## Concentrations

$N_C$  conduction band density of states,  $N_C = 2.91 \cdot 10^{19} \text{cm}^{-3}$  at 300 K,  $N_C \propto T^{3/2}$ .

$N_V$  valence band density of states,  $N_V = 1.06 \cdot 10^{19} \text{cm}^{-3}$  at 300 K,  $N_V \propto T^{3/2}$ .

$N_D$  dopant-donor concentration

$N_A$  dopant-acceptor concentration

$N_T$  deep-trap concentration

$D$  deep-donor concentration

$A$  deep-acceptor concentration

$n_i$  intrinsic density of charged carriers

$$n_i = 3.56 \cdot 10^{20} \left( \frac{T}{300 \text{ K}} \right)^{3/2} \exp \left( -\frac{1.206 \text{ eV}}{2kT} \right)$$

$n$  electron concentration

$p$  hole concentration

## Mass-action-law constants

$K_n$  electron equilibrium constant  $K_n = N_C \exp(E_T - E_C)/kT = n_i \exp(E_T - E_i)/kT$

$K_p$  hole equilibrium constant  $K_p = N_V \exp(E_V - E_T)/kT = n_i \exp(E_i - E_T)/kT$

## Trap Charge States

$T^{++}, T^+, T^0, T^-, T^-$  doubly positive, singly positive, neutral, singly negative, doubly negative state of the trap  $T$

$T_0$  Trap state with the least number electrons which is of interest or concentration thereof

$T_1$  Trap state with one more electron than  $T_0$  or concentration thereof

## Generation-Recombination Statistics

$e_n, e_p$  electron, hole emission rate (from a trap)

$r_n, r_p$  electron, hole capture coefficient (by a trap)

$r_A$  Auger electron-hole recombination coefficient,  $r_A = 1.66 \cdot 10^{-30} \text{cm}^6/\text{s}$  ( $\pm 15\%$ )

$G_n, G_p$  net electron, hole generation rate per unit volume

$\sigma_n, \sigma_p$  electron, hole capture cross section

$v_{th}$  thermal velocity,  $v_{th} = \sqrt{3kT/m^*}$

## References

- [A188] A.Ali et al., DESY 88-119 (1988), Hamburg
- [Am 84] W.von Ammon, H.Herzer, Nucl. Instr. and Meth. **226**, (1984) 94
- [Ap 79] W.D.Apel et al., CERN-EP/79-18 (1979), Geneva
- [Ap 87] V.D.Apokin et al., Sov. J. of Nucl. Phys. **46** (5) (1987)
- [APS 92] The American Physical Society, Phys. Rev. D **45**(11) Part II, 'Review of Particle Properties', (1992)
- [As 87] M.T.Asom et al., Appl. Phys. Lett. **51** (4) (1987) 256
- [Aw 88] O.O.Awaldekarim, B.Monemar, Journ. of Appl. Phys. **64**(11) (1988) 6301
- [Aw 89] O.O.Awaldekarim, B.Monemar, Journ. of Appl. Phys. **65**(12) (1989) 4779
- [Ba 86] S.K.Bains and P.C.Bansbury, Semicond. Sci. Technol. **2** (1987) 20
- [Bo 88] V.G.Bogdanov et al., Sov. J. of Nucl. Phys. **47**(6) (1988) 1148
- [Bo 91] E.Borchi et al., Nucl. Instr. and Meth. **A 310** (1991) 273
- [Bö 87] Appl. Phys. **A43** (1987) 29
- [Br 54] Brinkmann, Journ. of Appl. Phys. **25** (1954) 961
- [Br 82] S.D.Brotherton, P.Bradley, Journ. of Appl. Phys. **53**(8) (1982) 5720
- [Br 83] S.D.Brotherton et al., Journ. of Appl. Phys. **54**(9) (1983) 5113
- [Bu 83] M.J.Buckingham, Noise in Electronic Devices & Systems, Ellis Horwood Ltd., England (1983)
- [CHARM II] CHARM II collaboration, CERN-EP/88-87 (1988), Geneva and San Miniato Conf. 7-11 March 1988, Italy
- [CDF 90] CDF Collab., presented by K.Sliwa at XXVth Recontres de Moriond, Les Arcs, France, March 4-18, (1990) and FERMILAB-Conf.90/93-E, [E741/CDF] (1990)
- [Co 61] J.W.Corbett, G.D.Watkins, Phys.Rev.Lett. **7** (1961) 314
- [Co 65] J.W.Corbett, G.D.Watkins, Phys. Rev. **138**(2A) (1965) A 555
- [Di 87] H.Dietl et al., Nucl. Instr. and Meth. **A 253** (1987) 460
- [Do 58] G.Doetsch, Einführung in Theorie und Anwendung der Laplacetransformation, Birkhäuser Verlag Basel und Stuttgart (1958).
- [Dw 88] A.Dwurazny et al., DESY 88-70 (1988), Hamburg
- [Dw 88a] A.Dwurazny et al., private communication

- [Dy 91] F.Dydak, (Rapporteur), Conf. at Singapore, 2-8 Aug. 1990, CERN-PPE/91-14 (1991), Geneva
- [Ea 82] W.T.Eadie, D.Drijard, *Statistical Methods in Experimental Physics*, North-Holland Publishing, Amsterdam, (1971)
- [Ed 91] M.Edwards et al., *Nucl. Instr. and Meth. A* **310** (1991) 283
- [El 89] J.Ellis, G.L.Fogli, *Phys. Lett. B* **232**(1) (1989) 139
- [Ev 76] A.O.Evwaraye, E.Sun, *Journ. of Appl. Phys.* **47**(9) (1976) 3776
- [Fa 68] R.A.Faulkner, *Phys.Rev.* **175** (1968) 991
- [Fa 89] P.M.Fahey et al., *Rev. of Mod. Phys.* **61**(2) (1989) 289
- [Fr 90] E.Fretwurst et al., *Nucl. Instr. and Meth. A* **288** (1990) 1
- [Fr 90a] F.H.Fröhner, Kernforschungszentrum Karlsruhe, KfK 4655 (1990) and F.R.Fröhner, invited paper at IX. Summer School on Nuclear Physics..., Varna, Bulgaria, 28.Sept.-7.Oct. 1989
- [Fu 90] A.Fürtjes, DESY F35-90-02 (1990), Diplomarbeit, University Hamburg, internal communication
- [Gi 80] J.F.Gibbons, Ion Implantation, in *Handbook on Semiconductors Vol. 3*, 599-640, edited by T.S.Moss, North Holland Publishing (1980)
- [Gr 87] P.Grafström et al., SPS/EBS/Note 87-7 (1987) CERN, Geneva
- [H1 89] H1 Calorimeter Group, DESY 89-022 (1989), Hamburg
- [Ha 87] R.D.Harris et al., *Phys. Rev. B* **36**(2) (1987) 1094
- [Ha 89] M.Hasegawa et al., *Nucl. Instr. and Meth. A* **277** (1989) 395
- [Hi 88] H.Hirayama, W.R.Nelson, SSC-SR-1033, Appendix 7, Lawrence Berkeley Lab, (1988) and the erratum to this: SSCL-285 (1990)
- [Ho 80] L.M.Howe et al., *Nucl. Instr. and Meth.* **170** (1980) 419
- [Hü 89] M.W.Hüppi, *Journ. of Appl. Phys.* **68**(6) (1990) 2702
- [ICRP 73] ICRP report no. 21, Pergamon Press Ltd., Oxford (1973)
- [In 87] G.Ingelman et al., Proc. of the HERA workshop, DESY, Hamburg, Oct. 12-14 (1987)
- [Ja 75] J.D.Jackson, *Classical Electrodynamics*, 2nd ed., John Wiley and sons, New York (1979)
- [Ja 84] P.Jarron, M.Goyot, *Nucl. Instr. and Meth. A* **226** (1984) 156
- [Ja 85] D.B.Jackson, C.T.Sah, *Journ. of Appl. Phys.* **58**(6) (1985) 2225 and *Journ. of Appl. Phys.* **58**(3) (1985) 1270
- [Jo 70] N.L.Johnson, S.Kotz, *Continuous Univariate Distributions Vol. 1*, John Wiley & Sons, New York (1970)
- [Ke 84] J.Kemmer, *Nucl. Instr. and Meth.* **226** (1984) 89
- [Kr 89] H.W.Kraner et al., *Nucl. Instr. and Meth. A* **279** (1989) 266
- [Kr 90] H.Kröger, Diplomarbeit, University of Hamburg, 1990
- [La 74] D.V.Lang, *Journ. of Appl. Phys.* **45** (7) (1974) 3023
- [La 80] P.T.Landsberg, *Semiconductor Statistics*, in T.S.Moss (Ed.), *Handbook on Semiconductors*, North Holland Publishing (1980)
- [La 83] M.Lannoo, J.Bourgoin, *Point Defects in Semiconductors*, Springer Series in Solid-State Sciences nos. 22 and 35, (1981) (1983), Springer-Verlag Berlin
- [La 87] P.T.Landsberg, E.H.Shaban, *Journ. of Appl. Phys.* **61**(11) (1987) 5055
- [Li 90] G.Lindström et al., *Nucl. Instr. and Meth. A* **288** (1990) 1
- [Li 89] G.Lindström et al., DESY 89-105, (1989), and 'Workshop on Calorimetry for the SSC', Tuscaloosa/Alabama, March 1989
- [Li 91] Z.Li, W.Chen, and H.W.Kraner, *Nucl. Instr. and Meth. A* **308** (1991) 585
- [Lo 87] E.Lohrmann, ZEUS Note 87-053 (1987), DESY, Hamburg
- [Lo 90] A.Longoni, *Nucl. Instr. and Meth. A* **288** (1990) 35
- [McL 88] V.McLane et al., *Neutron Cross Section Curves*, National Nuclear Data Center, Brookhaven National Laboratory, Upton, New York, Academic Press, New York (1988)
- [Me 83] E.Meijer, H.G.Grimmeis, *Journ. of Appl. Phys.* **55**(12) (1984) 4266
- [Mo 83] P.M.Mooney, *Journ. of Appl. Phys.* **54**(1) (1983) 208
- [NA31 88] NA31 Collab., H.Burkhardt et al., *Phys. Lett. B* **206** (1988) 169
- [Na 88] M.Nakamura et al., *Nucl. Instr. and Meth. A* **270** (1988) 42
- [Oh 88] T.Ohsugi et al., *Nucl. Instr. and Meth. A* **265** (1988) 105
- [Ou 90] A.M.Ougouag et al., *IEEE Transact.Nucl.Sci.* **37**(6) (1990) 2219
- [Pa 68] W.K.H.Panofsky, Proc. of the Int. Conf. on High Energy Physics, Vienna 1968
- [Pa 79] B.Pajot et al., *Solid State Comm.* **31** (1979) 759
- [Pe 87] D.H.Perkins, *Introduction to High Energy Physics*, 2nd ed., Addison-Wesley Publishing Company, (1982)
- [Pl 80] W.A.Pliskin et al., *Passivation and Insulation in Handbook on Semiconductors Vol. 3*, 641-686, edited by T.S.Moss, North Holland Publishing (1980)

- [Po 84] D.Pons, Journ. of Appl. Phys. 55(10) (1984) 3644
- [Ru 66] G.Rudstam, Z. Naturforschg. 21a (1966) 1027
- [Sc 82] M.Scheffler, in Festkörperprobleme XXII (1982) pp 115, Ed. P.Grosse, Vieweg Verlag 1982
- [Si 90] E.G.Sieverts et al., Phys. Rev. B 41(13) (1990) 8630
- [Si 76] B.Sigfridsson et al., Journ. of Appl. Phys. Lett. 47 (10) (1976) 4611
- [SR 89] ZEUS Status Report, (1989), DESY, Hamburg
- [Sv 87] B.G.Svensson et al., Journ. of Appl. Phys. 62(7) (1987) 2758
- [Sv 88] J.H.Svensson et al., Phys. Rev. B 38(6) (1988) 4192
- [Sv 91] B.G.Svensson et al., Phys. Rev. B 43(3) (1991) 2292
- [Te 90] K.Tesch, J.M.Zazula, DESY HERA-90-18, (1990)
- [Th 82] W.R.Thurber et al., Journ. of Appl. Phys. 53(11) (1982) 7397
- [Ve 76] J.A.VanVechten, Phys. Rev. B 14(8) (1976) 3539
- [Ve 80] J.A.VanVechten, A Simple Man's View of the Thermochemistry of Semiconductors, in T.S.Moss (Ed.), Handbook on Semiconductors, North Holland Publishing (1980)
- [Wa 40] G.N.Watson, A Treatise on the theory of Bessel Functions, 2nd ed., The Syndics of the Cambridge University Press, London, (1944)
- [Wa 73] J.W.Walker, Phys. Rev. B 7(10) (1973) 4587
- [Wa 84] G.D.Watkins, in Festkörperprobleme XXIV (1984) 163, Ed. P.Grosse, Vieweg Verlag 1984
- [Wa 90] F.P.Wang et al., Journ. of Appl. Phys. 68(4) (1990) 1535
- [Wi 87] R.Wigmans, Nucl. Instr. and Meth. A 259 (1987) 389
- [Wo 91] R. van Woudenberg et al., in Proceedings of the Workshop 'Physics at HERA', Vol.2, pp. 739, ed. by W.Buchmüller and G.Engelmann, Oct. 1991, Hamburg
- [Wu 91] R.Wunstorff et al., Nuclear Physics B (Proc. Suppl.) 23 A (1991) 324
- [ZEUS 86] The Zeus Detector, Technical Proposal (1986), DESY, Hamburg
- [Zi 75] A. van der Ziel, Solid State Electr. 18 (1975) 969
- [Zi 89] E. Di Zitti et al., Journ. of Appl. Phys. 66(3) (1989) 1199
- [Zu 88] R.Zurmühl und S.Falk, Matrizen und ihre Anwendungen, 5te Auflage, Springer Verlag, Berlin; Teil 1 (1984) und Teil 2 (1986)

## Acknowledgements

It is my pleasant duty to here record my gratitude to Messr Prof. Lohrmann, Dr Wolf, and Dr Klanner who made this work possible in the first place and generously offered any support throughout the whole project.

Especially during its first phase the warm hospitality of the ZEUS-HES group is acknowledged and thanks are due to all colleagues I was fortunate to work with at the time. I am very much indebted to Dr Koch, who most of all let the experimentation at DESY and at CERN become excitement and adventure. From the beginning to the end, numerous discussions with him, rehearsals of talks I had to give, and his painstakingly careful reading of this work's manuscript have contributed to no small extent to the successful finishing of this thesis.

The second part of this work, which is concerned with radiation damage, could not have been begun had it not been for the encouragement and trust put in me by Messr E.Lohrmann, R.Klanner and A.Seidmann. And without their continuing support of the idea and considerable patience on their side, it could not have been brought to a good end.

I am also grateful to Mr Lenzen, who was constantly being of invaluable help in administrative and technical matters.

Special thanks are due to Mr Wierba, who designed the electronics of the HES prototype, and who afterwards came up with a positive answer to almost any question I had. Especially the idea to use a transimpedance converter in the CTR instrumentation is his contribution.

Thanks are due to Mr Dwurazny for performing two of the irradiations, the one with <sup>60</sup>Co at Krakow and the other with electromagnetic showers at Bonn.

