The Effect of Mechanical Stress on Bipolar Transistor Characteristics

J.F. Creemer

Stellingen

behorende bij het proefschrift

The Effect of Mechanical Stress on Bipolar Transistor Characteristics

Het effect van mechanische spanning
op de karakteristieken van
bipolaire transistors

Jan Fredrik CREEMER

Delft

14 januari 2002

1. In monokristallijn silicium is het piëzoresistieve effect veel sterker dan in metalen. Dit komt doordat het kristalrooster van silicium veel symmetrischer is en minder goed geleidt.

2. De deformatiepotentialen van silicium zijn al vijftig jaar het onderwerp van publicaties. Het waarom hiervan wordt duidelijk zodra men de potentialen gebruikt voor het voorspellen van metingen.

3. Leerboeken over de fysica van halfgeleidercomponenten zouden minder aandacht moeten besteden aan de drift-diffusievergelijkingen en meer aan de elektrochemische potentiaal.

4. Transistors hebben met mensen gemeen dat ze beter presteren bij een beetje stress, maar doorslaan bij teveel.

5. Als wiskunde net zoveel redundantie bevatte als taal, dan zouden lange afleidingen veel vaker tot correcte conclusies leiden.

6. In de politiek kunnen metingen op veel meer manieren worden uitgelegd dan in de wetenschap.

7. Het feit dat er van universitaire docenten geen didactische opleiding wordt gevraagd, ontslaat hen niet van de morele verplichting om die te hebben.

8. Het geloof in een maakbare samenleving is de laatste twintig jaar omgeslagen in een geloof in een maakbare levensloop.

9. Moderne kunst zou minder sceptisch worden bekeken als de makers af en toe een portret creëerden.

10. Ons voedsel is goedkoper dan goed voor ons is.

11. Een leuke eigenschap van computers is dat je er op jonge leeftijd al nostalgisch over kunt praten.

1. In monocrystalline silicon the piezoresistive effect is much larger than in metals. Essentially this is because the crystal lattice is more symmetric and conducts less well.

2. The deformation potentials of silicon have been the subject of publications for fifty years. The reason why becomes clear when one uses the potentials to predict measurements.

3. Text books on the physics of semiconductor devices should pay less attention to the drift diffusion equations and more to the electrochemical potential.

4. Transistors have in common with human beings that they perform better under some stress, but break down at high levels.

5. If mathematics contained as much redundancy as language, long derivations would more often lead to correct conclusions.

6. In politics, measurements can be explained in much more ways than in science.

7. The fact that university professors are not obliged to have any didactic training does not release them from the moral obligation to have it.

8. In the past twenty years, the belief that one can consciously shape one's society has changed into the belief that one can consciously shape one's life.

9. Modern art would be considered less sceptically if the makers occasionally created a portrait.

10. Our food is too cheap for our own good.

11. A nice property of computers is that one can discuss them nostalgically already at an early age.

The Effect of Mechanical Stress on Bipolar Transistor Characteristics

The Effect of Mechanical Stress on Bipolar Transistor Characteristics

Het effect van mechanische spanning
op de karakteristieken van
bipolaire transistors

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. F. Wakker,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 14 januari 2002 om 16.00 uur

door

Jan Fredrik CREEMER

elektrotechnisch ingenieur
geboren te Amsterdam.

Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. P. J. French
Prof. dr. ir. S. Middelhoek

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. P. J. French,	Technische Universiteit Delft, promotor
Prof. dr. ir. S. Middelhoek,	Technische Universiteit Delft, promotor
Prof. Y. Kanda, Dr. Sc.,	Toyo University, Japan
Prof. dr. rer. nat. G. Wachutka,	Technische Universität München, Duitsland
Prof. dr. A. G. R. Evans,	University of Southampton, Engeland
Dr. ir. G. C. M. Meijer,	Technische Universiteit Delft
Dr. ir. M. G. Middelhoek,	SystematIC design Delft

ISBN 90 5166 883 X

Uitgeverij Eburon
Postbus 2867
2601 CW Delft
info@eburon.nl
www.eburon.nl

Key words: bipolar transistor, stress, piezojunction effect, piezoresistance, minority carriers

Contents

1 Introduction **1**

1.1 Main characteristic of the bipolar transistor 1
1.2 Shifts in characteristics due to stress 2
1.3 Consequences for circuits 3
1.4 Consequences for sensors 4
1.5 Previous research on stress effects 5
1.6 Transistor modelling for design purposes 6
1.7 Scope of this thesis 7
1.8 Organisation of this thesis 8

2 Theory of Stress-Induced Changes **17**

2.1 Deformation of the energy bands 17
2.1.1 The $\mathbf{k}\hat{\mathbf{p}}$ method for deformed cubic crystals 19
2.1.2 The strain Hamiltonian 20
2.1.3 Strain matrix elements 22
2.1.4 Conduction band equations 24
2.1.5 Valence band equations 28
2.1.6 Bandgap definitions 36
2.1.7 Band edge diagrams 36
2.2 Microscopic changes in charge transport 39
2.2.1 Available transport models 39
2.2.2 Charge carrier concentrations 42
2.2.3 Electrical conductivity 51
2.2.4 Auxiliary relations 58
2.3 Macroscopic changes in charge transport 61
2.3.1 Phenomenological definition of conductivity 61
2.3.2 Piezoconductance 64
2.3.3 Piezoresistive and piezojunction effects 66
2.3.4 Evaluation of the first-order coefficients 69
2.3.5 Interpretation of the piezocoefficients 71
2.4 Modification of device characteristics 74
2.4.1 General assumptions 74
2.4.2 Resistors 75
2.4.3 Bipolar transistors 78

2.5 Conclusions 86

3 Measurements of Stress-induced Changes **99**
3.1 Principle of the measurement set-up 99
3.1.1 Devices and their stress dependence 100
3.1.2 Stress generation 104
3.1.3 Electrical characterization 106
3.2 Implementation of the measurement set-up 108
3.2.1 Overview 108
3.2.2 Devices under test 109
3.2.3 Beam carrier 114
3.2.4 Bending apparatus 120
3.2.5 Source-measurement units 124
3.3 Operation of the measurement set-up 125
3.3.1 Measurement procedure 125
3.3.2 Data processing 126
3.3.3 Model identification 126
3.4 Measurement results 128
3.4.1 Zero-stress operation 128
3.4.2 Stress-induced transistor changes 130
3.4.3 Piezojunction coefficients 131
3.4.4 Stress-induced resistor changes 136
3.4.5 Piezoresistive coefficients 137
3.5 Conclusions 139

4 Discussion and Conclusions **143**
4.1 Validity of the microscopic model 144
4.2 Validity of the macroscopic model 148
4.3 Validity of the measurements 150
4.4 Opportunities for electronic designers 153
4.5 Possible future work 155
4.6 Conclusions 157

A Stress, Strain, and Elasticity Tensors **161**
A.1 Stress definitions 161
A.2 Strain definitions 162
A.3 Tensor properties 164
A.3.1 Einstein summation convention 164
A.3.2 Scalar description in an anisotropic case 165
A.3.3 Transformation of axes 165
A.3.4 Principal values 166
A.4 Elasticity 167
A.5 Conclusions 168

B Band Structure of Cubic Semiconductors **171**
B.1 Overview of the $\mathbf{k}\hat{\mathbf{p}}$-method . . . 172
B.2 The $\mathbf{k}\hat{\mathbf{p}}$ Hamiltonian . . . 173
B.3 Löwdin perturbation theory . . . 174
B.4 Cubic symmetry . . . 176
B.5 Conduction band equations . . . 178
B.6 Valence band equations without spin . . . 182
B.7 Effects of spin-orbit coupling . . . 186
B.8 Valence band equations including spin . . . 190
B.9 Conclusions . . . 192

C Charge Transport in Stressed Silicon **197**
C.1 Charge concentration in nonequilibrium . . . 198
C.1.1 The relaxation time approximation . . . 198
C.1.2 Concentration integrals . . . 199
C.1.3 Effective density of states . . . 200
C.1.4 Density of states mass . . . 201
C.2 Charge transport equations . . . 203
C.2.1 Solution of the Boltzmann equation . . . 203
C.2.2 Current density and conductivity . . . 204
C.2.3 Mobility . . . 206
C.2.4 Generalised drift-diffusion equations . . . 207
C.3 Auxiliary relations . . . 209
C.3.1 The *pn* product . . . 209
C.3.2 Continuity equations . . . 210
C.3.3 The Poisson equation . . . 212
C.4 Conclusions . . . 212

D Current through a Bipolar Transistor **215**
D.1 The one-dimensional approximation . . . 216
D.2 Assumptions . . . 217
D.3 Current densities in the base . . . 219
D.4 Relation with the terminal voltages . . . 220
D.5 Collector and saturation current . . . 222
D.6 Conclusions . . . 222

List of symbols **225**

Acknowledgements **231**

About the Author **233**

List of Publications **235**

Summary **237**

Samenvatting **239**

Chapter 1

Introduction

Since its invention in 1947 the bipolar transistor has caused a revolution in signal processing. A major force behind this has been the enormous development of semiconductor technology. Due to this technology the dimensions of the transistor could shrink by more than a factor ten thousand. As a result, radios and other electronic equipment could be made much cheaper, more complex and lighter.

Semiconductor technology has also been employed to improve signal acquisition. Especially since 1980 this has resulted in micromachined sensors which are small and inexpensive and which are often integrated with standard electronics. Examples of this are the micromachined pressure sensors and the accelerometers used in the air bags of cars.

Technological developments have led to the spectacular improvement of the MOS transistor. Since 1980 it has taken over the digital signal-processing task from the bipolar transistor, which has led to an enormous improvement in computers. Nevertheless, the use of the bipolar transistor is still appreciated for analog signal processing due to its excellent analog characteristics.

1.1 Main characteristic of the bipolar transistor

The most valuable characteristic of a bipolar transistor is its amplification of a base-emitter voltage V_{be} to a collector current I_c. For forward bias this is expressed in terms of the well-known relationship

$$I_c = I_S \exp\left(\frac{qV_{be}}{k_B T}\right) \tag{1.1}$$

in which I_S is the saturation current, k_B the Boltzmann constant, q the elementary charge, and T the absolute temperature. This relationship is powerful due to its exponential nature. In addition, it is valid for many decades of magnitude of I_c. Finally, it yields a small-signal transconductance g_m which can be manipulated very accurately since it only depends on I_c and T. This can be

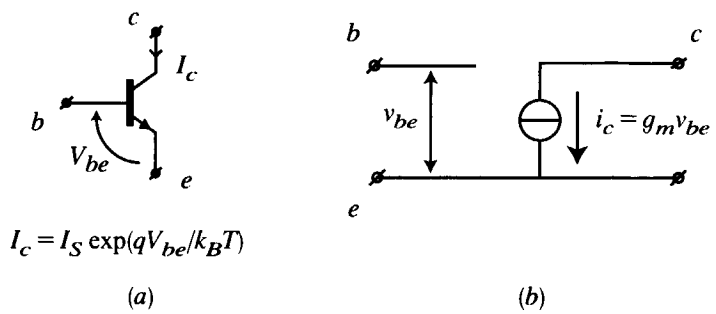

Figure 1.1: Network symbol (a) and ideal equivalent small-signal circuit (b) of a bipolar (npn) transistor.

seen when Equation 1.1 is inserted into the definition of the transconductance, $g_m = \partial I_c/\partial V_{be}$. The relationship is often resumed in the network models of Figure 1.1a and b. They are often used as a first approach in the design of analog circuits.

A real transistor has characteristics that are, of course, less perfect. These imperfections include the base current, the Early effect, the base resistance, and some parasitic capacitances. They are usually modelled by extending Figure 1.1b to a network like that given in Figure 1.2. Nevertheless, the imperfections rarely determine the main functioning of the circuit because in a good circuit design they are not allowed to become dominant.

1.2 Shifts in characteristics due to stress

Since the beginning of transistor physics it has been known that the current-voltage relationship not only depends on temperature, but also on mechanical

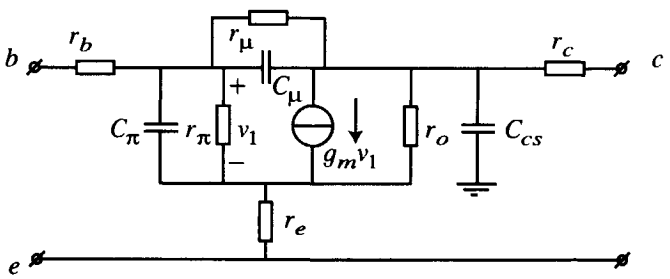

Figure 1.2: Small-signal equivalent circuit for bipolar transistor including imperfections [1].

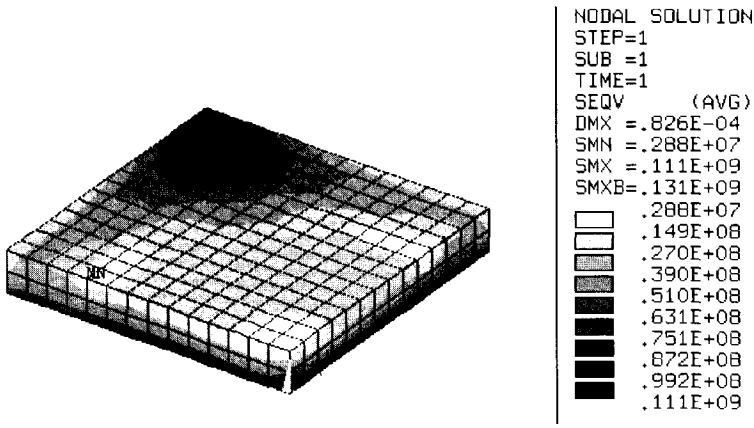

Figure 1.3: Equivalent stress distribution in a silicon chip mounted in an open quad flat package (OQFP). Units: Pa. Finite-element analysis by C. Cotofana Tomescu [7].

stress [2]. Stress acts on the relationship in Equation 1.1 through the saturation current I_S. This stress dependence is known as the piezojunction effect. It is related to the better-known piezoresistance effect [3] and the piezotunnelling effect [4, 5]. It is nonlinear and very anisotropic, meaning that it depends to a large extent on the orientation of the stress and current orientation with respect to the semiconductor crystal. For a stress of 100 MPa it typically changes I_S with five percent, corresponding to a change in V_{bc} of 1.2 mV.

Apart from the saturation current, stress also changes the values of the other transistor characteristics. In literature, for instance, base currents were reported that decrease up to a factor five for very high compressive stresses [6]. Changes in other characteristics have hardly been investigated.

1.3 Consequences for circuits

Stress is often present in integrated circuits as an unwanted by-product of packaging and processing. During the packaging the chip containing the circuit is glued on a frame and encapsulated in moulded plastic. When the materials harden and cool down they shrink. This can introduce stresses larger than 100 MPa; see Figure 1.3 and Reference [7]-[10]. During the processing, even higher stress levels may develop when silicon oxide and nitride layers are formed on top of the chip. It is especially around LOCOS that stress over 1 GPa can occur [11, 12]. Unfortunately, the stresses are not constant but may change during the lifetime of the circuit. These changes are sometimes accompanied by mechanical damage, but more often by a shift in the electrical characteristics of the circuit.

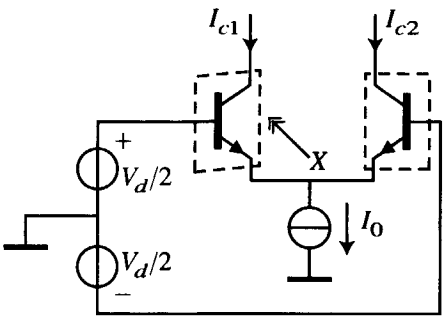

Figure 1.4: Emitter-coupled pair where the symmetry between the transistors is disturbed by a mechanical stress X.

The packaging and processing stresses can pose serious problems in analog circuits where the current-voltage relationship of Equation 1.1 needs to be known accurately. Good examples of this are found in bandgap temperature sensors, bandgap voltage references, and DA-converters [13, 14]. The stresses can also pose problems if they shift the transistor imperfections such as the base current to completely different values.

The problems can be illustrated by means of a popular subcircuit in analog electronics, the emitter-coupled pair. This circuit is represented in Figure 1.4. It should transfer the differential voltage V_d to a differential current I_d, which is the difference between the collector currents of both transistors, $I_{c1} - I_{c2}$. If V_d is zero, I_d should be zero as well. Normally, this value is obtained by forcing the base-emitter voltages to reach the same value and by matching the geometry and temperature as closely as possible. However, if one of the transistors is stressed, its I_S changes and therefore its collector current. This creates an asymmetry between the transistors, resulting in some offset value for I_d. As a result, the transfer function $I_d(V_d)$ changes. It should be noted, however, that the changes are limited here to even powers in V_d, due to the common bias current I_0.

1.4 Consequences for sensors

Stress in a transistor is not always undesirable, and it can also be introduced deliberately to measure some mechanical quantity. This idea has been utilised in a number of sensors [6], [15]-[18]. A good example of such a sensor is the 'classical' micromachined accelerometer of Figure 1.5. When the frame of this device is accelerated, the suspended mass in the middle tends to stay behind because of its inertia. This causes mechanical stress in the hinges. If the hinges contain transistors with a constant collector current, the stress is translated into a change in the base-emitter voltage or the collector current [16].

Most accelerometers today contain resistors instead of transistors in the

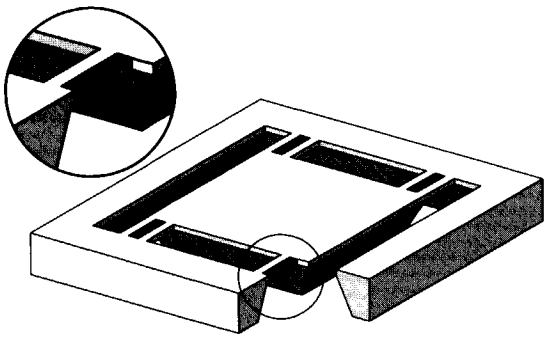

Figure 1.5: Classical micromachined accelerometer. The mass is suspended on hinges where transistors can be integrated instead of piezoresistors [16].

hinges, or have a capacitive read-out. The transistors, however, have a number of potential advantages. Just as the classical piezoresistors, they are fully compatible with the standard processing of electronics. But unlike them they have:

- A much larger internal resistance, and
- Smaller dimensions of the active area.

The larger internal resistance means that the sensor is less sensitive to the offset voltage of the amplifier, which is always needed to read out the signal. It also means that less power is needed to transfer the signal to the amplifier with equal accuracy. The smaller dimensions enable stress measurements to be more localised and therefore more precise. In addition, they save expensive chip area.

1.5 Previous research on stress effects

The piezojunction effect was first reported by Hall, Bardeen and Pearson in 1951 in relation to hydrostatic pressure on *pn*-junctions [2]. In 1962 Rindner found that the effect is much larger for anisotropic stress [19]. That stress was caused by pressing a stylus on a point of the *pn*-junction. The experiments with styli were frequently repeated and extended during the 1960s [20]-[38]. They were characterized by very high, compressive stresses with a large gradient [30]. The resulting changes were often spectacular, not only in the current-voltage relationship, but also where the base current, the breakdown voltage, and the generation-recombination currents were concerned [39]-[46]. Much smaller stresses were obtained in 1973 by Monteith and Wortman, who used cantilever beams instead of styli [47]. They observed that the changes under tensile stress differ from those under compressive stress.

The piezojunction effect was first explained in 1951 but only in relation to hydrostatic pressure [2]. The theory was based on bandgap widening caused by this isotropic stress. In 1964 the theory was extended by Wortman, Hauser, and Burger to anisotropic, high stresses in the GPa range [48]. This theory is based on the stress-induced lifting of degeneration of the band edges. It predicts a stress response which is independent of the current direction, and is equal for diodes, *npn,* and *pnp* transistors. The theory was further refined in 1967 by Kanda, who considered changes in the effective masses [49] for stresses over 1 GPa. Afterwards, the theory was modified by the addition of some less critical features, but those were rarely verified in practice [50]- [52], [18]. To the author's knowledge, it has never been extended to include low anisotropic stresses by considering *all* changes in the effective masses.

The piezojunction effect has often been applied in the design of mechanical sensors [53]-[60]. Initially, the sensors were mainly based on the pressing styli. This made them commercially not very successful because they were difficult to manufacture, sensitive to thermal expansion, and vulnerable to overload damage [6]. The invention of micromachining, however, has enabled new designs such as the accelerometer in Figure 1.5. So far, however, the new possibilities have only seldomly been explored [16]-[18].

1.6 Transistor modelling for design purposes

The description of the piezojunction effect is a part of transistor modelling, and it should therefore satisfy the same requirements. Transistor models are mainly used for the design of circuits and sensors. The purpose of such a design is to implement a signal processing function within a number of quality conditions [61]. An accelerometer, for instance, may be designed for a sensitivity of 1 V per g acceleration with a maximum offset of 0.1 mg and an operating temperature between -40 and +80 °C.

Electronic designing can be done at three levels [61]. On the highest level, it is a process of circuit (or sensor) *synthesis.* This consists of finding the optimum configuration between all device connections and dimensions with the same signal processing function. On an intermediate level, designing is a process of circuit *optimisation.* This concerns the study of one basic configuration with a number of variations. On the lowest level, designing involves circuit *analysis,* which is the study of one possible configuration.

Circuit design requires adequate models of the constituting components. The models should at least be able to reproduce the behaviour of the component with sufficient accuracy over the relevant working domain. This is not enough, however, for performing circuit *synthesis.* Synthesis involves comparing all possible configurations and therefore requires figures of their quality. These figures should have a physical meaning. In addition, they should describe the configuration at a high level of abstraction. They can only be derived from models which are mathematically simple, contain unique, physical parameters, and have a well-defined domain of validity.

The requirements for circuit synthesis are only satisfied by analytical models [61]. Analytical models are characterised by a set of closed, analytical equations describing the state of the component terminals as a function of the applied signals. They are based on *a priori* knowledge of the structure of the component and on the physical processes inside. A typical example is the Gummel-Poon model of a bipolar transistor, which has been implemented in SPICE circuit simulators. The analytical models of the piezojunction effect, however, have not been very practical so far.

1.7 Scope of this thesis

This thesis presents an analytical model of the piezojunction effect suitable for circuit and sensor design on all levels. Previous models of the piezojunction effect were quite successful when it came to predicting the experiments conducted with hydrostatic pressure and stress-inducing styli. However, they were not designed for the stresses which are currently important: stresses caused by packaging and processing, and stresses in micromachined mechanical sensors. Those stresses have the following characteristics:

- They can be tensile as well as compressive;
- They are often lower than 200 MPa in magnitude;
- They may have any orientation with respect to the axes of the crystal from which the transistor is fabricated;
- They are generally homogeneous on the scale of a transistor.

The previous models can only be valid in this stress domain if they are extended. This extension is the first concern of this thesis. The disadvantage of this approach is that it seriously increases the mathematical complexity of the model. The second aim is therefore to create a new model on a higher level of abstraction. The parameters of this model are still based on the physics of the old model. Even so, they are closely linked to measurable transistor characteristics. The structure of this model is very similar to that of the piezoresistive effect, which is well-known and widely used [3]-[63].

The approach of the modelling in this thesis is both theoretical and experimental. The main transistor equation is re-derived from the energy bands of the material and the structure of the device. All assumptions in this derivation are critically assessed on their validity under mechanical stress. Experimentally, the model is verified by fabricating *npn* and *pnp* transistors and characterising them under different bending stresses.

The work presented is restricted to silicon transistors as they are most commonly applied in analog circuits and sensors. It therefore focuses on their saturation current at normal, forward bias, and at low-level injection. It neglects the contribution of the generation and recombination currents. In addition, it is

restricted to extrinsic silicon at room temperature, where doping atoms are completely ionised.

1.8 Organisation of this thesis

The thesis starts with a theoretical chapter, followed by a chapter on the experiments. It continues with a discussion of the results, and with some general conclusions. It ends with four appendices supporting the theory. The derivations in these appendices can in principle be found in literature. However, they are quite specialistic and appear in a wide variety of notations and in many different contexts. Their inclusion in this thesis should therefore enhance the readability.

The theory chapter discusses the influence of mechanical stress on different levels of transistor modelling. Each level is described in a separate section, as shown in Figure 1.6. It arises from the foregoing level through the physics described in the appendices.

On the first level stress creates a mechanical strain. The relation between stress and strain is contained in the elasticity theory. On the second level, strain causes deformations in the electronic band structure of the crystal. This effect is most important at the extrema of the conduction and valence bands, on which the attention will be focused. On the third level, the band deformations change the transport properties of charge carriers, especially the conductivity. These changes are described with a revised microscopic model, and also with an entirely new macroscopic model. On the last level, the changing transport properties cause shifts in the transistor characteristics, in particular in the saturation current.

The measurements chapter discusses the experiments that were done to verify the new theory and to extract its parameters. It is divided in sections about the principle, the measurement set-up, its operation, and the results. The principles section describes the estimated stress response of the investigated devices, and the configuration of the measurement equipment. The set-up section discusses the actual design of the devices and their mounting on a beam carrier. It also shows the construction of the bending apparatus with which they are stressed. In addition, it discusses the properties of the source-measurement units. The results section presents graphs of the stress-induced current changes, for different crystal orientations. It also presents the extracted model parameters.

The discussion and conclusions chapter compares the measurements with the theory developed. It indicates how the results could be used by designers of circuits and sensors. It ends with some suggestions for future research.

The appendices, finally, contain derivations supporting the theory chapter. Appendix A resumes the definitions of stress, strain, and elasticity. In addition, it

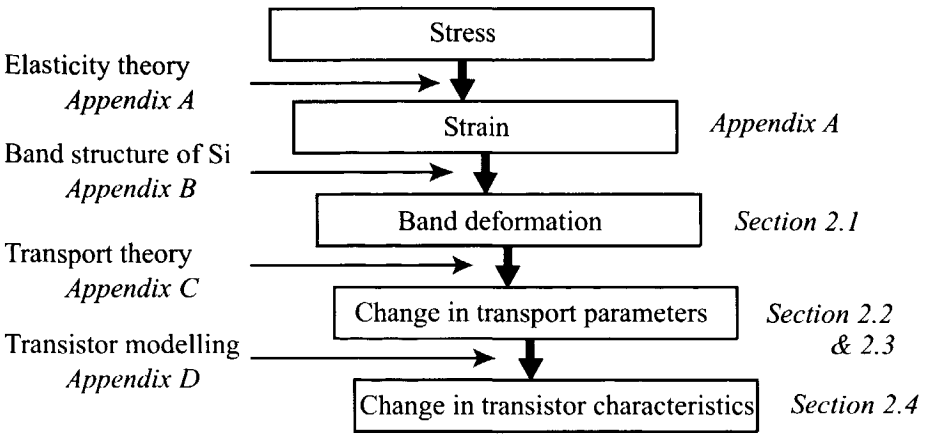

Figure 1.6: Structure of the theory chapter. The influence of stress on a transistor is modelled on different levels. Each level is treated in a separate section and supported by an appendix.

presents some mathematics on tensor properties. Appendix B derives the band equations of silicon from the Schrödinger equation with the aid of the $\mathbf{k}\hat{\mathbf{p}}$-theory. The band equations are used in Appendix C to derive the charge transport in silicon which is inhomogeneous and anisotropic. The parameters of this general case are used in Appendix D to model the saturation current of the bipolar transistor.

Bibliography

[1] P. R. Gray and R. G. Meyer, *Analysis and design of analog integrated circuits*. 3rd ed., Wiley, New York, 1993.

[2] H. H. Hall, J. Bardeen, and G. L. Pearson, The effects of pressure and temperature on the resistance of *p-n* junctions in germanium, *Phys. Rev.*, *84* (1951) 129-132.

[3] Y. Kanda, Piezoresistance effect of silicon, *Sensors and Actuators A, 28* (1991) 83-91.

[4] A. P. Friedrich, P. A. Besse, C. M. A. Ashruf, and R. S. Popovic, Characterization of a novel piezo-tunneling strain sensor, *Sensors and actuators A, 66* (1998) 125-130.

[5] A. P. Friedrich, *Silicon piezo-tunneling strain sensor, Series in Microsystems, Vol.2,* Hartung-Gorre, Konstanz, 1999.

[6] J. Matovic, Z. Djuric, N. Simicic, and A. Vijanic, Piezojunction effect based pressure sensor, *Electron. Lett., 29* (1993) 565-566.

[7] C. V. B. Cotofana Tomescu, *Low-cost sensor packaging*, Ph.D. Thesis of Delft University of Technology, Delft University Press, Delft, The Netherlands, 2001.

[8] S. A. Gee, W. F. van den Boogert, and V. R. Akylas, Strain-gauge mapping of die surace stresses, *IEEE Trans. Comp., Hybrids, Manuf. Techn., 12* (1989) 587-593.

[9] Y. Zou, J. C. Suhling, and R. C. Jaeger, Characterization of plastic packages using (100) silicon stress test chips, *Applicat. Experimental Mechanics Electron. Packag., 22* (1997) 15-21.

[10] B. Nysaether, A. Larsen, B. Liverod, and P. Ohlckers, Measurement of package-induced stress and thermal zero shift in transfer molded silicon piczoresistive pressure sensors, *J. Micromech. Microeng., 8* (1998) 168-171.

[11] P. Smeys, P. B. Griffin, Z. U. Rek, E. de Wolf, and K. C. Saraswat, Influence of process-induced stress on device characteristics and its impact on scaled device performance, *IEEE Trans. Electron Devices, 46* (1999) 1245-1252.

[12] H. Miura and Y. Tanizaki, Effect of process-induced mechanical stress on circuit layout, *Proc. Simul. Semicond. Dev. Processes 95, Vol. 6, 1995, pp. 147-150.*

[13] G. C. M. Meijer, *Integrated circuits and components for bandgap references and temperature transducers,* Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 1982, p.18.

[14] F. Fruett and G. C. M. Meijer, A test structure to characterize the piezojunction effect and its influence on silicon temperature sensors, *Proc. 13th Eur. Conf. Solid-State Transducers (Eurosensors XIII),* The Hague, The Netherlands, Sept. 12-15, 1999, pp. 53-56.

[15] Y. Kanda, Y. Kanazawa, T. Terada, and M. Maki, Improved Si piezotransistors for mechano-electrical transducers, *IEEE Trans. Electron Devices., ED-25* (1978) 813-817.

[16] B. Puers, L. Reynaert, W. Snoeys, W. M. C. Sansen, A new uniaxial accelerometer in silicon based on the piezojunction effect, *IEEE Trans. El. Dev., ED-35* (1988) 764-770.

[17] B. Puers and W. M. C. Sansen, New mechanical sensors in silicon by micromachining piezojunction transistors, *Dig. 3rd International Conference on Solid-State Sensors and Actuators (Transducers'87),* Tokyo, 1987, pp. 324-327.

[18] R. Schellin and R. Mohr, A monolithically-integrated transistor microphone: modelling and theoretical behaviour, *Sensors and Actuators A, 37-38* (1993) 666-673.

[19] W. Rindner, Resistance of elastically deformed shallow p-n junctions, *J. Appl. Phys., 33* (1962) 2479-80.

[20] W. Rindner and I. Braun, Resistance of elastically deformed shallow p-n junctions, II., *J. Appl. Phys., 34* (1963) 1958-1970.

[21] Y. Matukura, Anisotropic stress effect of silicon pn junctions, *Japan. J. Appl. Phys., 3* (1964) 256-261.

[22] Y. Matukura, Some factor influencing on the anisotropic stress effect of pn junctions, *Japan. J. Appl. Phys., 3* (1964) 516-520.

[23] K. Matsuo, Electric pulse response of P-N junctions under anisotropic stresses, *J. Phys. Soc. Japan, 19* (1964) 1490-1491.

[24] R. Edwards, Some effects of localized stress on silicon planar transistors, *IEEE Trans. Electron Devices, 11* (1964) 286-294.

[25] W. Touchy, Untersuchungen ueber Druckempfindlichkeit von Si-Transistorsystemen, *Z. Angew. Phys., 16* (1964) 430-434.

[26] T. Imai, M. Uchida, H. Sato and A. Kobayashi, Effect of uniaxial stress on germanium p-n junctions, *Japan. J. Appl. Phys., 4* (1965) 102-113.

[27] T. Imai and M. Uchida, Effect of uniaxial stress on germanium p-n junctions (II), *Japan. J. Appl. Phys., 4* (1965) 409-414.

[28] Y. Matukura, Uniaxial stress effect of Ge p-n junctions, *Japan. J. Appl. Phys., 4* (1965) 309.

[29] W. Rindner, G. Doering, and R. Wonson, Structural and operational characteristics of piezo-transistors and allied devices, *Solid-State Electron., 8* (1965) 227-240.

[30] K. Bulthuis, The effects of local pressure on silicon p-n junctions, *Philips Res. Repts., 20* (1965) 415-431.

[31] K. Bulthuis, The effect of local pressure on germanium p-n and p-s-n structures, *Philips Res. Repts., 21* (1966) 85-103.

[32] K. Bulthuis, Effect of local pressure on germanium p-n junctions, *J. Appl. Phys., 37* (1966) 2066-2068.

[33] A. L. Polyakova and V. V. Shklovskaya-Kordi, Influence of deformation on the properties of silicon p-n junctions, *Sov. Phys.-Solid State, 6* (1966) 163-168.

[34] R. H. Mattson, L. D. Yau, and J. R. DuBois, Incremental stress effects in transistors, *Solid-St. Electron., 10* (1967) 241-251.

[35] K. Bulthuis, Effect of uniaxial pressure along the main crystallographic axes for silicon and germanium, *Philips Res. Repts., 23* (1968) 25-47.

[36] A. A. Mahmoud, C. Calabrese, and J. R. Tudor, Stress effects on n-p-n transistor parameters, *Proc. IEEE, 59* (1971) 1264-1265.

[37] Z. Djuric, The effect of localised deformation upon the common emitter transistor current gain, *Solid-State Electron., 14* (1971) 627-637.

[38] G. Kaszynski and W. Ortmeyer, Über die Wechseldruckempfindlichkeit von Piezodioden und Piezotransistoren, *Fernmeldetechnik, 13* (1973) 205-210.

[39] A. Goetzberger and R. H. Finch, Lowering the breakdown voltage of silicon p-n junctions by stress, *J. Appl. Phys., 35* (1964) 1851-1854.

[40] W. Rindner, Breakdown voltages and currents in mechanically stressed Ge and Si diodes, *Appl. Phys. Lett., 6* (1965) 225-226.

[41] J. R. Hauser, An approximation for generation-recombination current in P-N junctions, *Proc. IEEE, 53* (1965) 743-744.

[42] Y. Matukura, Uniaxial stress effect on Ge grown junctions, *Japan. J. Appl. Phys., 4* (1965) 632-638.

[43] J. J. Wortman and J. R. Hauser, Effect of mechanical stress on p-n junction device characteristics. II. Generation-recombination current, , *J. Appl. Phys., 37* (1966) 3527-3530.

[44] J. R. Hauser and J.J. Wortman, Some effects of mechanical stress on the breakdown voltage of p-n junctions, *J. Appl. Phys., 37* (1966) 3884-3892.

[45] R. Birebent et J.-M. Simon, Interprétation des effets de la pression sur la tension de claquage des diodes à effet Zener, *C. R. Acad. Sc. Paris, Série B, 263* (1966) 1252-1253.

[46] H. Kressel and A. Elsea, Effect of generation-recombination centers on the stress-dependence of Si p-n junction characteristics, *Solid-St. Electron., 10* (1967) 213-224.

[47] L. K. Monteith and J. J. Wortman, Characterization of p-n junctions under the influence of a time varying mechanical strain, *Solid-St. Electron., 16* (1973) 229-237.

[48] J. J. Wortman, J. R. Hauser, and R. M. Burger, Effect of mechanical stress on p-n junction device characteristics, *J. Appl. Phys., 35* (1964) 2122-2131.

[49] Y. Kanda, Effect of stress on germanium and silicon p-n junctions, *Japan. J. Appl. Phys., 6* (1967) 475-486.

[50] Y. Kanda, Effect of stress on germanium and silicon p-n junctions II. Non-exhausted condition, *Japan. J. Appl. Phys., 7* (1968) 1464-1472.

[51] Y. Kanda, Effect of compressive stress on silicon bipolar devices, *J. Appl. Phys., 44* (1973) 389-393.

[52] W. Wlodarski and B. Moeschke, The effect of hydrostatic pressure on the characteristics of the forward biased p-n junctions, *Electron Technol., 13 (3)* (1980) 3-42.

[53] M. E. Sikorski, Transistor Microphones, *J. Audio Eng. Soc., 13* (1965) 207-217.

[54] W. H. Legat and L. K. Russell, A silicon p-n junction transducer, *Solid-St. Electron., 8* (1965) 709-714.

[55] F. Krieger and H.-N. Toussaint, A piezo-mesh-diode pressure transducer, *Proc. IEEE, 55* (1967) 1234-1235.

[56] O. Brünnert, R. Karmann, und P. Lohse, Über ein Mikrophon mit druckgesteuertem pn-Übergang, *NTZ, 20* (1967) 76-79.

[57] J. J. Wortman and L. K. Monteith , Semiconductor mechanical sensors, *IEEE Trans. Electron Devices., ED-16* (1969) 855-860.

[58] W. Wlodarski and J. Pintara, Application of the semiconductor p-n junction to measurements of rapidly varying pressures, *Bull. Acad. Pol. Sci. Ser. Sci. Tech., 21* (1973) 393-401.

[59] D. P. Jones, S. V. Ellam, H. Riddle, and B. W. Watson, The measurement of air flow in a forced expiration using a pressure-sensitive transistor, *Med. &Biol. Eng., 13* (1975) 71-77.

[60] W. Sansen, P. Vandeloo and B. Puers, A force transducer based on stress effects in bipolar transistors, *Sensors and Actuators, 3* (1982) 343-354.

[61] M. G. Middelhoek, *The identification of analytical device models,* Ph. D. Thesis of the Delft University of Technology, Delft University Press, Delft, The Netherlands, 1992.

[62] A. Belu-Marian, E. Candet and A. Devenyi, *Piezoresistive sensors, in: Thin film resistive sensors,* C. Ciureanu and S. Middelhoek ed., IOP Publishing, Bristol, 1992.

[63] S. Middelhoek and S. A. Audet, *Silicon Sensors, Ch. 3,* Academic Press, London, 1989.

Chapter 2

Theory of Stress-Induced Changes

The effect of mechanical stress on the electrical characteristics of semiconductor devices comprises several transduction steps. These consecutive steps are outlined in Section 1.8 and Figure 1.6. First, the stress generates a mechanical strain, which changes the energy bands of the material. Then, the bands modify the parameters of charge transport. Finally, the transport parameters shift the electrical terminal characteristics of resistors, diodes, and transistors.

In this chapter the successive transduction steps are modelled in detail. Where possible, these models are expressed as explicit analytical equations. Compared to data fitting models, for example, analytical models provide much more insight in the physical processes [1]. They are also more compact and can be evaluated rapidly. Finally, they have interface parameters with a physical meaning. As a result, the model equations can be used in the next transduction step: a model with a higher level of abstraction. This leads to parameters which can directly be verified by measuring voltages and currents.

2.1 Deformation of the energy bands

In solid-state physics it is fundamental that the conduction electrons of a crystal behave as quantum-mechanical waves subject to the periodic boundary condition [2, 3]. This means that the electron wave functions adapt the symmetry of the lattice. It also means that the wave numbers of the electrons are compatible with the interatomic distances of the crystal lattice, as sketched in Figure 2.1.a.

When the crystal is mechanically stressed, the interatomic distances are strained. This transduction is mathematically described in Appendix A. The resulting strain changes the periods of the electron waves, as shown in Figure 2.1.b. In most cases, it also reduces the symmetry of the wave functions to less symmetric states.

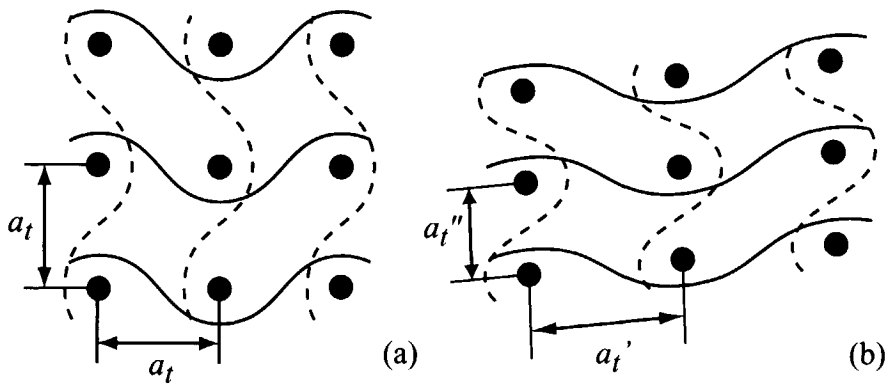

Figure 2.1: Periodic boundary condition in a crystal with a square symmetry without strain (a) and in the presence of strain (b). The electron waves have the periodicity of the crystal as well as its symmetry. The strain therefore changes the interatomic distance a_t and the wave numbers of the electrons.

The wave description of electrons is far too complex if large numbers of electrons are involved. In this case the description is often replaced by energy bands, in which the electrons can be considered as (semi-)classical particles. These bands represent the energy of independent electrons in a material as a function of the wave vector. They are also called band structure, band diagrams, or dispersion relations. Their exact calculation has occupied specialists for over fifty years (see e.g. Reference [4] and [5]).

However, for the modelling of most semiconductor properties it is sufficient to know only the small parts of the band structure where moveable charge is located. These parts are located around the minima of the conduction bands and the maxima of the valence bands, which are also called the band edges. The edges are separated by a bandgap E_G, as shown in Figure 2.2.a. The edges are also characterised by the curvature or effective mass. Their mathematical description is found in Appendix B.

In this section it is derived how the band edges are changed by mechanical stress. It will be shown that the stress shifts the band edges and deforms their curvature. In addition, it will be shown that stress can break the symmetry of the edges. In this case, the edges are no longer at the same level, which means that the bandgap is no longer uniquely defined. This is shown in Figure 2.2.b. It has a strong influence on the electrical properties of the material. For that reason, the classical definition of the bandgap will be reconsidered, just as the band edge diagram as a function of position.

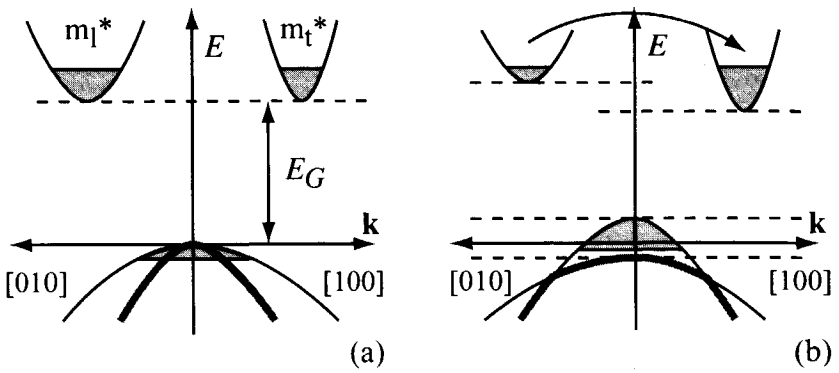

Figure 2.2: (a) Schematic band structure of silicon in the stress-free and (b) in the stressed case. The conduction band edges on top are separated by an energy gap E_G from the valence band edges on the bottom. They are partially filled with electrons and holes, respectively. The application of stress shifts and deforms the edges. This lifts their degeneracy, resulting in different gaps.

2.1.1 The kp̂ method for deformed cubic crystals

The shape of the energy bands around the edges can be mathematically determined by the **kp̂** method. This method is explained in Appendix B for the unstressed case and is worked out for silicon and germanium. It is based on perturbation theory and the symmetry properties of the semiconductor crystal. The theory requires a set of unperturbed orthogonal states, for which the electron wave functions $\psi_{n\mathbf{k}_0}(\mathbf{x})$ of the band maximum or minimum are chosen. These functions are supposed to be perturbed by a wave vector slightly off the extremum, by the magnetic coupling between the spin and orbital angular momenta, and by the strain.

The **kp̂** method was extended to deformed crystals by the formulation of the deformation potential theory. For the conduction bands this theory was proposed by Bardeen and Shockley [9] and further developed by Herring and Vogt [10, 51]. For the more complicated case of the valence bands it was initiated by Adams [11]. It was completed simultanuously by Pikus and Bir [6], and by Kleiner and Roth [7]. The formulation of Pikus and Bir is used in the following two sections.

The deformation potential theory consists of three major steps. Firstly, the strain perturbation Hamiltonian is defined. Secondly, the symmetry of the Hamiltonian matrix elements is evaluated and, finally, the secular determinant is solved. This results in a set of n energy eigenvalues E_n, depending on the wave number $\mathbf{k}$ and the strain e_{ij}. These eigenvalues form the band equations around the band edges. They contain matrix elements which are material parameters and can be found in literature.

2.1.2 The strain Hamiltonian

The influence of strain on a crystal is to change its periodicity, and thereby the periodicity of the electron eigenfunctions. As a result, the functions ψ'_n of the strained crystal have another period than those of the unstrained crystal, ψ_n. It is therefore impossible to express the ψ'_n as a linear combination of the ψ_n, although this is required by perturbation theory.

The periodicity can be restored by transforming the coordinate system of the strained crystal into that of the unstrained crystal. Any position vector x'_i is then expressed in terms of the strain tensor e_{ij} and the unstrained vector x_j (see Appendix A):

$$x'_i = (\delta_{ij} + e_{ij})\, x_j \tag{2.1}$$

where δ_{ij} is the Kronecker delta and where the Einstein convention has been used to indicate the summation of the terms with dummy subscripts. In the following, vectors will be indicated by both boldface symbols and subscript notation. Boldface symbols are more compact, whereas subscripts provide more insight when vectors are combined with tensors. The transformation from strained to unstrained coordinates has consequences for the electron functions, the full Hamiltonian, and the Hamiltonian on the basis of modulating functions used in the $\mathbf{k\hat{p}}$ method.

Transformation of the electron functions

The electron functions can always be written as a Bloch function $\psi'_{n\mathbf{k}'}(\mathbf{x}')$, where the accents indicate the coordinates of the strained crystal lattice. The Bloch function is by definition the product of a plane wave and a modulating function with the periodicity of the lattice $u'_{n\mathbf{k}'}$:

$$\psi'_{n\mathbf{k}'}(\mathbf{x}') = e^{i\mathbf{k}'\cdot\mathbf{x}'} u'_{n\mathbf{k}'}(\mathbf{x}') \tag{2.2}$$

The position vector $\mathbf{x}' = x'_i$ in this equation can be transformed to the unstrained system with the aid of Equation 2.1. This yields:

$$\begin{aligned} \psi'_{n\mathbf{k}'}(\mathbf{x}') &= e^{ik'_i(\delta_{ij}+e_{ij})x_j} u'_{nk'_i}((\delta_{ij} + e_{ij})\, x_i) \\ &= e^{ik_j x_j} u'_{nk_j}(x_j, e_{ij}) \\ &= \psi'_{n\mathbf{k}}(\mathbf{x}) \end{aligned} \tag{2.3}$$

which is true if the wave vector $\mathbf{k}$ in the unstrained system is equal to:

$$k_j = (\delta_{ij} + e_{ij})\, k'_i \tag{2.4}$$

and if the wave vector in the index of the modulating function follows that of the plane wave.

Transformation of the full Hamiltonian

The electron functions $\psi'_{n\mathbf{k}'}(\mathbf{x}')$ are the eigenfuctions of the full Hamiltonian in the strained crystal $\widehat{H}'(\mathbf{x}')$. This Hamiltonian can also be transformed to the axes system of the unstrained crystal. For this purpose it is written as:

$$\widehat{H}'(\mathbf{x}') = \frac{\widehat{p}'^2(\mathbf{x}')}{2m_0} + U'(\mathbf{x}') \tag{2.5}$$

in which $\widehat{p}'^2$ is the square of the strained momentum operator $\widehat{p}'_i$, m_0 is the free electron mass, and U' is the potential of the electron in the strained crystal. The momentum and its square are transformed by using their defining functions, which contain the position vector x'_i:

$$\widehat{p}'_i = -i\hbar\frac{\partial}{\partial x'_i} = -i\hbar\frac{\partial x_j}{\partial x'_i}\frac{\partial}{\partial x_j} = (\delta_{ij} - e_{ij})\,\widehat{p}_j \tag{2.6}$$

$$\widehat{p}'^2 = \widehat{p}^2 - 2\widehat{p}_i\widehat{p}_j e_{ij} \tag{2.7}$$

The second-order terms in strain have been neglected here. The potential energy $U'(\mathbf{x}')$, in addition, can be transformed by using a first-order Taylor expansion:

$$U'(x'_i) = U'((\delta_{ij} + e_{ij})\,x_j) = U(x_j) + U_{ij}(x_j)\,e_{ij} \tag{2.8}$$

in which the tensor $U_{ij}(x_j)$ is defined as the first-order derivative of the potential with respect to strain, at zero strain:

$$U_{ij}(x_j) = \left.\frac{\partial U(x_j)}{\partial e_{ij}}\right|_{e_{ij}=0} \tag{2.9}$$

The full Hamiltonian $\widehat{H}'$ can therefore be expressed in unstrained coordinates as a combinaton of the regular Hamiltonian $\widehat{H}$ and a strain perturbation Hamiltonian $\widehat{H}_e$:

$$\widehat{H}'(\mathbf{x}')\,\psi'_{n\mathbf{k}'} = \left(\widehat{H} + \widehat{H}_e\right)(\mathbf{x})\,\psi'_{n\mathbf{k}} \tag{2.10}$$

in which:

$$\widehat{H} = \frac{\widehat{p}^2}{2m_0} + U \tag{2.11}$$

$$\widehat{H}_e = -\frac{1}{m_0}\widehat{p}_i\widehat{p}_j e_{ij} + U_{ij}e_{ij} \tag{2.12}$$

It should be noticed that the influence of spin-orbit coupling has not been taken into account here. In silicon, however, the coupling is small enough to be considered as a separate perturbation. Under this assumption it is calculated in Appendix B. Moreover, it appears that its inclusion does not significantly change the shape of the following equations [6, 8]. It will therefore be added to the results in Section 2.1.5.

Transformation of the modulating function Hamiltonian

The transformation from strained to unstrained coordinates is finally applied to $\widehat{H}'_u(\mathbf{x}')$, which is the Hamiltonian defined on a basis of the modulating parts $u'_{n\mathbf{k}_0}(\mathbf{x})$ of Bloch functions. In the unstrained case the transition to the basis of modulating functions yields an extra $\mathbf{k}\widehat{\mathbf{p}}$-Hamiltonian $\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}}$, as shown in in Appendix B. For the strained case the transition yields some extra Hamiltonians, which can be shown by similar steps. The full Hamiltonian of Equation 2.10 is therefore operated on the Bloch function of Equation 2.2, after which the plane wave part is removed. Then, the wave vector $\mathbf{k}_0$ of the band extremum is introduced. This yields:

$$\widehat{H}'_u(\mathbf{x}')\,u'_{n\mathbf{k}_0} = \left(\widehat{H}_0 + \widehat{H}_{0e} + \widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}} + \widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}e}\right)(\mathbf{x})\,u'_{n\mathbf{k}_0} \tag{2.13}$$

in which $\widehat{H}_0$ is the Hamiltonian of the unperturbed electrons on the basis of $u'_{n\mathbf{k}_0}(\mathbf{x})$, and $\widehat{H}_{0e}$, $\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}}$, and $\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}e}$ are perturbations on the same basis. More specificly, $\widehat{H}_{0e}$ is the strain Hamiltonian, $\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}}$ is the $\mathbf{k}\widehat{\mathbf{p}}$-Hamiltonian, and $\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}e}$ is the influence of strain on the $\mathbf{k}\widehat{\mathbf{p}}$-Hamiltonian. They are given, respectively, by:

$$\widehat{H}_0 = \frac{\hat{p}^2}{2m_0} + U + \frac{\hbar}{m_0}k_i^0\hat{p}_i + \frac{\hbar^2 k_0^2}{2m_0} \tag{2.14}$$

$$\widehat{H}_{0e} = -\frac{1}{m_0}\widehat{p}_i e_{ij}\widehat{p}_j + U_{ij}e_{ij} - \frac{\hbar}{m_0}k_i^0 e_{ij}\hat{p}_j \tag{2.15}$$

$$\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}} = \frac{\hbar}{m_0}\left(k_i - k_i^0\right)\hat{p}_i \tag{2.16}$$

$$\widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}e} = -\frac{\hbar}{m_0}\left(k_i - k_i^0\right)e_{ij}\hat{p}_j \tag{2.17}$$

Usually, however, it is allowed to neglect higher-order terms and only retain the terms which are quadratic in $\mathbf{k}$, linear in e_{ij}, and linear in both $\mathbf{k}$ and e_{ij}. To a good approximation, Equation 2.13 can therefore be written as:

$$\widehat{H}'_u(\mathbf{x}')\,u'_{n\mathbf{k}_0} = \left(\widehat{H}_0 + \widehat{H}_e + \widehat{H}_{\mathbf{k}\widehat{\mathbf{p}}}\right)(\mathbf{x})\,u'_{n\mathbf{k}_0} \tag{2.18}$$

in which $\widehat{H}_e$ is the strain Hamiltonian of Equation 2.12.

2.1.3 Strain matrix elements

The perturbation Hamiltonians of the preceding section can be used to evaluate the matrix elements of the secular determinant in the case of strain. In Appendix B, Section B.3, this determinant is found in its most general form. To the second order in $\mathbf{k}$ it is equal to:

$$\det\left(H_{mn} + \sum_{b\neq m,n}^{B}\frac{H_{mb}H_{bn}}{E'_u - H_{bb}} - E'_u\delta_{mn}\right) = 0 \tag{2.19}$$

where E'_u is the energy E_u reduced with the amount $(\hbar^2/2m_0)(k^2 - k_0^2)$. The matrix element H_{mn} is in the present case equal to the mean value of $\hat{H}'_u$, the Hamiltonian with respect to the functions $u_{m\mathbf{k}_0}$ and $u_{n\mathbf{k}_0}$, which are two modulating functions of electrons in the band extrema of the unstrained crystal. This means that H_{mn} is defined as:

$$H_{mn} = \int u^*_{m\mathbf{k}_0} \hat{H}'_u u_{n\mathbf{k}_0} d\mathbf{x} = \langle m| \hat{H}'_u |n\rangle \tag{2.20}$$

In Equation 2.19, the functions denoted by m and n correspond to the degenerate band edges for which a description is to be found. The functions b, however, correspond to the bands which lie relatively far away and are of no particular interest.

First-order elements

The first-order elements H_{mn} can be evaluated by using Equation 2.18 and writing:

$$H_{mn} = H^0_{mn} + H^{\mathbf{k}\hat{\mathbf{p}}}_{mn} + H^e_{mn} \tag{2.21}$$

The first term here is equal to the energy eigenvalue E_0 when $m = n$, because the $u_{m\mathbf{k}_0}$ are the orthogonal eigenfunctions of the Hamiltonian of the unperturbed electrons $\hat{H}_0$. The second term vanishes because its integrand is odd in $\mathbf{x}$. This is the result of the oddness of the momentum operator $\hat{p}_i = -i\hbar\partial/\partial x_i$, and the fact that the degenerate band edges have eigenfunctions m, n with the same parity, odd or even. The third term, however, should be retained. With the aid of Equation 2.12 this matrix element of $\hat{H}_e$ can be defined as:

$$H^e_{mn} = \langle m| - \frac{1}{m_0}\hat{p}_i\hat{p}_j + U_{ij} |n\rangle e_{ij} \tag{2.22}$$

$$\equiv \Xi^{mn}_{ij} e_{ij} \equiv \mathbf{H}_e \tag{2.23}$$

where Ξ^{mn}_{ij} is called the *first-order deformation potential tensor* [9, 10] and $\mathbf{H}_e$ is defined as the strain matrix, with the dimensions $m \times n$. An entire first-order element can therefore be written as:

$$H_{mn} = E_0\delta_{mn} + \Xi^{mn}_{ij} e_{ij} \tag{2.24}$$

Second-order elements

In contrast to the first-order elements, the product of second-order elements $H_{mb}H_{bn}$ is essentially the same as in the absence of strain. To derive this, H_{mb} is again expanded into its components with the aid of Equation 2.18. The first term of this expansion is H^0_{mb}. It vanishes because the functions b are unequal to m by definition, whereas they are all eigenfunctions of $\hat{H}_0$. The *product*

$H_{mb}H_{bn}$ therefore only consists of the remaining terms, the matrix elements of the $\mathbf{k}\widehat{\mathbf{p}}$- and strain Hamiltonians:

$$H_{mb}H_{bn} = H_{mb}^{\mathbf{k}\widehat{\mathbf{p}}}H_{bn}^{\mathbf{k}\widehat{\mathbf{p}}} + H_{mb}^{\mathbf{k}\widehat{\mathbf{p}}}H_{bn}^{e} + H_{mb}^{e}H_{bn}^{\mathbf{k}\widehat{\mathbf{p}}} + H_{mb}^{e}H_{bn}^{e} \tag{2.25}$$

The first term of this expression is equal to that of the $\mathbf{k}\widehat{\mathbf{p}}$ method in the absence of strain, and leads to the inverse effective mass tensors M_{ij} and D_{ij}^{mn} (see Appendix B, Section B.5 and B.6). The last term can be neglected because it is only quadratic in strain. The second and third term vanish if the crystal has a centre of inversion, such as silicon. This can be shown as follows.

Inversion symmetry implies that the matrix elements remain invariant when the position vector $\mathbf{x}$ is replaced by its inverse $-\mathbf{x}$ in the integrand. For the strain matrix element H_{mb}^{e} it should therefore be true that:

$$\begin{aligned} H_{mb}^{e} &= e_{ij}\int u_{m\mathbf{k}_0}^{*}(\mathbf{x})\left(-\frac{1}{m_0}\widehat{p}_i\widehat{p}_j + U_{ij}\right)(\mathbf{x})\,u_{b\mathbf{k}_0}(\mathbf{x})\,d\mathbf{x} \\ &= e_{ij}\int u_{m\mathbf{k}_0}^{*}(-\mathbf{x})\left(-\frac{1}{m_0}\widehat{p}_i\widehat{p}_j + U_{ij}\right)(-\mathbf{x})\,u_{b\mathbf{k}_0}(-\mathbf{x})\,d\mathbf{x} \end{aligned} \tag{2.26}$$

The momentum operator $\widehat{p}_i$ is odd in $\mathbf{x}$, whereas the strain e_{ij} and the strain derivative of the potential U_{ij} are even. The above equation is therefore only true if the functions m and b have the same parity, odd or even, or if the integral vanishes. For the $\mathbf{k}\widehat{\mathbf{p}}$ matrix element, on the other hand, the inversion symmetry implies that:

$$\begin{aligned} H_{bn}^{\mathbf{k}\widehat{\mathbf{p}}} &= \frac{\hbar}{m_0}k_i\int u_{b\mathbf{k}_0}^{*}(\mathbf{x})\,\widehat{p}_i(\mathbf{x})\,u_{n\mathbf{k}_0}(\mathbf{x})\,d\mathbf{x} \\ &= -\frac{\hbar}{m_0}k_i\int u_{b\mathbf{k}_0}^{*}(-\mathbf{x})\,\widehat{p}_i(-\mathbf{x})\,u_{n\mathbf{k}_0}(-\mathbf{x})\,d\mathbf{x} \end{aligned} \tag{2.27}$$

This is only true if the modulating functions b and n have opposite parity, or if the integral vanishes. In the *product* $H_{mb}^{e}H_{bn}^{\mathbf{k}\widehat{\mathbf{p}}}$ the functions m and n represent the same degenerate band edge and are of equal parity. If the function b also has that parity, the element $H_{bn}^{\mathbf{k}\widehat{\mathbf{p}}}$ will vanish. However, if it has the opposite parity, the element H_{mb}^{e} was seen to vanish. The product will therefore vanish in any case. As a result, the entire product of second-order elements $H_{mb}H_{bn}$ in Equation 2.25 reduces to:

$$H_{mb}H_{bn} = H_{mb}^{\mathbf{k}\widehat{\mathbf{p}}}H_{bn}^{\mathbf{k}\widehat{\mathbf{p}}} \tag{2.28}$$

which is the same expression as in the unstrained case.

2.1.4 Conduction band equations

The second-order description of the conduction band near the band edge can now be found by using that the edge is nondegenerate at $\mathbf{k}_0$. This can be seen from Figure 2.3.a and it means that the indices m and n can only be 1. It is

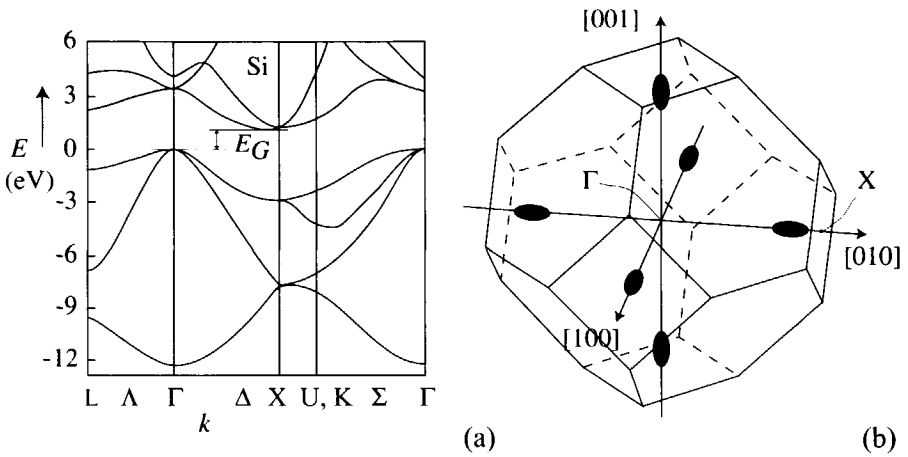

Figure 2.3: Energy bands of stress-free silicon in the first Brillouin zone: (a) Complete energy vs. wave vector diagram $E(k)$ for different directions. E_G is the gap between the valence and the conduction bands; (b) Equi-energy surfaces of the six conduction band edges in **k**-space. After Madelung [12].

combined with the expressions of the nonzero matrix elements of the secular determinant of Equation 2.19. This yields the equation:

$$E_{C0} + \Xi_{ij}^{11} e_{ij} + \sum_{b \neq m,n}^{B} \frac{H_{1b}^{\mathbf{k}\hat{\mathbf{p}}} H_{b1}^{\mathbf{k}\hat{\mathbf{p}}}}{E'_u - H_{bb}} - E'_C = 0 \tag{2.29}$$

where the reduced energy E'_u has been called the reduced conduction band energy E'_C, and where the unperturbed energy E_0 has been called the conduction band edge E_{C0}. As a result of the symmetry of the silicon crystal, the conduction band has six equivalent edges throughout the first Brillouin zone, as shown in Figure 2.3.b. They are located along the <100>directions and are indicated in the following by the superscript m, $m \in \{1, 2..6\}$. The conduction band equations can be rewritten in analogy to the strain-free case of Appendix B, Equation B.26. This gives:

$$E_C^m(\mathbf{k}) = E_{C0}^m + \Xi_{ij}^m e_{ij} + \frac{\hbar^2}{2} M_{ij}^m k_i k_j, \qquad i,j \in \{1,2,3\} \tag{2.30}$$

where M_{ij}^m is the inverse effective mass tensor of the conduction band and where the wave vector k_i is measured from the band edge at $\mathbf{k}_0^m$.

Symmetry of the deformation potential tensor

In analogy to the inverse effective mass tensors M_{ij}^m, the deformation potential tensors Ξ_{ij}^m of the silicon conduction band contain only two independent, nonzero

Table 2.1: Deformation potentials of silicon according to different authors. All potentials are given in eV, except α which is given in m_0^{-1}.

Ξ_d	Ξ_u	α	$(\Xi_d + \Xi_u/3 - a)$	a	b	d	**Ref.**
1.1	10.5		2.5	2.1	−2.33	−4.75	[5]
1.03	8.47		1.79	2.06	−2.27	−3.69	[14]
1.13	9.2		1.72	2.46	−2.35	−5.32	[15, 16]
				2.1	−1.5	−3.4	[17]
	8.1		1.6		−2.14	−5.1	[18]
	8.6		1.5		−2.10	−4.85	[19]
		86.8			−1.36	−3.09	[20, 21]

elements. These elements can be derived by applying the symmetry properties of the crystal to the tensor, as shown in Appendix B, Section B.5. For a band edge along the [100] direction this yields:

$$\Xi^1_{ij} = \begin{bmatrix} \Xi_l & 0 & 0 \\ 0 & \Xi_t & 0 \\ 0 & 0 & \Xi_t \end{bmatrix}; \quad M^1_{ij} = \begin{bmatrix} m_l^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 0 \\ 0 & 0 & m_t^{-1} \end{bmatrix} \tag{2.31}$$

where Ξ_l and Ξ_t are the longitudinal and transverse deformation potentials, respectively, and m_l and m_t the longitudinal and transverse effective masses. It is more common, however, to use the constants of Herring and Vogt [10, 13] who defined the *dilational deformation potential* Ξ_d and the *shear deformation potential* Ξ_u:

$$\Xi_d = \frac{1}{2}\left(\Xi^1_{22} + \Xi^1_{33}\right) \tag{2.32}$$

$$\Xi_u = \frac{1}{2}\left(2\Xi^1_{11} - \Xi^1_{22} - \Xi^1_{33}\right) \tag{2.33}$$

Numerical values of these potentials are available from measurements as well as from more fundamental calculations. Examples of these are given in Table 2.1. It should be noted that they show a rather large spread.

Shifts of the band edges

The effect of the strain term in Equation 2.30 is to generate an additional energy term independent of $\mathbf{k}$. This term can therefore be considered as a shift ΔE^m_{C0} of the band edge E^m_{C0}:

$$\Xi^m_{ij} e_{ij} = \Delta E^m_{C0}\left(e_{ij}\right) \tag{2.34}$$

The strain-induced energy shift is generally different for the various conduction band edges. If K^m_i is a unit vector in $\mathbf{k}$-space pointing to $\mathbf{k}^m_0$, the energy shift of an arbitrary band m can be expressed as [10]:

$$\Delta E_{C0}^{m} = \left(\Xi_d \delta_{ij} + \Xi_u K_i^m K_j^m\right) e_{ij} \tag{2.35}$$

It can be seen from this equation that the shifts are equal for the pairs of bands which are opposite in **k**-space. Band 1 is therefore equal to Band 4, Band 2 is equal to Band 5, and Band 3 is equal to Band 6, even under stress.

It is interesting to consider the shifts of the conduction band edges of silicon for three special types of strain. These types correspond to similar types of stress; see Appendix A.

Pure dilatation, firstly, is the strain for which all diagonal elements of e_{ij} are equal to a magnitude e and all off-diagonal elements are zero. In that case, Equation 2.35 predicts that all band edges are shifted with the same amount $(3\Xi_d + \Xi_u)\, e$. The result is that all equi-energy surfaces in Figure 2.4.a are equally enlargened.

Uniaxial strain along the [001]-axis, secondly, corresponds to a strain tensor with only one nonzero element, e_{33}. Its effect on the band edges in that direction is different from the effect on the edges in the perpendicular directions:

$$(\Xi_d + \Xi_u)\, e_{33} = \Delta E_{C0}^{3} \tag{2.36}$$

$$\Xi_d e_{33} = \Delta E_{C0}^{1} = \Delta E_{C0}^{2} \tag{2.37}$$

As a result, strain shifts the conduction band edges to different levels, which was already sketched in Figure 2.2. This is equivalent to the picture that the equi-energy surfaces of the [001]-edges are much more enlargened than the other edges, which is sketched in Figure 2.4.

Shear strain, finally, is described by a tensor with off-diagonal elements only. Such a strain has no influence on the band edge energies of silicon, as can be seen from Equation 2.35.

Deformation of the equi-energy surfaces

The conduction band expression of Equation 2.30 suggests that strain only shifts the band edges, and that shear strain has no influence. Experiments and theory have shown, however, that this is not entirely true for silicon [21, 22, 23]. Shear strain does have some influence, which becomes apparant through a modification of the inverse effective mass tensor.

The cause of the mass changes is that the minima of the conduction band, Δ_1, are relatively close to a higher-lying band, $\Delta_{2'}$. This is shown in Figure 2.5. The figure also shows that both bands touch each other at the points X at the zone boundary, which is a result of the cubic symmetry of the crystal. Therefore, the electron wave functions of Δ_1 and $\Delta_{2'}$ are coupled, and determine each other's curvature. This coupling is still perceptible at the minima of Δ_1 at $\mathbf{k}_0^m$.

When the crystal is subjected to a orthorhombic distortion e_{23}, the degeneracy is lifted at X in the [100]-direction. The two bands split with an amount $2E_1$, as shown in Figure 2.5.b. The coupling between the wave functions is thereby

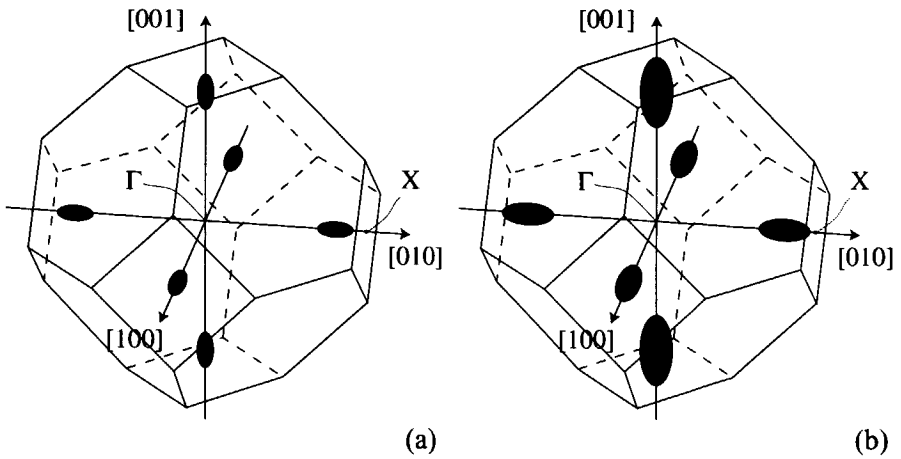

Figure 2.4: Equi-energy surfaces of the conduction band edges of silicon in the first Brillouin zone; (a) in the absence of strain; (b) in the presence of uniaxial strain along the [001]-axis.

partially removed. This results in a change of the curvature, and therefore in a modification of the effective mass. For a minimum in the [100]-direction, this can be described by the following inverse effective mass tensor:

$$M^1_{ij}(e_{kl}) = \begin{bmatrix} m_l^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 2\alpha e_{23} \\ 0 & 2\alpha e_{23} & m_t^{-1} \end{bmatrix} \tag{2.38}$$

where α is a proportionality constant. The value of α has been measured with the aid of cyclotron resonance and was found to be $(86.8 \pm 5.0)\ m_0^{-1}$ [21]. The minima in other directions are not affected by the strain e_{23}, but are sensitive to shear perpendicular to their longitudinal axis.

The introduction of off-diagonal elements in the effective mass tensor means that the principal axes of the mass ellipsoid no longer coincide with the crystal axes. Instead, the transverse axes are turned over $\pi/4$. Moreover, they are different from each other, which is illustrated in Figure 2.6.

2.1.5 Valence band equations

The second-order description of the valence bands is found by supposing that the edges are threefold degenerate at $\mathbf{k}_0$ in the unperturbed case. This is worked out in Appendix B, Section B.6. The eigenfunction indices m and n therefore range between 1 and 3. Also for this case a collection can be made of the nonzero matrix elements entering the secular determinant of Equation 2.19. The result

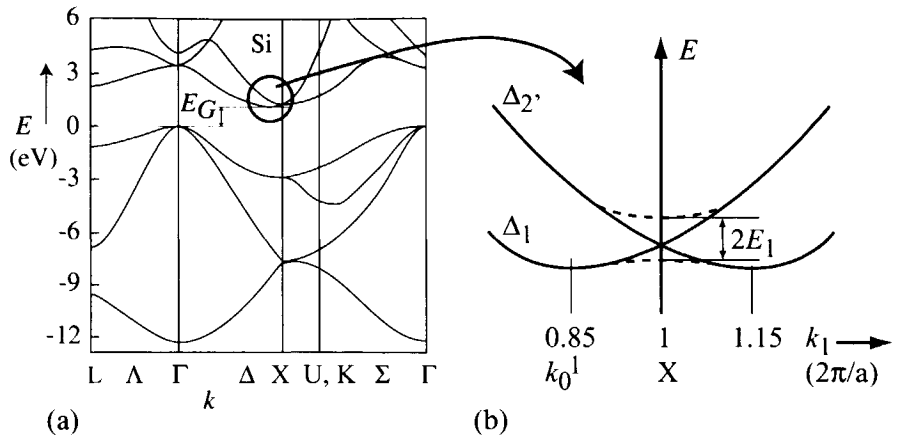

Figure 2.5: Degeneracy of the conduction band Δ_1 with the band $\Delta_{2'}$ at the point X: (a) Complete band diagram of the first Brillouin zone; (b) Detail of the degeneracy at the zone boundary, as well as its lifting with an energy $2E_1$ by a shear strain e_{23} [21].

is:

$$\det\left[\Xi_{ij}^{mn}e_{ij} + \sum_{b\neq m,n}^{B}\frac{H_{mb}^{\mathbf{k}\hat{\mathbf{p}}}H_{bn}^{\mathbf{k}\hat{\mathbf{p}}}}{E_u' - H_{bb}} - (E_u' - E_0)\,\delta_{mn}\right] = 0 \tag{2.39}$$

With the aid of the inverse effective mass tensor D_{ij}^{mn} this can also be written in a form showing the mathematical similarity between the strain- and the wave vector-dependent parts:

$$\det\left[\Xi_{ij}^{mn}e_{ij} + D_{ij}^{mn}k_ik_j - (E_u - E_0)\,\delta_{mn}\right] = 0 \tag{2.40}$$

where E_u is the unreduced energy eigenvalue. This, finally, can be written in a more abstract matrix notation:

$$\det\left[\mathbf{H}_e + \mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}} - E\delta_{mn}\right] = 0 \tag{2.41}$$

where E is an abbreviation of the energy difference $E_u - E_0$. The matrix $\mathbf{H}_e$ is called the strain Hamiltonian matrix.

It is seen in Section B.7, however, that the valence bands are also considerably influenced by the spin-orbit coupling. This coupling doubles the number of eigenfunctions to six, and therefore doubles the rank of $\mathbf{H}_e$ and $\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}$. Moreover, it adds a first-order perturbation term to the Hamiltonian. The coupling has no apparent influence, however, on the $\mathbf{k}\hat{\mathbf{p}}$ and strain perturbation terms. Its effect on the secular determinant of Equation 2.41 is to add the spin-orbit matrix $\mathbf{H}_{SO}$, derived in Section B.7:

$$\det\left[\mathbf{H}_e^{6\times6} + \mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}^{6\times6} + \mathbf{H}_{SO} - E\delta_{mn}\right] = 0 \tag{2.42}$$

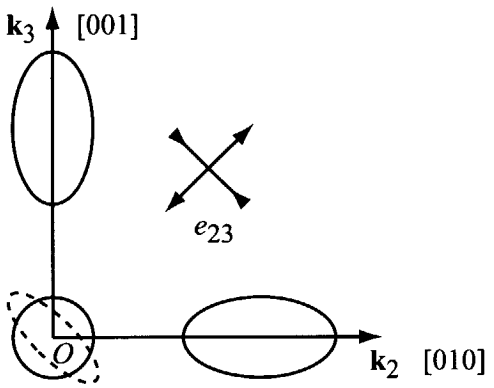

Figure 2.6: Cross section of three perpendicular conduction band ellipsoids under a shear strain e_{23}. The ellipsoids are projected onto the k_2k_3-plane. The strain changes the transverse axes of the ellipsoid in the k_1-direction, whereas it leaves unchanged the perpendicular ellipsoids [23].

$$\mathbf{H}_{SO} = \frac{\Delta_{SO}}{3} \begin{bmatrix} \mathbf{0} & -i\widehat{\sigma}_3 & i\widehat{\sigma}_2 \\ i\widehat{\sigma}_3 & \mathbf{0} & -i\widehat{\sigma}_1 \\ -i\widehat{\sigma}_2 & i\widehat{\sigma}_1 & \mathbf{0} \end{bmatrix} \tag{2.43}$$

where Δ_{SO} is the spin-orbit splitting energy and the $\widehat{\sigma}_i$ are the Pauli matrices. The rank of the matrices $\mathbf{H}_e$ and $\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}$ is explicitly indicated here. The matrices are obtained from the corresponding 3×3 matrices by multiplying each element with the 2×2 unity matrix $\mathbf{I}_2$.

Symmetry of the deformation potential tensor

The strain Hamiltonian matrix is considerably simplified by subjecting its deformation potential tensor Ξ_{ij}^{mn} to the symmetry properties of the crystal. For this purpose it is noticed that Ξ_{ij}^{mn} in Equation 2.40 appears mathematically in the same way as the effective mass tensor D_{ij}^{mn}. They therefore obey the same symmetry properties, which are developed for D_{ij}^{mn} in Appendix B, Section B.6. This yields the following $\mathbf{k}\hat{\mathbf{p}}$ matrix:

$$\begin{aligned} \mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}} &= D_{ij}^{mn} k_i k_j \\ &= \begin{bmatrix} Lk_1^2 + M\left(k_2^2 + k_3^2\right) & Nk_1k_2 & Nk_1k_3 \\ Nk_1k_2 & Lk_2^2 + M\left(k_1^2 + k_3^2\right) & Nk_2k_3 \\ Nk_1k_3 & Nk_2k_3 & Lk_3^2 + M\left(k_1^2 + k_2^2\right) \end{bmatrix} \end{aligned} \tag{2.44}$$

Table 2.2: Other definitions of the deformation potentials appearing in literature.

This work	[6]	[7]	[24]	[25]
l	$a+2b$	$D_d^v - \frac{4}{3}D_u$	$(d_1+2d_3)/\sqrt{3}$	
m	$a-b$	$D_d^v + \frac{2}{3}D_u$	$(d_1-d_3)/\sqrt{3}$	
n	$\sqrt{3}d$	$-2D'_u$	$\sqrt{3/2}d_5$	
Ξ_d			$d_1^1/\sqrt{3} - d_1^3/\sqrt{6}$	$E_1 - E_2/3$
Ξ_u			$\sqrt{3/2}d_1^3$	E_2

where L, M, and N are the valence band parameters. It also yields the following strain matrix:

$$\mathbf{H}_e = \Xi_{ij}^{mn} e_{ij} \tag{2.45}$$
$$= \begin{bmatrix} le_{11} + m(e_{22}+e_{33}) & ne_{12} & ne_{13} \\ ne_{12} & le_{22} + m(e_{11}+e_{33}) & ne_{23} \\ ne_{13} & ne_{23} & le_{33} + m(e_{11}+e_{22}) \end{bmatrix}$$

where l, m, and n are called the first-order deformation potentials of the valence bands. These potentials are defined as:

$$l = \Xi_{11}^{11}, \quad m = \Xi_{22}^{11}, \quad n = \Xi_{12}^{12} + \Xi_{12}^{21} \tag{2.46}$$

Often, they appear in literature as linear combinations of the parameters a, b, and d, or as a combination of D_d^v, D_u, and D'_u. Their conversion is given in Table 2.2, whereas numerical values are shown in Table 2.1. It should be noted that the conversion between the symbols is prone to mistakes, especially when the strain is defined as the so-called 'engineering strain' or 'conventional strain', instead of the usual 'tensor strain' (see Appendix A of Reference [20]).

Solution of the secular determinant

The solutions of the secular determinant in Equation 2.42 represent the valence bands in the presence of strain. In other words, they are the eigenvalues of the matrix $\mathbf{H} = \mathbf{H}_e^{6\times6} + \mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}^{6\times6} + \mathbf{H}_{SO}$ [26]:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} - i\widehat{\sigma}'_3 & h_{13} + i\widehat{\sigma}'_2 \\ h_{12} + i\widehat{\sigma}'_3 & h_{22} & h_{23} - i\widehat{\sigma}'_1 \\ h_{13} - i\widehat{\sigma}'_2 & h_{23} + i\widehat{\sigma}'_1 & h_{33} \end{bmatrix} \tag{2.47}$$

of which the elements are new 2×2 submatrices. In this Hamiltonian matrix the $\widehat{\sigma}'_i$ denote the reduced Pauli matrices and are equal to $\widehat{\sigma}_i \Delta_{SO}/3$. The diagonal elements h_{ii} are defined by:

$$h_{ii} = \left[Lk_i^2 + M\left(k_j^2 + k_k^2\right) + le_{ii} + m(e_{jj}+e_{kk})\right]\mathbf{I}_2 \tag{2.48}$$

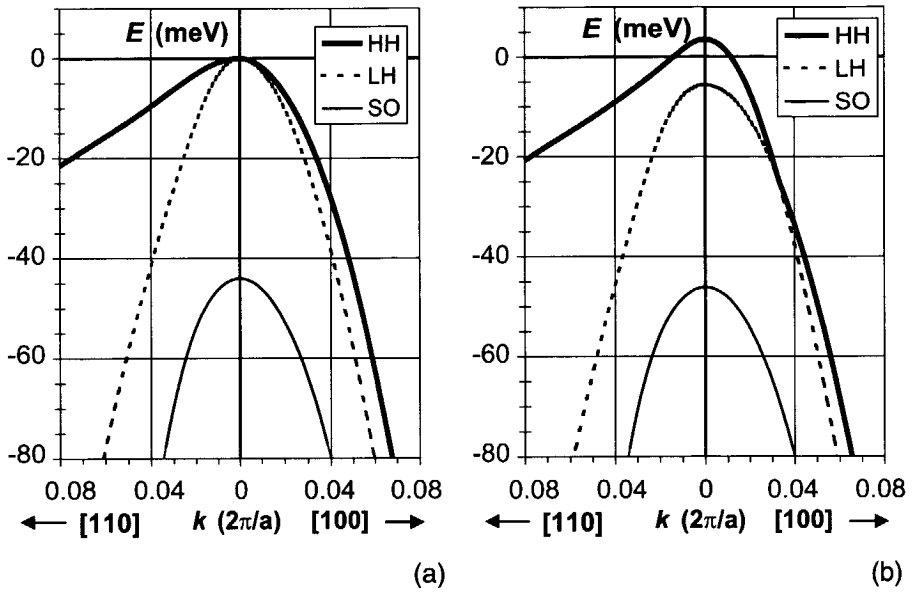

Figure 2.7: Valence band dispersion relations of Si in the [100]- and [110]-directions: (a) in the absence of stress, and (b) in the presence of 200 MPa uniaxial compressive stress along the [100]-direction. E is the energy, k the wave number, and a the atomic constant. HH, LH, and SO indicate the heavy-hole, light-hole, and split-off bands, respectively.

in which (i, j, k) is a cyclic permutation of $(1, 2, 3)$. Finally, the off-diagonal elements h_{ij} are defined as:

$$h_{ij} = (Nk_ik_j + ne_{ij})\,\mathbf{I}_2, \qquad i \neq j \tag{2.49}$$

The eigenvalues of $\mathbf{H}$ are given by the same expressions as in Appendix B, Section B.8 [26, 27, 28]. They are called the heavy-hole band $E_V^{\mathrm{HH}}(k_i, e_{ij})$, the light hole band $E_V^{\mathrm{LH}}(k_i, e_{ij})$, and the split-off band $E_V^{\mathrm{SO}}(k_i, e_{ij})$, respectively.

When the valence band functions are plotted the influence of stress and strain becomes immediately apparent, for instance in Figure 2.7. The most striking feature of this figure is the splitting of the degenerate band edges. This was to be expected from the fact that the strain destroys the cubic symmetry of the crystal, which makes the degeneracy no longer necessary. Another feature is that the shapes of the bands are deformed.

Shifts of the band edges

The splitting of the valence band edges is plotted in Figure 2.8 as a function of uniaxial stress. It is interesting to see that the heavy- and light-hole band edges do not cross each other at the origin. Instead, they repel each other,

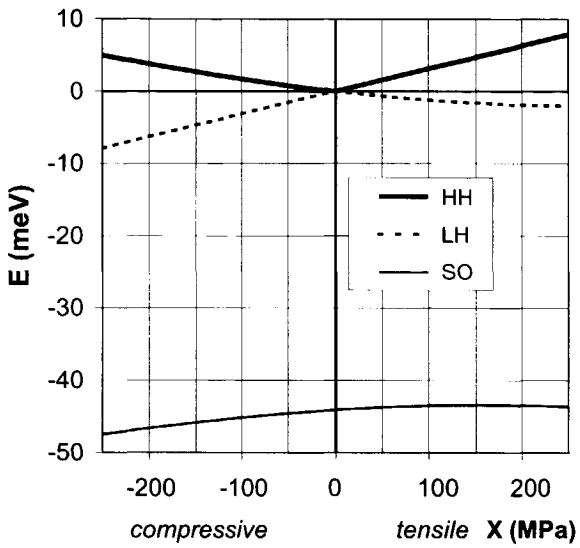

Figure 2.8: Edges of the valence bands E^u_{V0} as a function of a uniaxial stress along the [100]-axis. The edge of the heavy-hole band is indicated by 'HH', that of the light-hole band by 'LH', and that of the split-off band by 'SO'.

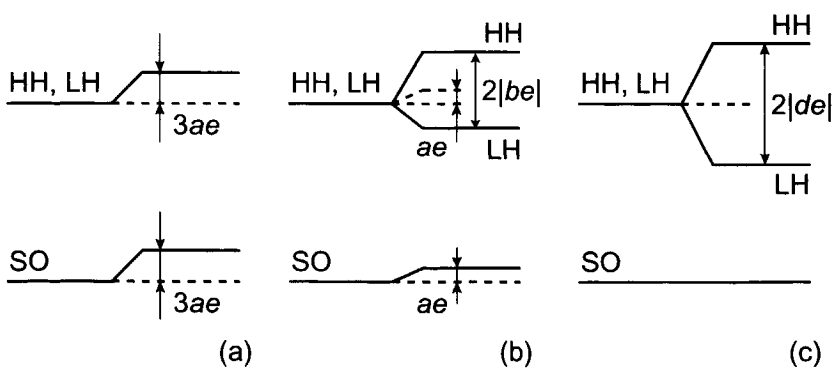

Figure 2.9: Schematic of the shifting of the valence band edges E^u_{V0} as a result of a strain e. (a): pure dilatation, (b): normal uniaxial strain, (c): pure shear.

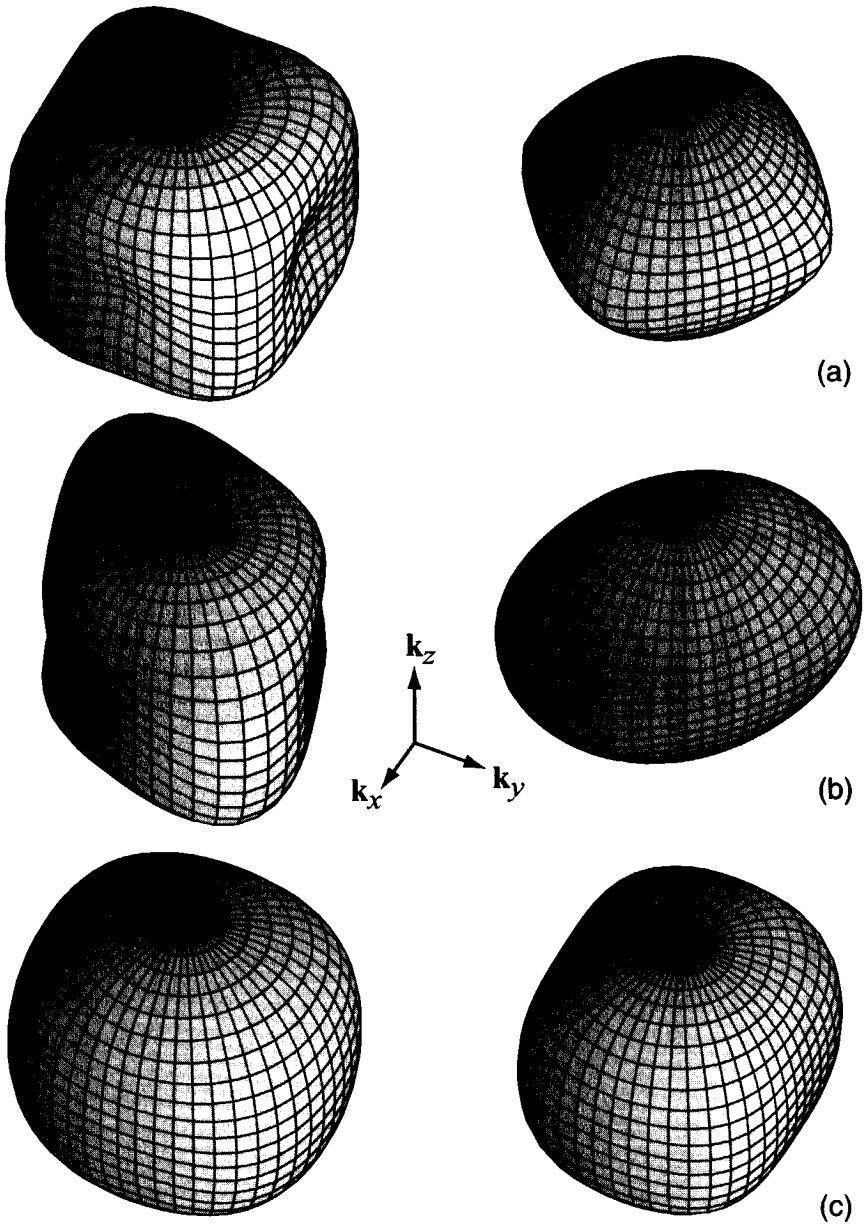

Figure 2.10: Valence band equi-energy surfaces for different stresses, at 1 meV below the band edge. Left: heavy-hole band; Right: light-hole band. (a): stress-free case, (b): 20 MPa shear stress in the [110]-direction, (c): -50 MPa normal stress in the [100]-direction.

bend rather sharply, and take over each other's trajectory. This effect is due to the quantum-mechanical interaction between the states of the bands, which is allowed due to the equal parity and quantum number of the states [29, 30].

The shifting of the band edges can be studied more in detail to find the role of the different deformation potentials. This is done by analysing the band functions E_V^{HH}, E_V^{LH}, and E_V^{SO} at $\mathbf{k} = 0$. Three special strains are again considered: pure dilatation, normal uniaxial strain, and shear strain. Their effects are sketched in Figure 2.9.

Pure dilatation appears to cause a rigid shift of all band edges with an amount $3ae$, where a is the deformation potential $(l + 2m)/3$, and e the magnitude of the dilatation. It does not cause any edge splitting. The reason for this is that it leaves intact the cubic symmetry of the crystal.

A normal uniaxial strain e also causes a rigid shift of all band edges, with an amount ae. More important, however, is that it splits the edges of the heavy- and light-hole bands with an amount $2|be|$, where b is the deformation potential $(l - m)/3$. An interesting feature of the splitting is the absolute value function. As a result, the heavy-hole band always stays above the light-hole band, whether the strain is positive or negative. The splitting is linear in e for strains well below $\Delta_{SO}/3b$. Usually, this is the case because the limit corresponds to 6 me (millistrain) and a stress of about 1 GPa. This is four times higher than the range considered in this thesis.

Pure shear strain, finally, also splits the edges of the heavy- and light-hole bands. This splitting amounts to $2|de|$, where d is the deformation potential $n/\sqrt{3}$. The linearity of the splitting holds for $e \ll \Delta_{SO}/3d$. The rigid shift of all edges is negligible under the same condition.

Deformation of the equi-energy surfaces

The strain not only shifts the valence band edges, but also deforms their equi-energy surfaces. Examples of this are given in Figure 2.10. The deformation of the valence band surfaces is strong, much stronger than for the conduction bands. The reason is that in the absence of strain the valence bands are very close to each other and thereby influence each other's shape. As a result, the equi-energy surfaces are warped instead of spherical. In the presence of strain, however, the bands are split, which partially decouples them and decreases the warping. For very high strains, the bands are even completely decoupled and have ellipsoidal equi-energy surfaces, just as the conduction bands.

Another interesting phenomenon is that for specific orientations, strain seems to press the light-hole band through the heavy-hole band. This is shown in Figure 2.7.b, for $\mathbf{k}$ between 0 and $0.25 \times 2\pi/a$ in the [100]-direction. Within this interval, the heavy- and light-hole bands have interchanged their shape with respect to the strain-free case. The transition with the shape outside the interval is sharp, with a discontinuous derivative. This is a result of the quantummechanical origin of the bands.

The deformation of the equi-energy surfaces appears to be the main cause of the piezoresistive effect for holes. The piezoresistive effect for electrons, on

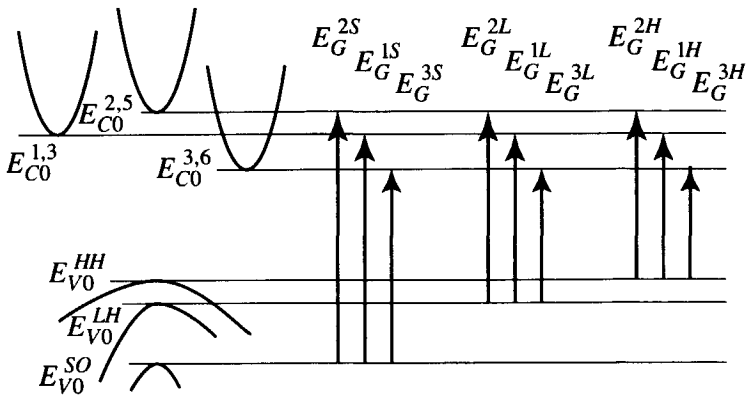

Figure 2.11: Schematic representation of the different bandgaps E_G^{mu} that may arise in silicon when it is subject to stress.

the contrary, is dominated by the shifts of their conduction band edges.

2.1.6 Bandgap definitions

In the preceeding sections it was seen that anisotropic stress lifts the degeneracy of the band edges. This makes it difficult to speak about 'the' bandgap of silicon, although this is a widely employed concept in electronics and in charge transport theory.

Usually, the bandgap E_G is defined as the energy necessary to excite an electron from the maximum of the heavy- and light-hole valence bands to one of the six equivalent minima of the conduction band. The gap with the split-off band is generally neglected.

Stress, however, yields three different minima and three different maxima, if the split-off band is also considered. This gives *nine* different gaps, which is illustrated in Figure 2.11. It is therefore necessary to specify the gaps with the band coefficients m and u.

2.1.7 Band edge diagrams

In the calculation of charge transport the band edge energies often need to be known as a function of the position, such as in Section 2.2 and Appendix C. These functions can be represented graphically in a band edge diagram, of which an example is shown in Figure 2.12. The course of the edges is determined by both the crystal potential and the external fields of force, such as stress and the electrostatic potential ϕ. The band edges need to be related to this potential by means of a reference energy level.

In stressed material the reference energy level cannot be chosen in the way adapted in most textbooks. There, the reference is usually defined by the 'in-

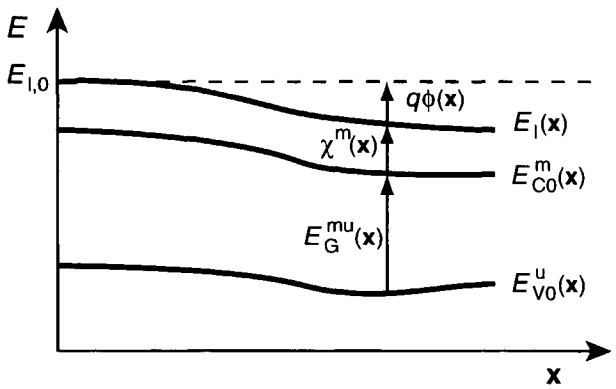

Figure 2.12: Band edge structure of a nonhomogeneous semiconductor as a function of position $\mathbf{x}$. Indicated are the local vacuum level E_l, the edge E_{C0}^m of conduction band m, and the edge E_{V0}^u of valence band u. After Marshak and Van Vliet [31].

trinsic level', which is calculated from the bandgap and the effective densities of states of the bands. It was shown in the previous section, however, that the bandgap is not unique under stress. The situation becomes even worse when the stress is inhomogeneous and the edges vary with position even in the absence of an electrostatic potential.

For stressed material it is therefore much better to choose the reference as the energy of an electron in vacuum. This approach is also used to describe heterojunctions, heavily doped regions, and non-isothermal conditions. As a result, the other quantities in the band edge diagram also need to be redefined. This is done here according to Marshak and Van Vliet [31] and relates the energy bands to the electrostatic potential:

- $E_{l,0}$ represents the energy of an electron at rest, free from the crystal potential, and on a position where the electrostatic potential ϕ is by definition zero.

- $\phi(\mathbf{x})$ is the electrostatic potential and is forced on the crystal from both external sources and internal space charge. It is a continuous function of the position $\mathbf{x}$.

- $E_l(\mathbf{x})$ is called the local vacuum level. It represents the energy of an electron at rest, free from the crystal potential, but with an electric field energy $q\phi(\mathbf{x})$. It is therefore suitable to serve as a reference level in charge transport calculations. Since $\phi(\mathbf{x})$ is continuous, $E_l(\mathbf{x})$ is continuous as well.

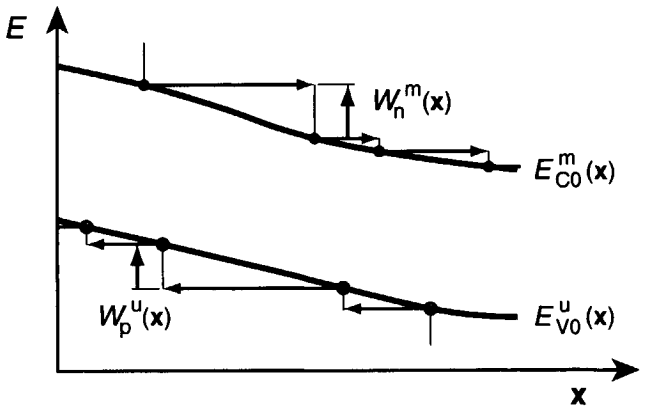

Figure 2.13: Kinetic energy of electrons W_n^m and holes W_p^u which are moving through the bands m and u of a semiconductor. The energy is continuously increased by the gradients in the band edges, and is decreased from time to time by scattering events.

- $\chi^m(\mathbf{x})$ is the electron affinity of conduction band m. It is the energy necessary to free an electron from the conduction band edge E_{C0}^m to the local vacuum level $E_l(\mathbf{x})$.
- $E_{C0}^m(\mathbf{x})$ is the energy of the conduction band edge m, thus at the minimum $\mathbf{k}_0^m$ in k-space. Partially, it depends on the electrostatic potential and partially on the crystal potential. It represents the total potential energy of an electron in that band on that position.
- $E_G^{mu}(\mathbf{x})$ is the bandgap between the conduction band m and the valence band u. It is equal to the energy that needs to be supplied to create an electron-hole pair.
- $E_{V0}^u(\mathbf{x})$, finally, is the edge of valence band u and represents the total potential energy of a hole in that band on that position.

With the aid of those definitions it is possible to define a Hamiltonian operator describing the total energy of electrons and holes in a band. By inspection of Figure 2.12 the local *potential energy* can be written as:

$$E_{C0}^m(\mathbf{x}) = E_{l,0} - \chi^m(\mathbf{x}) - q\phi(\mathbf{x}) \tag{2.50}$$

$$E_{V0}^u(\mathbf{x}) = E_{l,0} - \chi^m(\mathbf{x}) - E_G^{mu}(\mathbf{x}) - q\phi(\mathbf{x}) \tag{2.51}$$

Due to their velocity, electrons and holes also have a *kinetic energy,* denoted by $W_n^m(\mathbf{k},\mathbf{x})$ and $W_p^u(\mathbf{k},\mathbf{x})$, respectively. The kinetic energy of an electron is increased continuously by the gradient in the band edge. However, it is decreased from time to time by scattering on phonons, ionised impurities, and

lattice imperfections. This is sketched in Figure 2.13. Therefore, the *total energy* of electrons and holes in a band is equal to:

$$E_C^m(\mathbf{x},\mathbf{k}) = E_{C0}^m(\mathbf{x}) + W_n^m(\mathbf{x},\mathbf{k}) = \widehat{H}_C^m \tag{2.52}$$

$$E_V^u(\mathbf{x},\mathbf{k}) = E_{V0}^u(\mathbf{x}) - W_p^u(\mathbf{x},\mathbf{k}) = \widehat{H}_V^u \tag{2.53}$$

where $\widehat{H}_C^m$ is the semiclassical Hamiltonian for the conduction band m, and $\widehat{H}_V^u$ is the Hamiltonian for the valence band u.

2.2 Microscopic changes in charge transport

The stress-induced changes in the energy bands must have consequences for the ease with which charge carriers can be transported through the material. The equations for this transport can be derived in essentially the same way as for unstressed material (see e.g. [2, 33]). Within the derivation, however, care should be taken not to assume that the material is isotropic or homogeneous. Instead, one should use the definition of the bandgaps and the band edge structure of Section 2.1.6 and 2.1.7 [31]. This is worked out in Appendix C, while the results are applied throughout the following sections.

This section starts with an overview of the models from which charge transport can be calculated, and the assumptions on which they are based. Then it presents equations for charge concentration under stress. It develops the conductivity in terms of band parameters, which is in fact the most important transport parameter for the piezojunction effect. Finally, it discusses the stress effects on auxiliary parameters, such as the pn-product, the carrier lifetime, and the dielectric constant.

2.2.1 Available transport models

For the description of electron and hole transport, models of different complexity and precision exist. Four possible models are given in Figure 2.14 [34]. Each model arises from the preceding one and includes additional assumptions. It is useful to be aware of them to know the limits of the validity of a model's predictions.

Quantum transport theory

The most fundamental theory on electron and hole motion is the quantum transport theory. In this theory, electrons are considered as wave functions moving in the periodic potential of the crystal lattice, their motion being determined by the Schrödinger equation. However, if the lattice is much larger than the De Broglie wavelength, it can contain a huge number of electrons which all interact with each other. It is therefore impossible to describe the motion of each electron separately.

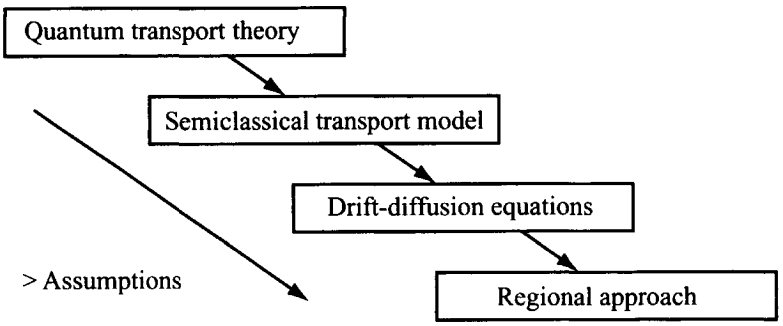

Figure 2.14: Models of charge carrier transport in semiconductors. Each model arises from the previous one and includes further assumptions [34].

Semiclassical transport model

The large number of electrons makes a statistical description necessary, which is provided by the semiclassical transport model. In this model, the electron wave functions are considered as particles. These particles have an effective mass m^* and a crystal momentum $\hbar\mathbf{k}$ centred around some $\hbar\mathbf{k}_0$. It is supposed that:

- The electrons remain in the same band m of the material.
- Their energy is determined by the band equation $E^m(\mathbf{k})$ (see Section 2.1 and Appendix B).
- They move, between collisions, according to the classical equations of motion.
- They have a distribution of momentum given by Fermi-Dirac statistics.

The development of the distribution in time and space is given by the Boltzmann transport equation (see Appendix C.2). By using this equation a number of additional assumptions is made:

- Quantum effects can be neglected;
- Electrons and holes form a gas of free and independent particles;
- The fields and concentration gradients only vary slowly with respect to the mean path between two collisions;
- Collisions are instantaneous;
- The electric field is low enough for scattering being independent of it.

Unfortunately, the exact distribution function is very difficult to calculate and the result is not practical for the description of transistor behaviour. It may serve, nevertheless, as a starting point for the derivation of simpler equations: the drift-diffusion equations.

Drift-diffusion equations

The drift-diffusion or Shockley equations can be derived from a first-order perturbation of the distribution function, as shown in Appendix C. They give the current density vectors of electrons and holes $\mathbf{J}^{\mathrm{n}}$ and $\mathbf{J}^{\mathrm{p}}$, respectively, as a function of the electric field $\mathbf{F}^{\mathrm{el}}$ and the gradients in the concentrations n and p with respect to the position $\mathbf{x}$. Conventionally, the equations are written as:

$$\mathbf{J}^{\mathrm{n}} = qn\mu^{\mathrm{n}}\mathbf{F}^{\mathrm{el}} + qD^{\mathrm{n}}\nabla n \tag{2.54}$$

$$\mathbf{J}^{\mathrm{p}} = qp\mu^{\mathrm{p}}\mathbf{F}^{\mathrm{el}} - qD^{\mathrm{p}}\nabla p \tag{2.55}$$

where q is the elementary charge, $\mu^{\mathrm{n,p}}$ are the electron and hole mobilities, respectively, and $D^{\mathrm{n,p}}$ are the electron and hole diffusion constants. The equations have widely been applied for the design of semiconductor devices because they enable accurate modelling of their most important properties. In addition, they are remarkably successful in incorporation quantum phenomenon such as band-to-band tunnelling.

To derive both the drift-diffusion equations and expressions for μ^{n}, μ^{p}, D^{n}, and D^{p} the semiclassical transport model is used in combination with a number of further assumptions. Most of them can be justified quite easily:

- The material has a homogeneous temperature;
- The scattering of electrons and holes is elastic;
- The mean time between two collisions τ only depends on the wave vector $\mathbf{k}$ through the energy E: $\tau\left[E\left(\mathbf{k}\right)\right]$;
- The scattering hardly changes the shape of the distribution function;
- The spatial variation of τ is small compared to the mean free path.

Unfortunately, the equations are also based on assumptions which in general are invalid under stress:

- The semiconductor is isotropic;
- The semiconductor is homogeneous.

Finally, they are based on some assumptions we often prefer not to make even in the case of unstressed material:

- The material is nondegenerate;
- The energy bands $E\left(\mathbf{k}\right)$ are parabolic.

As a result of the above limitations the conventional drift-diffusion equations should be amended. This is done in Appendix C, while the results are resumed in Section 2.2.3.

Regional approach

The simplest transport model of Figure 2.14 is the regional approach, often used for manual calculations of semiconductor device behaviour. In this approach a device is considered to be built up of one-dimensional regions, which are assumed to be either quasi-neutral (zero field) or depleted (zero free carrier concentration). This allows for each region an analytical solution of the drift-diffusion equation, after which the solutions of all regions can be combined by requiring continuity at the interfaces.

The regional approach provides insight in many aspects of the device behaviour while it is remarkably precise in view of the large number of assumptions it contains. It leads to the compact analytical models desired in circuit design, such as the current-voltage relation derived in Appendix D. In particular, it leads to the model of transistor behaviour under stress presented in Section 2.4.

2.2.2 Charge carrier concentrations

The concentration of charge carriers in a band is important for the transport properties of the material. A high concentration facilitates the charge transport and increases the conductivity. It is, however, also important to know the concentration in one band relative to that in other bands. The reason is that each band imposes a specific effective mass to its charge carriers. This mass contributes more to the average mass with respect to all bands if the band is filled to a higher level. Mathematically, the concentration gives an expression for the Fermi level which can be used to calculated the conductivity.

General expressions

It was argued in Appendix C that the concentration of electrons in a band depends on its density of states and also on the distribution of electrons across those states. In the relaxation time approximation the distribution is assumed to be given by Fermi-Dirac statistics, even outside thermodynamic equilibrium. In that case, the electron concentration n^m due to a band m and the hole concentration p^u due to a band u are given by:

$$n^m(\mathbf{x}) = \int_C \frac{d\mathbf{k}}{4\pi^3} \left[1 + \exp\left(\frac{E_C^m(\mathbf{k},\mathbf{x}) - E_{Fn}(\mathbf{x})}{k_B T}\right)\right]^{-1} \tag{2.56}$$

$$p^u(\mathbf{x}) = \int_C \frac{d\mathbf{k}}{4\pi^3} \left[1 + \exp\left(\frac{E_{Fp}(\mathbf{x}) - E_V^u(\mathbf{k},\mathbf{x})}{k_B T}\right)\right]^{-1} \tag{2.57}$$

where the integrals are taken over the first Brillouin zone, E_C^m and E_V^u are the band dispersion relations, k_B is the Boltzmann constant and T the absolute temperature. E_{Fn} and E_{Fp} are the electron and hole quasi Fermi levels. They represent the total concentration of a carrier type, summed over all bands, and not necessarily in thermal equilibrium. In equilibrium, however, E_{Fn} and E_{Fp} are equal.

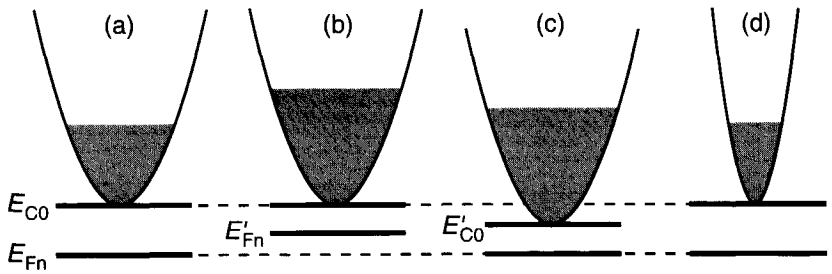

Figure 2.15: Changes in a conduction band which influence the electron concentration. The concentration is represented by the size of the gray area. (a) Original situation, (b) shift of the quasi-Fermi level, (c) shift of the band edge, (d) deformation of the curvature.

The concentration in a band is influenced by mechanical stress through the band equation as well as through the (quasi-)Fermi level. It was shown in Section 2.1 that this may deform the band and shift it with respect to the Fermi level. This is illustrated in Figure 2.15. The average effect over all bands is reflected by the Fermi level, which therefore also shifts with stress. This Fermi level shift depends on whether the charge carrier is a majority or a minority in the material, which will be developed in Section 2.3.1.

Expressions for nondegenerate material

The mathematical analysis of the concentrations under stress is much simplified when the material considered is nondegenerate, i.e. when the distance between the bands and the quasi Fermi levels is larger than 2.3 $k_B T$. For silicon this means that the doping concentration is less than 10^{18} cm^{-3} and that the carrier concentrations due to injection remain below that value (low-level injection). Neither of these assumptions is very restrictive for the treatment of the piezo-junction effect. In that case the Fermi-Dirac statistics differ less than ten percent from Maxwell-Boltzmann statistics, meaning that Equation 2.56 and 2.57 can be replaced by:

$$n^m(\mathbf{x}) \equiv \int_C \frac{d\mathbf{k}}{4\pi^3} \exp\left(\frac{E_{Fn}(\mathbf{x}) - E_C^m(\mathbf{k},\mathbf{x})}{k_B T}\right) \tag{2.58}$$

$$p^u(\mathbf{x}) \equiv \int_C \frac{d\mathbf{k}}{4\pi^3} \exp\left(\frac{E_V^u(\mathbf{k},\mathbf{x}) - E_{Fp}(\mathbf{x})}{k_B T}\right) \tag{2.59}$$

For nondegenerate material it is customary to introduce the concept of the *effective density of states* (DOS). According to Appendix C, Section C.1.3 the effective densities of states of a conduction band m and a valence band u are

defined as:

$$N_C^m \equiv \int_C \frac{d\mathbf{k}}{4\pi^3} \exp\left(\frac{-W_n^m(\mathbf{k},\mathbf{x})}{k_B T}\right), \quad N_V^u \equiv \int_C \frac{d\mathbf{k}}{4\pi^3} \exp\left(\frac{-W_p^u(\mathbf{k},\mathbf{x})}{k_B T}\right) \tag{2.60}$$

where W_n^m and W_p^u are the electron and hole kinetic energies. These energies are defined as the total energy minus the energy of the band edge (see Section 2.1.4).

The effective density of states N_C^m is a measure of the volume of conduction band m and only depends on its *shape*. When it is used to express the electron concentration, it separates the influence of the shape from that of the band *edge* E_{C0}^m and the quasi Fermi level E_{Fn}. For the hole concentration a similar separation is made, yielding:

$$n^m(\mathbf{x}) = N_C^m(\mathbf{x}) \exp\left(\frac{E_{Fn}(\mathbf{x}) - E_{C0}^m(\mathbf{x})}{k_B T}\right) \tag{2.61}$$

$$p^u(\mathbf{x}) = N_V^u(\mathbf{x}) \exp\left(\frac{E_{V0}^u(\mathbf{x}) - E_{Fp}(\mathbf{x})}{k_B T}\right) \tag{2.62}$$

The separation is useful in the stress-free case and in the description of the conduction bands under stress. However, it appears to be confusing in the description of the valence bands under stress, as will be shown in the following. It is therefore useful to introduce the concept of the available density of states.

Definition of the 'available density of states'

The *available density of states* is a new quantity with a meaning similar to that of the effective density of states. For a conduction band m and a valence band u the available densities of states F_{C}^m and F_{V}^u are defined as:

$$F_{\mathrm{C}}^m(\mathbf{x}) \equiv \int_C \frac{d\mathbf{k}}{4\pi^3} \exp\left(\frac{-E_C^m(\mathbf{k},\mathbf{x})}{k_B T}\right) = \exp\left(\frac{-E_{Fn}(\mathbf{k},\mathbf{x})}{k_B T}\right) n^m(\mathbf{x}) \tag{2.63}$$

$$F_{\mathrm{V}}^u(\mathbf{x}) \equiv \int_C \frac{d\mathbf{k}}{4\pi^3} \exp\left(\frac{E_V^u(\mathbf{k},\mathbf{x})}{k_B T}\right) = \exp\left(\frac{E_{Fp}(\mathbf{k},\mathbf{x})}{k_B T}\right) p^u(\mathbf{x}) \tag{2.64}$$

The available DOS include both the shape and the level of the band. They relate the quasi-Fermi levels to the concentration in the bands. However, they do not explicitely incorporate the energy of the band edges. Together with the quasi Fermi levels they entirely determine the charge carrier concentration due to the band. In nondegenerate material, the total electron concentration n and hole concentration p are therefore given by:

$$n(\mathbf{x}) = \exp\left(\frac{E_{Fn}(\mathbf{x})}{k_B T}\right) F_{\mathrm{C}}(\mathbf{x}), \qquad F_{\mathrm{C}}(\mathbf{x}) \equiv \sum_{m=1}^{6} F_{\mathrm{C}}^m(\mathbf{x}) \tag{2.65}$$

$$p(\mathbf{x}) = \exp\left(\frac{-E_{Fp}(\mathbf{x})}{k_B T}\right) F_{\mathrm{V}}(\mathbf{x}), \qquad F_{\mathrm{V}}(\mathbf{x}) \equiv \sum_{u=1}^{3} F_{\mathrm{V}}^u(\mathbf{x}) \tag{2.66}$$

where F_C and F_V are defined as the *total available density of states* of the conduction and valence bands, respectively. It should be noted that the ratio of the different F_V^u is equal to the ratio of the contributions of each valence band to the total hole concentration.

Strictly speaking, the definition of the available densities of states is unnecessary for the calculations in the following. It will appear, however, that it has three important advantages over the effective density of states:

- It allows a very compact description of the piezojunction effect;
- It shows the tight relationship between the piezojunction and the piezoresistive effects;
- It does not reflect any quantum effects under the influence of stress.

Changes through band edge shifts

Mechanical stress influences the charge carrier concentrations most directly by shifting the levels of the band edges. Those shifts were described in Section 2.1.4 by means of the deformation potential theory. Usually they are in the order of meV for the stress range considered.

According to Equation 2.61, a stress X_{ij} shifts a conduction band edge with ΔE_{C0}^m and thereby changes the electron concentration of band m with the following factor:

$$\frac{n^m(\mathbf{x},X_{ij})}{n^m(\mathbf{x},0)} = \exp\left(\frac{-\Delta E_{C0}^m(\mathbf{x},X_{ij})}{k_B T}\right) \tag{2.67}$$

For small stresses this change is linear. Over the stress range considered, however, n^m can change between -50 to +100%, which means that the first-order approach should be abandoned. Similar changes occur in the hole concentrations and the valence band edges. However, for those bands the edge shifts are tightly coupled to the changes in the effective DOS.

Changes through variations in the effective densities of states

Stress not only influences the charge carrier concentrations through band edge shifts, but also through the effective densities of states. For the stress range considered this influence is small for the conduction bands. However, it is highly significant for the valence bands. For both bands it is calculated by evaluating the integrals of Equation 2.60.

Conduction bands The conduction band density of states can be evaluated analytically. The method of this evaluation is in principle the same as in Appendix C.1.4, but requires a few modifications. It was seen in Section 2.1.4 that

the kinetic energy W_n^1 for an electron in a band along the k_1-axis is given by:

$$W_n^1 = \frac{\hbar^2}{2}\left[M_{11}^1\left(k_1 - k_1^0\right)^2 + \sum_{i,j=2,3} M_{ij}^1 k_i k_j\right] \tag{2.68}$$

where its effective mass tensor M_{ij}^1 is equal to:

$$M_{ij}^1 = \begin{bmatrix} m_l^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 2\alpha e_{23} \\ 0 & 2\alpha e_{23} & m_t^{-1} \end{bmatrix} \tag{2.69}$$

and depends on the strain e_{ij}. Because of the off-diagonal terms in M_{ij}^1 it is not directly possible to apply a Herring transformation and switch to spherical coordinates. The mass tensor should therefore first be diagonalised to $\widetilde{M}_{kl}$:

$$M_{ij}^1 = a_{ik}\widetilde{M}_{kl}a_{jl} \tag{2.70}$$

where a_{ik} is a orthogonal transformation matrix of which the columns are the orthonormalised eigenvectors of M_{ij}^1, and where $\widetilde{M}_{kl}$ has the eigenvalues of M_{ij}^1 on its main diagonal. Those eigenvalues are:

$$\lambda_1 = \frac{1}{m_l}, \quad \lambda_2 = \frac{1}{m_t} - 2\alpha e_{23}, \quad \lambda_3 = \frac{1}{m_t} + 2\alpha e_{23} \tag{2.71}$$

The transformation a_{ik} can also be applied to the wave vector, yielding a new orthogonal basis: $q_l = a_{jl}\left(k_j - k_j^0\right)$. The energy of Equation C.15 can then be written as:

$$W_n^1 = \frac{\hbar^2}{2}\widetilde{M}_{kk}q_k^2 \tag{2.72}$$

This equation allows the Herring transformation $\kappa_i = \widetilde{M}_{ii}^{1/2}q_i$.

The definition of the DOS in Equation 2.60 also contains the infenitesimal volume element $d\mathbf{k}$. This element remains unchanged by the diagonalisation to the basis $\mathbf{q}$. The reason is that a_{ik} is a unitary transformation, having a Jacobian determinant equal to one. On the contrary, the volume element is changed by the Herring transformation to the variable κ. This means that it can be written as:

$$d\mathbf{k} = \det\left(a_{ij}\right)^{-1} d\mathbf{q} = \det\left(\widetilde{M}_{ij}^{\frac{1}{2}}\right)^{-1} d\kappa = \left(m_{dC}^1\right)^{\frac{3}{2}} d\kappa \tag{2.73}$$

where m_{dC}^1 is defined as the density of states effective mass of the band. This mass is given by:

$$m_{dC}^1 = \sqrt[3]{\frac{m_l m_t m_t}{1 - \left(2\alpha e_{23} m_t\right)^2}} \tag{2.74}$$

Table 2.3: Total effective density of states of the valence bands N_V, density-of-states effective mass m^*_{dV}, and intrinsic carrier concentration n_{i0}, calculated with the valence band parameters of different authors. The values of the lower two rows have been found by Green and serve as a reference. The calculational accuracy was 0.1 percent.

N_V $(10^{19}\ \mathrm{cm}^{-3})$	m^*_{dV} (m_0)	n_{i0} $(10^{10}\ \mathrm{cm}^{-3})$	**Source of parameters**
1.872	0.822	0.787	Stickler *e.a.* [39, 43]
2.060	0.877	0.827	Dexter *e.a.* [40]
2.269	0.935	0.873	Fischetti and Laux [5]
2.777	1.070	0.972	Hensel [41]
3.10	1.15	1.07	Green, theory [45]
2.9(7)	1.10(18)	-	Green, experiments [45]

The remainder of the evaluation is the same as in Appendix C.1.4, leading to the following expression for the DOS of Band 1:

$$N_C^1 = 2\left(\frac{m_{dC}^1 k_B T(\mathbf{x})}{2\pi\hbar^2}\right)^{3/2} \tag{2.75}$$

The above effective density of states can be evaluated numerically with the values of Table 2.1 and B.2. In the strain-free case it yields $4.48 \times 10^{18}\ \mathrm{cm}^{-3}$ at $T = 300$ K. The total density of states of all six bands is therefore $2.75 \times 10^{19}\ \mathrm{cm}^{-3}$, which is close to the literature value of $2.86 \times 10^{19}\ \mathrm{cm}^{-3}$ [45].

In the case of nonzero strain the change in N_C^1 is only of second order. In addition, it remains inferior to 0.06%, because the strain does not exceed 0.2% for the stress range considered. As a result, the change in effective density of states has much less effect on the electron concentration than the band edge shifts.

Valence bands For the valence bands the density of states integrals of Equation 2.60 are too complex to be evaluated analytically. This evaluation was therefore done by numerical means. For this purpose, the integrand was implemented in Fortran 90 [35] and integrated over the wave vector space using a triple Romberg integration routine [36, 37, 38]. A schematic flow chart of this program is given at the end of this chapter. The evaluation time was 11 s on a personal computer with a Pentium II-400 MHz processor and for a numerical accuracy better than 0.1%.

The results of the evaluation are given in Table 2.3 for the input parameters of different authors. They include the total DOS N_V at zero stress, which is the sum of the densities of states of the different bands N_V^u. The results also include the values of the total DOS effective mass m^*_{dV} and the pn product n_{i0}^2. The latter quantity depends on the effective DOS of both the valence and the conduction bands, as will be explained in Section 2.2.4.

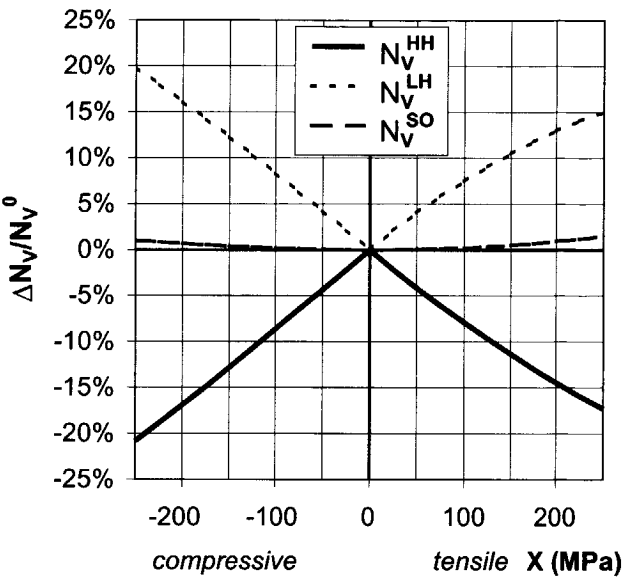

Figure 2.16: Changes in the effective densities of states of the valence bands of silicon calculated as a function of uniaxial stress in the [100] direction of the crystal. The band parameters for those calculations were taken from Hensel [41], and the deformation potentials from Friedel [14]. $T = 296$ K.

The results in the table vary with about 30%, depending on which set of band parameters is used. Nevertheless, the values obtained are comparable with the theoretical and experimental values reported by Green in a review article [45]. Green's values are best approached by the band parameter set of Hensel [41].

The same numerical routines can be used to calculate the stress-induced changes in the N_V^u of the different bands change under stress. Typical curves of those changes are shown in Figure 2.16. They are almost linear for the heavy and light hole band , and up to 20% in magnitude for the considered stress domain. As a result, they influence the hole concentrations with an amount comparable to the effect of the band edge shifts [42, 43, 44]. The curve of the split-off band DOS, however, only changes with a negligible amount.

The curves in Figure 2.16 show some unnatural behaviour: they have a sharp peak in the origin. A closer analysis learns that this peak results from the shifts of the band edges. A similar peak was seen already in Figure 2.8 where the band edges are plotted as a function of stress. This peak could be attributed to the quantum-mechanical interaction between the heavy- and light-hole states. As pointed out in Section 2.2.1, such quantum effects are rather unwelcome at this stage of modelling of charge transport.

All peaks cancel each other when the hole concentration p^u is calculated

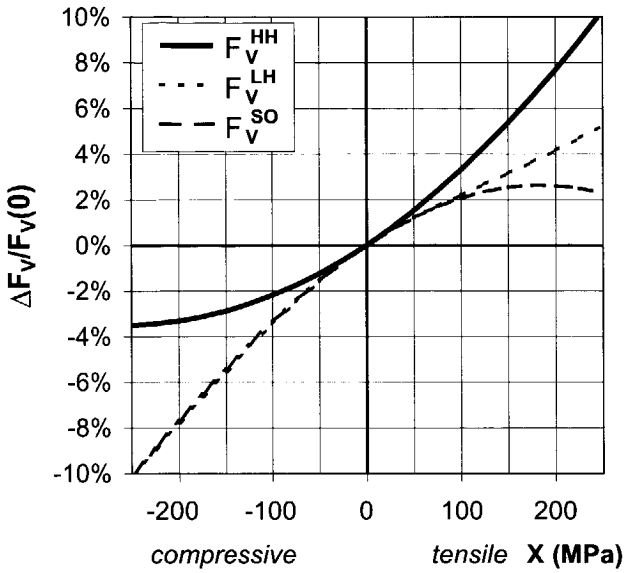

Figure 2.17: Relative changes in the available densities of states of the three valence bands, as a function of normal uniaxial stress. The band parameters for the calculations were taken from Hensel [41], and the deformation potentials from Friedel [14]. $T = 296$ K.

with the aid of Equation 2.62. This can be understood from the expression of p^u in Equation 2.59 where the band edge is not explicitly present. Apparently, the peaks arise from the separation of this equation into a shape- and an edge-dependent part. For this reason the effective density of states concept obscures the understanding of the influence of stress on the hole concentration.

Changes through variations in the available densities of states

Quantum-mechanical effects such as the peaks in the effective DOS are avoided by using the concept of the *available* density of states F_V, defined previously in this section. Its numerical evaluation yields three interesting results.

The first interesting result is the ratio between the different F_V^u, which is equal to the ratio of the hole concentrations provided by the bands. The total F_V equals 2.57×10^{19} cm^{-3} in the stress-free case, at 296 K and with the band parameters of Hensel [41]. The heavy-hole band contributes 88% to this number, the light-hole band 11%, and the split-off band 1%.

The second interesting result is the behaviour of the different F_V^u under stress. This is plotted in Figure 2.17 for the case of normal uniaxial stress and in Figure 2.18 for pure shear stress. It can be seen that the plots do not show any sharp peaks, in contrast to the plots of the effective densities of states. It

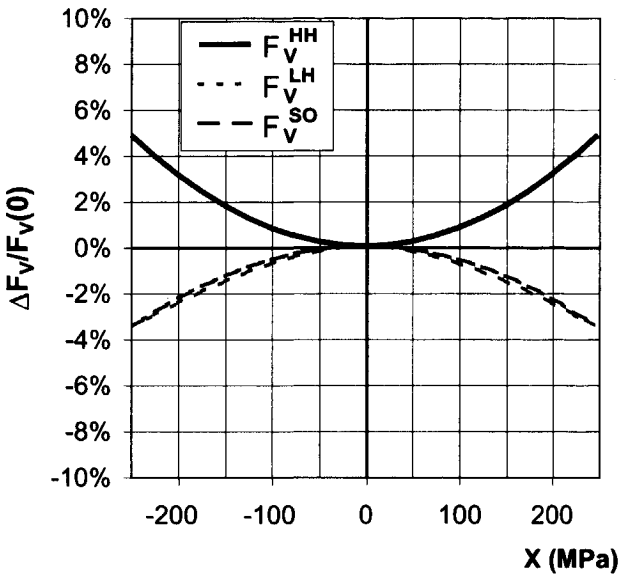

Figure 2.18: Relative changes in the available densities of states of the three valence bands, as a function of pure shear stress. The band parameters for the calculations were taken from Hensel [41], and the deformation potentials from Friedel [14]. $T = 296$ K.

can also be seen that the available densities of states change up to 10% in the stress range considered. For normal stress they increase with increasing stress, whereas for pure shear they vary almost parabolically with the extrema in the origin. Finally, it can be seen that the changes are unequal for the different bands. This modifies the ratio between the charge carrier concentrations of the bands. More precisely stated: stress always increases the percentage of heavy holes in the total concentration.

The third interesting result of the numerical evaluation is the stress-induced change in the *total* available density of states F_{V}. This change represents the shifting of the Fermi level ΔE_{Fp} due to a stress X, as can be derived from Equation 2.64 and 2.66:

$$\frac{F_{\mathrm{V}}(X)}{F_{\mathrm{V}}(0)} = \exp\left(\frac{E_{Fp}(X) - E_{Fp}(0)}{k_B T}\right) \approx \frac{\Delta E_{Fp}(X)}{k_B T} \tag{2.76}$$

The latter equality is valid if the shift is small compared to the thermal energy, which is generally the case in this work. The changes in F_{V} are very near to the changes in $F_{\mathrm{V}}^{\mathrm{HH}}$ depicted in Figure 2.17 and 2.18 because the heavy-hole band contains most of the holes.

2.2.3 Electrical conductivity

Electrical conductivity is the transport property on which the characteristics of resistors and transistors depend most directly. It is also the stress dependence of the conductivity which leads to the piezoresistive and piezojunction effects.

The conductivity relates the electrical current density at a point to its driving forces. Usually, this relation is given by the drift-diffusion equations. However, since these equations rely on the assumptions of isotropy and homogeneity, they are invalid in stressed silicon. Their derivation should therefore be reconsidered, which is done in Appendix C. It yields equations for the electron and hole current density in terms of the band parameters, which can be expressed in two forms. In the most general form, the currents are expressed as functions of the gradients in the quasi Fermi levels. In the other form, the currents are given by the drif-diffusion equations extended with some new terms.

General expressions

In its first form the electron current density vector J_i^m due to the conduction band m is expressed as a function of the gradient in the electron quasi Fermi level E_{Fn} with respect to the position $\mathbf{x} = x_j$. Similarly, the hole current density vector J_i^u due to the valence band u is a function of the gradient in the hole quasi Fermi level E_{Fp}. It was derived in Appendix C that:

$$J_i^m(\mathbf{x}) = \sigma_{ij}^m(\mathbf{x}) \frac{\partial E_{Fn}}{q\partial x_j}(\mathbf{x}) \ , \qquad J_i^u(\mathbf{x}) = \sigma_{ij}^u(\mathbf{x}) \frac{\partial E_{Fp}}{q\partial x_j}(\mathbf{x}) \tag{2.77}$$

In those equations q is the unit charge, σ_{ij}^m is the contribution of the conduction band m to the total conductivity tensor of the material, whereas σ_{ij}^u is the contribution of the valence band u to the total conductivity. The subscript notation applied here is explained in Appendix A. The *band conductivities* are related to the energy bands E_C^m and E_V^u by the integrals

$$\sigma_{ij}^m(\mathbf{x}) = -\frac{q^2}{\hbar^2} \int_C \frac{d\mathbf{k}}{4\pi^3} \tau_m(\mathbf{k},\mathbf{x}) \frac{\partial E_C^m(\mathbf{k},\mathbf{x})}{\partial k_i} \frac{\partial E_C^m(\mathbf{k},\mathbf{x})}{\partial k_j} \frac{\partial f_n(E_C^m, E_{Fn})}{\partial E_C^m} \tag{2.78}$$

$$\sigma_{ij}^u(\mathbf{x}) = -\frac{q^2}{\hbar^2} \int_C \frac{d\mathbf{k}}{4\pi^3} \tau_u(\mathbf{k},\mathbf{x}) \frac{\partial E_V^u(\mathbf{k},\mathbf{x})}{\partial k_i} \frac{\partial E_V^u(\mathbf{k},\mathbf{x})}{\partial k_j} \frac{\partial f_p(E_V^u, E_{Fp})}{\partial E_V^u} \tag{2.79}$$

in which τ_m and τ_u are the electron and hole relaxation times of band m and u, respectively, and f_n and f_p are the Fermi-Dirac distribution functions in nonequilibrium.

The relations of Equation 2.77 are especially useful for analytical determination of device characteristics using the regional approach (see Section 2.2.1 and Appendix D). Together with the device geometry they form a boundary value problem [46, 47]. The solution of this problem includes the difference between the quasi Fermi levels of the device terminals, which is nothing more than the applied voltage.

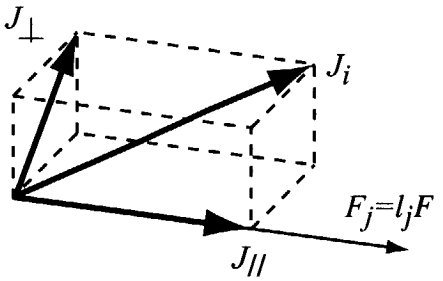

Figure 2.19: Current vector J_i as the result of a gradient in the quasi-Fermi level F_j in the case of a mobility tensor with off-diagonal components. J_i can be split into the vectors $J_{//}$ and $J_\perp$ which are parallel and perpendicular to F_j, respectively.

Symmetry properties of the conductivity tensors

It is worth noting that the conductivities in the previous section are represented as *tensors*. This means for example that the direction of the current density need not necessarily be parallel to the gradient in the quasi Fermi level, see Figure 2.19. It is a result of a fundamental postulate of crystal physics, known as *Neumann's principle* [48]:

> *The symmetry elements of any physical property of a crystal must include the symmetry elements of the point group of the crystal.*

In the absence of stress, silicon is a cubic crystal. It has the symmetry elements given in Appendix B.4, which means that its properties are isotropic and can be represented by a single number. In the presence of stress, however, the cubic symmetry is usually perturbed. The property tensors therefore have no longer equal numbers on their diagonal, and off-diagonal elements appear. In that case the full tensor notation should be retained.

Expressions for nondegenerate material; mobility

In analogy to the charge carrier concentration, the mathematical analysis of the conductivity is considerably simplified when the material is nondegenerate. In that case the Maxwell-Boltzmannn distribution function may be applied and Equation 2.78 and 2.79 reduce to:

$$\sigma_{ij}^m = \exp\left(\frac{E_{Fn}(\mathbf{x})}{k_B T}\right) G_{ij}^m(\mathbf{x}) \tag{2.80}$$

$$G_{ij}^m(\mathbf{x}) \equiv \frac{q^2}{\hbar^2 k_B T} \int_C \frac{d\mathbf{k}}{4\pi^3} \tau_m \frac{\partial E_C^m}{\partial k_i} \frac{\partial E_C^m}{\partial k_j} \exp\left(\frac{-E_C^m}{k_B T}\right) \tag{2.81}$$

$$\sigma_{ij}^{u} = \exp\left(\frac{-E_{Fp}(\mathbf{x})}{k_B T}\right) \tag{2.82}$$

$$G_{ij}^{u}(\mathbf{x}) \equiv \frac{q^2}{h^2 k_B T}\int_C \frac{d\mathbf{k}}{4\pi^3}\tau_u \frac{\partial E_V^u}{\partial k_i}\frac{\partial E_V^u}{\partial k_j}\exp\left(\frac{E_V^u}{k_B T}\right) \tag{2.83}$$

in which G_{ij}^{m} and G_{ij}^{u} are *defined* as the parts of the conductivities which are independent of the quasi Fermi levels.

For semiconductors it is customary to split the conductivity into a charge carrier concentration and a *mobility* (see Appendix C.2.3). The mobility represents the scattering per charge carrier. For a conduction band m and a valence band u the respective mobilities μ_{ij}^{m} and μ_{ij}^{u} are defined as:

$$\mu_{ij}^{m}(\mathbf{x}) \equiv \frac{\sigma_{ij}^{m}(\mathbf{x})}{q n^{m}(\mathbf{x})}\,, \qquad \mu_{ij}^{u}(\mathbf{x}) \equiv \frac{\sigma_{ij}^{u}(\mathbf{x})}{q p^{u}(\mathbf{x})} \tag{2.84}$$

In a semiconductor the mobility is a much more constant parameter than the conductivity because the concentration may vary considerably. In nondegenerate material the mobility is even the part of the conduction which is independent from the quasi Fermi level, which can be concluded from combining Equation C.41 with Equation 2.81 and 2.65:

$$\mu_{ij}^{m} = \frac{1}{q}\frac{G_{ij}^{m}}{F_{\mathrm{C}}^{m}}\,, \qquad \mu_{ij}^{u} = \frac{1}{q}\frac{G_{ij}^{u}}{F_{\mathrm{V}}^{u}} \tag{2.85}$$

It is useful to also define the *'average' mobilities* μ_{ij}^{n} and μ_{ij}^{p} of the electrons and holes, respectively. These quantities are generally referred to in literature as 'the' mobility of a charge carrier in a material. They are defined by the weighted average of the separate band mobilities:

$$\mu_{ij}^{\mathrm{n}}(\mathbf{x}) \equiv \frac{\sigma_{ij}^{\mathrm{n}}(\mathbf{x})}{q n(\mathbf{x})} = \frac{\sum_{m=1}^{6} n^{m}\mu_{ij}^{m}}{\sum_{m=1}^{6} n^{m}} \tag{2.86}$$

$$\mu_{ij}^{\mathrm{p}}(\mathbf{x}) \equiv \frac{\sigma_{ij}^{\mathrm{p}}(\mathbf{x})}{q p(\mathbf{x})} = \frac{\sum_{u=1}^{3} p^{u}\mu_{ij}^{u}}{\sum_{u=1}^{3} p^{u}} \tag{2.87}$$

Stress-induced changes through the electron mobility

The conductivity of electrons is influenced by mechanical stress, which can be connected to the band edge shifts and band deformations. By definition, the conductivity in Equation C.41 depend on the electron concentration n^{m} and the mobility μ_{ij}^{m} of the considered band. The stress influence on the concentrations n^{m} was already discussed in Section 2.2.2. It was shown to act mainly through the shifts in the band edges.

If the *concentration* in a band m increases, it also increases the weight of the corresponding mobility within the *average mobility* of Equation 2.86. If the electrons in this band are 'light' for a particular field direction, their corresponding mobility is high, and the average mobility will increase.

For a separate conduction band m the *band mobility* also changes under the influence of a stress X_{kl}. It can be calculated from F_C^m and G_{ij}^m, according to Equation 2.85. The former quantity was evaluated in Section 2.2.2, whereas the latter can be found by solving the integral of Equation 2.80. For this purpose use is made of the energy band equations of Equation 2.30 or C.15.

Apart from the energy bands, the integral of Equation 2.80 contains the electron relaxation time τ_m. As shown in Appendix C.2.3, the relaxation time $\tau_m(\mathbf{k}, \mathbf{x})$ is often described by a small power of the kinetic energy s:

$$\tau_m\left(W_n^m(\mathbf{k}, \mathbf{x})\right) = \tau_m^0(\mathbf{x}) \left(\frac{W_n^m(\mathbf{k}, \mathbf{x})}{k_B T}\right)^s \tag{2.88}$$

where $s = -1/2$ to 0 for nondegenerate doping levels. The description should be treated with care because of the underlying assumptions. It firstly assumes that an electron after scattering has a completely arbitrary direction (relaxation time approximation) [2, 49]. This is not probable since the bands are anisotropic. Secondly, it assumes that the probability of intervalley scattering is independent of the energy level [50, 51, 10]. This is not true, and the levels vary with stress. Nevertheless, the description is often quite successful, especially in predicting the first-order piezoresistive coefficients, which will be discussed later [28, 42].

After the substitution of the relaxation time, the integrals of Equation 2.80 can be calculated by combining a diagonalisation and a Herring transformation as in Section 2.2.2 and Appendix C.2.3. This yields the mobility μ_{ij}^1 of a conduction band ellipsoid along the k_1 axis:

$$\mu_{ij}^1 = \frac{4q\tau_n^0}{3\sqrt{\pi}} M_{ij}^1 = \frac{4q\tau_n^0}{3\sqrt{\pi}} \begin{bmatrix} m_l^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 2\alpha e_{23} \\ 0 & 2\alpha e_{23} & m_t^{-1} \end{bmatrix} \tag{2.89}$$

The mobilities μ_{ij}^m of the other bands are calculated in a similar way.

The equation shows that the *diagonal* elements of the mobility tensors are strain-independent. The *off-diagonal* elements, however, are nonzero in the presence of strain, and depend linearly on strain and therefore on stress. With a maximum strain of 0.1% they can become as large as 7% of the element along the longitudinal axis.

Equation 2.86 and 2.89 lead to two conclusions about the influence of stress on the *average* electron mobility. Firstly, the diagonal elements of the average mobility are only changed when the ratio between the *concentrations* in the different bands is changed. This changes the weigth of the contribution of each band, expressed in Equation 2.86. Secondly, the off-diagonal elements of the average mobility is changed because the *mobilities* of the bands are modified by stress. It appears in practice that the influence on the main diagonal elements is about five times larger than the influence on the off-diagonal elements.

The influence of stress on the average mobility tensor elements has been resumed in Figure 2.20. It can be seen that a normal uniaxial stress X_{11} has a

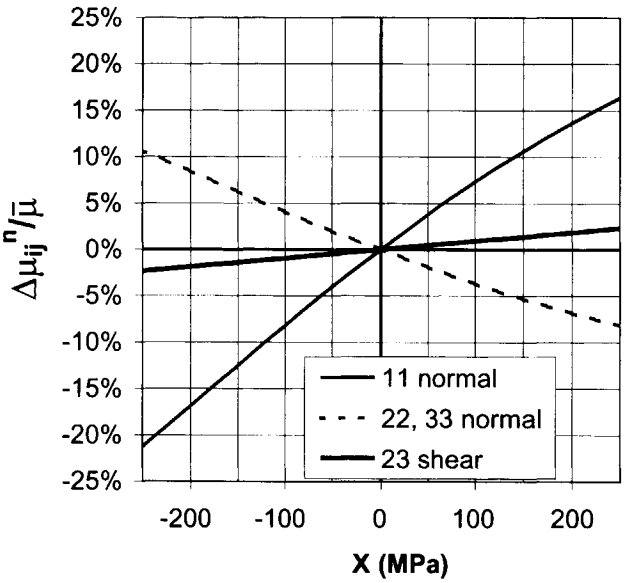

Figure 2.20: Calculated changes in the tensor elements of the average electron mobility as the result of a normal uniaxial stress X_{11} and a pure shear stress X_{23}. Only the nonzero changes are plotted. The changes are normalised with respect to the trace of the zero-stress mobility tensor $\bar{\mu}$. The band parameters were taken from Hensel [41], and the deformation potentials from Friedel [14]. $T = 296$ K.

large influence on the mobility in the stress direction, μ_{11}^{n}, of about 20% on the stress range considered. This influence is approximately linear. The same stress has a smaller and opposite influence on the mobility elements perpendicular to it, μ_{22}^{n} and μ_{33}^{n}. It has no influence on the off-diagonal mobility elements. However, a pure shear stress X_{23} does influence those elements, but only by a few percent. The shear stress does not influence the main diagonal mobility elements.

Stress-induced changes through the hole mobility

The influence of a stress X_{kl} on the hole mobility can be evaluated numerically only, by using Equation 2.87 and 2.85. Rearranging these equations yields:

$$\mu_{ij}^{p}(X_{kl}) = \frac{1}{q}\frac{\sum_{u=1}^{3} G_{ij}^{u}}{\sum_{u=1}^{3} F_{V}^{u}} \tag{2.90}$$

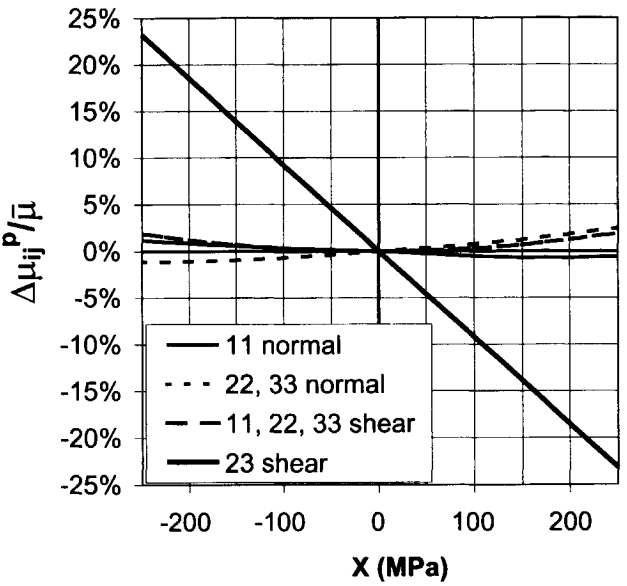

Figure 2.21: Calculated changes in the tensor elements of the average hole mobility as the result of a normal uniaxial stress X_{11} and a pure shear stress X_{23}. Only the nonzero changes are plotted. The changes are normalised with respect to the trace of the zero-stress mobility tensor $\bar{\mu}$. The band parameters were taken from Hensel [41], and the deformation potentials from Friedel [14]. $T = 296$ K.

The quantities G_{ij}^{u} and F_{V}^{u} can be expressed in terms of the energy bands by using Equation 2.83 and 2.64.

$$G_{ij}^{u}\left(X_{kl}\right) = \frac{q}{\hbar^2 k_B T}\int_C \frac{d\mathbf{k}}{4\pi^3}\tau_u \frac{\partial E_V^u}{\partial k_i}\frac{\partial E_V^u}{\partial k_j}\exp\left(\frac{E_V^u}{k_B T}\right) \tag{2.91}$$

$$F_{\mathrm{V}}^{u}\left(X_{kl}\right) = \int_C \frac{d\mathbf{k}}{4\pi^3}\exp\left(\frac{E_V^u}{k_B T}\right) \tag{2.92}$$

The above integrals were numerically solved by using the energy bands of Equation B.77 to B.79 in combination with the matrix elements of Equation 2.48 and 2.49. The derivatives were calculated analytically. The method of solution was the same as in Section 2.2.2. Each solution took approximately ten minutes of computing time.

The results of the calculations are shown in Figure 2.21. They show that a pure shear stress X_{23} has a large influence on the off-diagonal element μ_{23}^{p} of the mobility tensor: 23% on the considered stress range. This influence is approximately linear. The influence on the main diagonal elements is small: only a few percent, and of second order only. A normal uniaxial stress X_{11} also

has only a small influence on the mobility tensor. It changes only the main diagonal mobility elements, with a few percent on the stress range considered. The changes are approximately parabolic.

Generalised drift-diffusion equations

The electron and hole currents are a function of the quasi Fermi levels, but can also be described by the drift-diffusion equations. In the drift-diffusion equations, the gradients in the quasi Fermi level are split up into several driving fields. These fields are the sources of either drift or diffusion of the charge carriers. In Appendix C they are derived for the generalised band edge structure of Section 2.1.7. This yields:

$$J_i^m = \sigma_{ij}^m \left[-\frac{\partial \phi}{\partial x_j} - \frac{1}{q}\frac{\partial \chi^m}{\partial x_j} + \frac{k_B T}{q}\left(\frac{d}{dx_j} \ln n^m - \frac{d}{dx_j} \ln N_C^m \right) \right] \tag{2.93}$$

$$J_i^u = \sigma_{ij}^u \left[-\frac{\partial \phi}{\partial x_j} - \frac{1}{q}\frac{\partial \chi^m}{\partial x_j} - \frac{1}{q}\frac{\partial E_G^{mu}}{\partial x_j} - \frac{k_B T}{q}\left(\frac{d}{dx_j} \ln p^u - \frac{d}{dx_j} \ln N_V^u \right) \right] \tag{2.94}$$

where $-\partial\phi/\partial x_j$ is the gradient in the electrostatic potential and equal to the electrostatic field. The symbol χ^m indicates the electron affinity, E_G^{mu} the bandgap between the conduction band m and the valence band u, and N_C^m and N_V^u are the effective densities of states of those bands, respectively.

There are three important differences between these equations and the conventional ones of Equation 2.54 and 2.55:

- The current densities are now defined *for each band separately.* The total current density is the sum of the contributions of each band, which are generally different.
- The current densities are now a function of a mobility *tensor* instead of a mobility scalar.
- The current densities are no longer determined by the electrostatic field and the concentration gradients only, but also by the gradients in χ^m, E_G^{mu}, N_C^m and N_V^u. This means, for instance, that holes tend to move in the direction of an increasing effective density of states.

Changes in the electron affinity due to mechanical stress have rarely been investigated. Some experiments indicate that they can be neglected compared to the band gap changes [52, 53].

The generalised drift-diffusion equations can be applied in all calculations of device characteristics where the classical drift-diffusion equations are used in the absence of stress. An important application could be found in Poisson solvers. These programs contain a physical model of the device consisting of

the drift-diffusion equations and the Poisson equation. They divide the device into regions with the aid of a mesh. Then they solve the equations with a finite-element method to find all local current densities, potentials, and charge carrier concentrations. A Poisson solver with stress as an input variable has already been developed [54], but does not yet incorporate an accurate transport model.

2.2.4 Auxiliary relations

The behaviour of semiconductor devices is mainly determined by the charge transport equations, but often some auxiliary relations are needed to calculate it completely. These relations include the *pn* product, the continuity equations, and the Poisson equation, and are discussed in Appendix C.3. Each of them contains parameters which are influenced by stress. These influences are discussed in the following.

pn product in thermodynamic equilibrium

The *pn* product in thermodynamic equilibrium,also called the *intrinsic carrier concentration,* is a doping-independent material constant in the case of nondegeneracy. It is often used to calculate the minority charge carrier concentration across a junction [55], and thereby enters the expression of the transistor collector current (see Appendix D). It is highly sensitive to the absolute temperature, a property that is often applied in temperature sensors [56, 57]. In addition, it can provide useful information on the band structure of the material when it is determined experimentally [31].

As argued in Appendix C.3.1, the *pn* product in stressed material can be defined in three ways because of the inequality of the bands. Firstly, it can apply to only one conduction band m and one valence band u. With the aid of Equation 2.61, 2.62, 2.63, and 2.64 it can be written as:

$$(n_{i0}^{mu})^2 = n_0^m p_0^u = N_C^m N_V^u \exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_B T}\right) = F_C^m F_V^u \tag{2.95}$$

This expression strongly depends on mechanical stress through the band edges and the densities of states, as was shown in the preceding sections. Secondly, the *pn* product can be defined for one conducion band with respect to all valence bands, or *vice versa*:

$$(n_{i0}^m)^2 = n_0^m \sum_{u=1}^{3} p_0^u = N_C^m \sum_{u=1}^{3} N_V^u \exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_B T}\right) = F_C^m F_V \tag{2.96}$$

$$(n_{i0}^u)^2 = p_0^u \sum_{m=1}^{6} n_0^m = N_V^u \sum_{m=1}^{6} N_C^m \exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_B T}\right) = F_V^u F_C \tag{2.97}$$

These expressions are used in Section 2.4 for the calculation of the piezojunction effect. Thirdly, the *pn* product can be defined for all conduction and valence

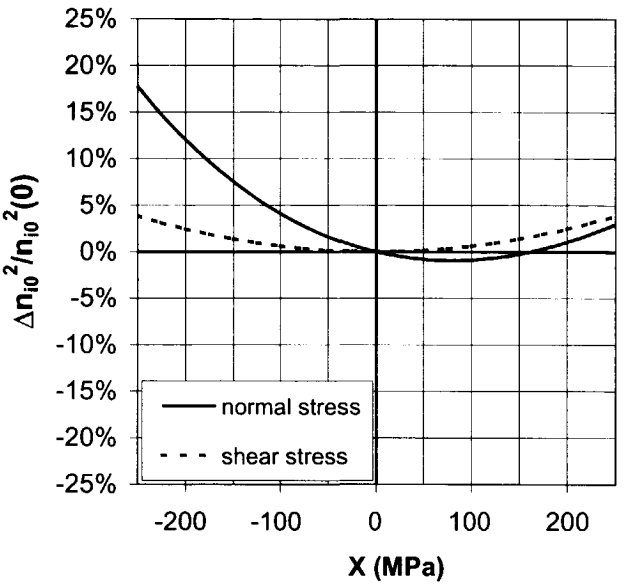

Figure 2.22: Changes in the pn product of silicon, calculated for normal uniaxial stress and for pure shear stress. The band parameters were taken from Hensel [41], and the deformation potentials from Friedel [14]. $T = 296$ K.

bands together, which corresponds to its classical definition:

$$n_{i0}^2 = np = \sum_{u=1}^{3} N_C^m N_V^u \exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_B T}\right) = F_C F_V \tag{2.98}$$

The stress sensitivity of this overall pn product can be calculated with the use of the equations for F_C and F_V given in Section 2.2.2. The changes in F_C are known in analytical form, while those in F_V need to be calculated numerically. This has been done for a normal uniaxial stress X_{11} and a pure shear stress X_{23}, and the results are displayed in Figure 2.22.

The figure has several interesting features. Firstly, it shows that the overall pn product is highly sensitive to stress, as could be expected from its dependence on the bandgaps and band curvatures. This sensitivity is in the same order of magnitude as that of the average mobilities shown in Figure 2.20 and 2.21 , for the considered stress range. It is much larger for normal stress than for shear stress. In addition, it results in curves which are approximately parabolic in this range. The curve for shear stress is symmetric around the origin and corresponds to a first-order sensitivity which is zero.

Lifetime

The lifetime of minority charge carriers appears in the continuity equations of charge transport, as described in Appendix C.3.2. In transistors it determines the base current and the collector current in very low forward bias.

The concept of lifetime can be quite complicated in practice. Its value depends on the operation of the device in which it is measured, and on the measurement technique itself [58]. In addition, there is a difference between bulk and surface lifetime. There is also a difference between the lifetimes of generation and recombination processes. Finally, recombination occurs through different mechanisms, of which the Shockley-Read-Hall and the Auger mechanisms are the most important in silicon.

It is therefore not surprising that lifetime changes through mechanical stress are not very well known. Surface recombination is generally supposed to be independent of stress [59, 60]. Bulk recombination in forward-biased diodes is supposed to be insensitive to hydrostatic pressure, at least compared to the stress sensitivity of the conductivity [61].

In some experiments, however, bulk recombination has been shown to change considerably under uniaxial stress [59, 60, 62]. These changes amount to a few percent for compressive stresses of 200 MPa, and vary between -85 and 300% for 1 GPa. They depend on the doping type and the stress orientation. Partially, they can be explained from the Shockley-Read-Hall model by changes in the energy bands and shifts of the levels of the recombination centres. The remaining changes are attributed to the capture cross-sections of the centres.

However, the material in the cited experiments is quite different from that used in transistors. Its doping is much lower, meaning that Auger recombination can be neglected. Even more important is that the concentration of recombination centres is probably much higher than the majority concentration. In that case the recombination strongly depends on the shifts of the energy levels, which can normally be neglected [63].

In theory, it is quite probable that lifetimes depend on stress even at low concentration of recombination centres. In Appendix C.3.2 it is shown that Shockley-Read-Hall recombination at low-injection conditions depends on the thermal velocity of electrons. This velocity depends on the energy bands, which change with stress. In addition, the capture cross sections and the Auger coefficients can be derived from the quantummechanical behaviour of electrons. They depend on the transition probabilities between electron wave functions, which were shown to change under stress in Section 2.1.

Dielectric constant

The dielectric constant of silicon appears in the Poisson equation, relating the space charge to the electric field (see Appendix C.3.3). It is important for the calculation of junction capacitances, the drain current of MOS transistors, and also for the refraction index of optoelectronic devices.

The 'constant' indeed hardly varies with the frequency of the electric field from zero up to the near infrared [65, 64]. In this spectrum silicon is transparent. At higher frequencies, photons of the field have enough energy to create electron-hole pairs. This increases the absorption very rapidly, and therefore the imaginary part of the dielectric constant. Through the Kramers-Kronig relations it also changes the real part of the constant, equivalent with the refractive index [68].

Stress influences the dielectric constant of silicon in two respects. Firstly, it causes anisotropy, making the silicon birefringent. The reason for this is that stress breaks the crystal symmetry. In that case Neumann's principle says that the constant is no longer a scalar but a second-order tensor ϵ_{ij} (see Section 2.2.3).

Secondly, stress changes the coefficients of ϵ_{ij}. This is called the photoelastic, elastooptic, or piezooptic effect [68]. It has been modelled extensively, especially for optical frequencies [64, 66, 67]. At those frequencies the effects are quite large. For lower frequencies, however, the changes are much smaller: in the order of 10^{-12} Pa^{-1}. As a result, they can be neglected in most electrical devices since they are a factor 10^3 lower than the changes in the conductivity.

2.3 Macroscopic changes in charge transport

In the previous section the current densities were expressed in terms of charge carrier concentrations and mobilities of separate energy bands. From a theoretical point of view this is convenient, but in experimental situations it is preferable to use a phenomenological description. Such a description relates the total current density to observable, macroscopic quantities such as the applied voltage and the stress. It contains parameters which are material properties, such as the conductivity, the piezoresistive coefficients, and the piezojunction coefficients.

This section starts with the phenomenological definition of conductivity and explains how it is related to the microscopic definition. Then it describes the concept of piezoconductance. This concept is related to the piezoresistive and piezojunction effects. The symmetry of these effects is discussed, yielding the tensor elements with which they are described. Finally, these elements are calculated from the microscopic definition. Some of them are the piezoresistive coefficients and are well-known from literature. Their calculation constitutes a check of the models presented.

2.3.1 Phenomenological definition of conductivity

Ohm's law

Macroscopically, the conductivity of a material is defined by Ohm's law, describing the electric current generated by applying a voltage. More precisely, it describes the *incremental* change in the current due to an infinitesimal increment of the voltage *difference*. The increments, however, are generally consid-

ered equal to the variables themselves because the relation is linear up to high electric fields (in silicon some kV/cm [69]). To avoid the effects of the material geometry, Ohm's law is often written in its *local form*, relating the current density as a result of an infinitesimal voltage difference per unit length:

$$J_i(\mathbf{x}) = \sigma_{ij}(\mathbf{x}) \frac{\partial V}{\partial x_j}(\mathbf{x}) \tag{2.99}$$

Here, J_i is the current density vector, σ_{ij} the conductivity tensor, and $\partial V/\partial x_j$ the voltage gradient. This gradient need not necessarily be parallel to the current.

Coupling with the quasi Fermi levels and the energy bands

Ohm's law can be coupled with the microscopic conductivity by using *the definition of the voltage*. Originally, the voltage is defined as the difference in the electrochemical potentials, or quasi Fermi levels, of the charge carriers involved. These charge carriers consist of electrons and holes. It is therefore possible to write Ohm's law as:

$$J_i = J_i^{\mathrm{n}} + J_i^{\mathrm{p}} = \sigma_{ij}^{\mathrm{n}} \frac{\partial E_{Fn}}{q \partial x_j} + \sigma_{ij}^{\mathrm{p}} \frac{\partial E_{Fp}}{q \partial x_j} \tag{2.100}$$

This can be split up further into the contributions of each conduction band m and valence band u:

$$\sigma_{ij}^{\mathrm{n}} = \sum_{m=1}^{6} \sigma_{ij}^{m}, \qquad \sigma_{ij}^{\mathrm{p}} = \sum_{u=1}^{3} \sigma_{ij}^{u} \tag{2.101}$$

which are worked out in microscopic terms in Section 2.2.3.

It is stressed here that the conductivity is *not* defined with respect to the electrostatic field, as done in many textbooks. It is, on the contrary, defined with respect to the Fermi level gradient. As outlined in Section 2.2.3, this gradient contains the electrostatic field *and* the other fields causing a current, such as the gradient in the electron concentration.

Coupling with the mobility and the *pn* product

It was seen in the previous section that the conductivities in the microscopic definitions depend on the quasi Fermi levels or the concentration of the free charge carriers. The hole conductivity, for instance, could be written for nondegenerate material as:

$$\sigma_{ij}^{\mathrm{p}} = qp\mu_{ij}^{\mathrm{p}} = qF_{\mathrm{V}} \exp\left(\frac{-E_{Fp}}{k_B T}\right) \mu_{ij}^{\mathrm{p}} \tag{2.102}$$

where F_{V} was defined in Equation 2.66 as the total valence band available density of states and μ_{ij}^{p} the average hole mobility. Either the concentration

or the quasi Fermi level E_{Fp} should therefore be found in order to make a link between the macroscopic and the microscopic descriptions.

In practice, the charge carrier concentration is often fixed by the doping concentration. The reason is that the material is generally *extrinsic*. This means that the concentration of the doping atoms is much higher than the pn product of charge carriers, whereas the temperature is high enough to ionise all the atoms. In such a material, a p-type doping concentration N_A is equal to the hole concentration p in thermodynamic equilibrium or at low-injection conditions:

$$N_A = p = \exp\left(\frac{-E_{Fp}}{k_B T}\right) F_{\mathrm{V}} \tag{2.103}$$

In this case, where the holes are the *majority* charge carriers, the hole conductivity can be found by combining Equation 2.102 and 2.103:

$$\sigma_{ij}^{\mathrm{p,maj}} = qN_A\mu_{ij}^{\mathrm{p}} \tag{2.104}$$

Things change, however, when the holes are the *minority* charge carriers. In that case the doping concentration N_D does not fix the hole concentration, but only the electron concentration n and the electron quasi Fermi level E_{Fn}:

$$N_D = n = \exp\left(\frac{E_{Fn}}{k_B T}\right) F_{\mathrm{C}} \tag{2.105}$$

Fortunately, the quasi Fermi levels of electrons and holes are equal at thermodynamic equilibrium: $E_{Fn} = E_{Fp}$. For this special case, Equation 2.105 can be combined with Equation 2.102 to yield the *minority hole conductivity:*

$$\sigma_{ij}^{\mathrm{p,min}} = \frac{q}{N_D} F_{\mathrm{C}} F_{\mathrm{V}} \mu_{ij}^{\mathrm{p}} = \frac{q}{N_D} n_{i0}^2 \mu_{ij}^{\mathrm{p}} \tag{2.106}$$

In this equation, n_{i0}^2 is the classical pn product defined in Equation 2.98. The minority conductivity now entirely consists of parameters which can be evaluated in the presence of stress, as done in the previous section.

The above equations can be generalised to the conductivity of any type of charge carrier, n or p. They apply to a specific position $\mathbf{x}$. The result is that the *majority conductivity* can be written as:

$$\sigma_{ij}^{\mathrm{maj}}(\mathbf{x}) = qN_{\mathrm{maj}}(\mathbf{x})\,\mu_{ij}^{\mathrm{maj}}(\mathbf{x}) \tag{2.107}$$

where N_{maj} is the majority doping concentration. The *minority conductivity*, on the other hand, is given by:

$$\sigma_{ij}^{\mathrm{min}}(\mathbf{x}) = \frac{q}{N_{\mathrm{maj}}(\mathbf{x})} n_{i0}^2(\mathbf{x})\,\mu_{ij}^{\mathrm{min}}(\mathbf{x}) \tag{2.108}$$

These equations will be useful in the calculation of the piezoresistive and piezojunction effects.

2.3.2 Piezoconductance

Necessity of the concept

In the previous two equations the conductivities were modelled as a function of the doping concentration, the mobility, and the *pn* product. The latter two quantities directly depend on the shapes of the energy bands and can be calculated for material with and without mechanical stress.

The model, however, is not very practically manageable when the response of various stresses needs to be calculated. This is firstly the result of the *pn* product and the mobility of the valence bands which need to be calculated numerically for each stress value and orientation. As a result, it seriously slows down iterative processes such as the optimisation of the stress sensitivity or the identification of model parameters from measurements.

The model is also impractical because it is difficult to verify with experimental data presented in this work. These data are only available for linear and quadratic parts of the curves which are a function of stress. As a result, it is hard to identify the band parameters and to validate the model structure, which consists of more complicated functions than linear and quadratic polynomials.

The problems are usually overcome by expanding the stress dependence of the conductivity into a MacLaurin series. This is suggested by the measurements, but also by the theory in the preceeding section. It was shown there for electrons that the conductivity of a band depends on the mobility, which depends on stress through some linear and quadratic terms. In addition, it depends on the band population, which depends on the bandgaps and is thereby exponential in stress.

The *series expansion* of the conductivity is much easier to evaluate than its original equation. Once the first- and second-order derivatives are known, the influence of a new stress tensor can be calculated in a fraction of a second. The calculation is simplified even further by using the *a priori* knowledge about the material. This drastically reduces the number of derivatives required.

The use of a series expansion can overcome the lack of experiments at high stress levels by neglecting higher-order terms. The lower-order terms alone are enough to describe all situations at the stress range considered. In addition, these terms can be derived from the original microscopic model of the conductivity by differentiation.

The stress-derivatives of the conductivity are called the *piezoconductive tensors* of the first, second, etc. order. They are closely related to the well-known piezoresistive coefficients. It will appear, however, that they are also related to the piezojunction coefficiens describing the stress-induced changes in the behaviour of bipolar transistors.

Mathematical formulation

The piezoconductance can be formulated mathematically by defining the conductivity in the macroscopic way of the previous section. Only one charge carrier

type is considered for simplicity, and the gradient in the quasi Fermi level is abbreviated by the field F_j. An increment in this field, dF_j, yields the following increment in the current density dJ_i:

$$dJ_i = \sigma_{ij} dF_j \tag{2.109}$$

The conductivity in this equation depends on the actual field F_j and the stress tensor X_{kl}. It is fixed to the reference state at $(F_j, X_{kl}) = (0, 0)$ by expanding dJ_i into a MacLaurin series [70, 71]. Up to the third order this yields:

$$\begin{aligned} dJ_i\,(F_j, X_{kl}) = {} & \left.\frac{\partial J_i}{\partial F_j}\right|_0 dF_j + \left.\frac{\partial J_i}{\partial X_{kl}}\right|_0 dX_{kl} + \\ \frac{1}{2!}\Bigg(& \left.\frac{\partial^2 J_i}{\partial F_j \partial F_m}\right|_0 dF_j dF_m + \left.\frac{\partial^2 J_i}{\partial X_{kl} \partial X_{mn}}\right|_0 dX_{kl} dX_{mn} + \left.\frac{2 \cdot \partial^2 J_i}{\partial F_j \partial X_{kl}}\right|_0 dF_j dX_{kl}\Bigg) \\ + \frac{1}{3!}\Bigg(& \left.\frac{\partial^3 J_i}{\partial F_j \partial F_m \partial F_n}\right|_0 dF_j dF_m \partial F_n + \left.\frac{\partial^3 J_i}{\partial X_{kl} \partial X_{mn} \partial X_{op}}\right|_0 dX_{kl} dX_{mn} dX_{op} \\ & + 3 \left.\frac{\partial^3 J_i}{\partial F_j \partial X_{kl} \partial X_{mn}}\right|_0 dF_j dX_{kl} dX_{mn} + 3 \left.\frac{\partial^2 J_i}{\partial F_j \partial F_m \partial X_{kl}}\right|_0 dF_j dF_m dX_{kl}\Bigg) \end{aligned}$$

For crystals with a centre of symmetry such as silicon all tensors of odd rank are zero [48, 72]. In addition, the applied field is low enough to neglect the terms nonlinear in F_j. As a result, the MacLaurin series reduces to:

$$dJ_i = \left.\frac{\partial J_i}{\partial F_j}\right|_0 dF_j + \left.\frac{\partial^2 J_i}{\partial F_j \partial X_{kl}}\right|_0 dF_j dX_{kl} + \frac{1}{2} \left.\frac{\partial^3 J_i}{\partial F_j \partial X_{kl} \partial X_{mn}}\right|_0 dF_j dX_{kl} dX_{mn}$$

The partial derivatives in this expression are generally *defined* as:

$$\left.\frac{\partial J_i}{\partial F_j}\right|_0 \equiv \sigma_{ij}^0, \text{ first-order stress-free conductivity} \tag{2.110}$$

$$\left.\frac{\partial^2 J_i}{\partial F_j \partial X_{kl}}\right|_0 \equiv \Pi_{ijkl}, \text{ first-order piezoconductive tensor} \tag{2.111}$$

$$\frac{1}{2} \left.\frac{\partial^3 J_i}{\partial F_j \partial X_{kl} \partial X_{mn}}\right|_0 \equiv \Pi_{ijklmn}, \text{ second-order piezoconductive tensor} \tag{2.112}$$

With zero as the reference, the increments are equal to the variables themselves. Therefore the previous expressions can be combined to yield the low-field conductivity expanded to the second order in stress:

$$\sigma_{ij} = \sigma_{ij}^0 + \Pi_{ijkl} X_{kl} + \Pi_{ijklmn} X_{kl} X_{mn} \tag{2.113}$$

By a further extension of the series it is possible to define higher-order tensors [50, 73, 74, 75]. The above expansion, however, is sufficiently accurate for most practical purposes on the stress range considered (−200 to 200 MPa). In addition, the determination of higher-order tensors from either theory or experiments is complicated and vulnerable to variations in the input data.

2.3.3 Piezoresistive and piezojunction effects

General expressions

Starting from the concept of piezoconductance, both the piezoresistive and piezojunction effects can now be defined. For this purpose it is firstly necessary to define resistance as the inverse of conductance. This means that the resistivity tensor ρ_{ij} satisfies the relation:

$$\sigma_{ij}\rho_{jk} = \delta_{ik} \tag{2.114}$$

where δ_{ik} is the second-order unity tensor. Secondly, it should be noted that resistance can be defined for majority as well as for minority charge carriers in a semiconductor device.

The piezoresistive effect can now be defined as the sensitivity of the *majority* resistance to mechanical stress. This definition is broader than usual [50]. It covers, however, the same phenomenon, namely the stress-induced change in the value of a resistor, which will be shown in the next section.

The piezojunction effect, on the other hand, can be defined as the the sensitivity of the *minority* resistance to stress. This definition is new. It will be shown in the next section that it covers the stress-induced change in the saturation current of a bipolar transistor, with which it is usually associated.

Just as the piezoconductance, the piezoresistive and piezojunction effects can be described by MacLaurin series in stress. Up to the second order this yields:

$$\rho_{ij}^{\mathrm{maj}} = \rho_{ij}^{0,\mathrm{maj}} + \pi_{ijkl}X_{kl} + \pi_{ijklmn}X_{kl}X_{mn} \tag{2.115}$$

$$\rho_{ij}^{\mathrm{min}} = \rho_{ij}^{0,\mathrm{min}} + \zeta_{ijkl}X_{kl} + \zeta_{ijklmn}X_{kl}X_{mn} \tag{2.116}$$

in which π_{ijkl} and π_{ijklmn} are the piezoresistive tensors of first and second order, respectively. The elements of those tensors are called the piezoresistive coefficients. Likewise, the parameters ζ_{ijkl} and ζ_{ijklmn} can be called the piezojunction tensors. Their elements are called the piezojunction coefficients.

Equivalent effects

Piezoconductance is equivalent to the concepts of elastoconductance and elastoresistance, which are also encountered in literature [51, 50]. The prefix 'elasto-' stands for the influence of mechanical strain, whereas 'piezo-' stands for that of stress. In silicon, however, the strain e_{kl} and stress X_{ij} are related to each other in a unique way. This relation is highly linear up to the fracture stress:

$$e_{ij} = S_{ijkl}X_{kl} \tag{2.117}$$

where S_{ijkl} is the compliance tensor (see Appendix A).

Reduced-index notation

Equation 2.115, 2.116 and 2.117 yield a very large number of terms when fully written down. It is therefore customary to write their tensors in a reduced notation. This reduction takes advantage of all possible symmetries of the stress, the resistivity, the equations, and the material [48, 75, 76, 50].

The first symmetry is that of the stress and the resistivity. For a body which is not accelerated the stress tensor is symmetric with respect to its main diagonal; see Appendix A. This reduces the number of independent tensor components from nine to six. The resistivity tensor is also symmetric, which appears from its microscopic definition in Equation 2.78 and 2.79. This makes it possible to take the tensor subscripts in pairs and convert them to single values. In this procedure the following rule is used:

$$11 \rightarrow 1;\ \ 22 \rightarrow 2;\ 33 \rightarrow 3;\ \ 23\text{-}32 \rightarrow 4;\ \ 13\text{-}31 \rightarrow 5;\ \ 12\text{-}21 \rightarrow 6 \qquad (2.118)$$

although other conversion rules occur in literature [77, 8]. The result of this operation is that a stress tensor X_{kl} reduces to a six-vector X_P, and a resistivity tensor ρ_{ij} reduces to a six-vector ρ_O. Care should be taken with the strain, however, which reduces from e_{ij} to $e_O/2$ if $i \neq j$. This factor two in the shear elements is present for historical reasons and has already led to much confusion.

The second symmetry to be exploited is that products like $X_{kl}X_{mn}$ are equal to $X_{mn}X_{kl}$. The subscripts in an accompanying tensor as π_{ijklmn} can therefore be interchanged as well, and equivalent terms can be taken together. This leads to the following reduction in which distinction is made between main-diagonal and off-diagonal subscrips [28, 70]:

Piezoresistive	*Piezojunction*	*Compliance*	*Conditions*
$\pi_{OP} = \pi_{iijj}/\overline{\rho}$	$\zeta_{OP} = \zeta_{iijj}/\overline{\rho}$	$S_{OP} = S_{iijj}$	$O \wedge P \in \{1,2,3\}$
$\pi_{OP} = 2\pi_{ijkl}/\overline{\rho}$	$\zeta_{OP} = 2\zeta_{ijkl}/\overline{\rho}$	$S_{OP} = 2S_{ijkl}$	$O \vee P \in \{4,5,6\}$
$\pi_{OP} = 2\pi_{ijkl}/\overline{\rho}$	$\zeta_{OP} = 2\zeta_{ijkl}/\overline{\rho}$	$S_{OP} = \mathbf{4}S_{ijkl}$	$O \wedge P \in \{4,5,6\}$
$\pi_{OPQ} = \pi_{ijkkll}/\overline{\rho}$	$\zeta_{OPQ} = \zeta_{ijkkll}/\overline{\rho}$		$P \wedge Q \in \{1,2,3\}$
$\pi_{OPQ} = 2\pi_{ijklmn}/\overline{\rho}$	$\zeta_{OPQ} = 2\zeta_{ijklmn}/\overline{\rho}$		$P \vee Q \in \{4,5,6\}$
$\pi_{OPQ} = 4\pi_{ijklmn}/\overline{\rho}$	$\zeta_{OPQ} = 4\zeta_{ijklmn}/\overline{\rho}$		$P \wedge Q \in \{4,5,6\}$

The tensors in this table are normalised with the mean unstressed resistivity value $\overline{\rho}$, which is defined as the trace of the resistivity matrix of the associated charge carrier: $\overline{\rho} = \mathrm{Tr}(\rho_{ij}^0)/3$. As a result of the reductions, the MacLaurin series of Equation 2.115 and 2.116 can be abbreviated as:

$$\left.\begin{array}{l} \rho_O^{\mathrm{maj}} = \rho_O^{0,\mathrm{maj}} + \pi_{OP}X_P + \pi_{OPQ}X_PX_Q \\ \rho_O^{\mathrm{min}} = \rho_O^{0,\mathrm{min}} + \zeta_{OP}X_P + \zeta_{OPQ}X_PX_Q \end{array}\right\} \quad O,P,Q \in \{1..6\} \qquad (2.119)$$

and the stress-strain relation as $e_O = S_{OP}X_P$.

The third kind of symmetry is that of the atomic lattice of the material [8]. In Section 2.2.3 Neumann's principle was presented, which implies that the

physical properties of monocrystalline silicon must at least have the symmetry elements of the cubic point group of the crystal. In practice this means that the property tensors are as symmetric as the corresponding products of the indices [48]. The element ζ_{2213}, for instance, should transform like the product $x_2^2 x_1 x_3$. This function is even in x_2 and odd in x_1 and x_3.

When some of the symmetries of Appendix B.4 are applied to the piezoresistive and piezojunction tensors, it can be shown that the zero-stress resistivity ρ_{ij}^0 is isotropic. As a result, it can be written as $\overline{\rho}\delta_{ij}$. The same applies of course to the zero-stress conductivity σ_{ij}^0, which reduces to $\overline{\sigma}\delta_{ij}$.

In addition, the cubic symmetry strongly reduces the number of first-order coefficients. It can be shown that among the 36 elements of π_{OP}, ζ_{OP} and S_{OP} only three constants per tensor are independent. They are called:

$\pi_{11} = \pi_{OO}$	$O \in \{1,2,3\}$
$\pi_{12} = \pi_{OP}$	$O \wedge P \in \{1,2,3\}, O \neq P$
$\pi_{44} = \pi_{OO}$	$O \in \{4,5,6\}$
$0 = \pi_{OP}$	otherwise

For ζ_{OP} and S_{OP} the reduction is similar.

Finally, the cubic symmetry elements dramatically limits the number of second-order coefficients. When they are applied to the 216 coefficients of π_{OPQ} and ζ_{OPQ}, it appears that only nine of them are independent. These are called [50]:

$\pi_{111} = \pi_{OOO}$	$O \in \{1,2,3\}$
$\pi_{122} = \pi_{OPP}$	$O \wedge P \in \{1,2,3\}, O \neq P$
$\pi_{112} = \pi_{OOP}$	$O \wedge P \in \{1,2,3\}, O \neq P$
$\pi_{123} = \pi_{OPQ}$	$O \wedge P \in \{1,2,3\}, O \neq P \neq Q$
$\pi_{144} = \pi_{OPP}$	$O \in \{1,2,3\}, P \in \{4,5,6\}, P - O = 3$
$\pi_{166} = \pi_{OPP}$	$O \in \{1,2,3\}, P \in \{4,5,6\}, P - O \neq 3$
$\pi_{414} = \pi_{OPO}$	$O \in \{4,5,6\}, P \in \{1,2,3\}, O - P = 3$
$\pi_{616} = \pi_{OPO}$	$O \in \{4,5,6\}, P \in \{1,2,3\}, O - P \neq 3$
$\pi_{456} = \pi_{OPQ}$	$O \wedge P \in \{4,5,6\}, O \neq P \neq Q$
$0 = \pi_{OPQ}$	otherwise

The coefficients of the second-order piezojunction tensor ζ_{OPQ} reduce in the same way.

With the use of the reduced-index notation it is possible to find convenient relations between the piezoresistive and piezojunction tensors on the one hand, and the piezoconductive tensors on the other. For this purpose the expansions of Equation 2.113, 2.115, and 2.116 are inserted into the definition $\sigma_{ij}\rho_{jk} = \delta_{ik}$. By a comparison of equal powers in stress it appears that:

$$\pi_{OP} = -\Pi_{OP}^{\text{maj}}, \qquad \pi_{OPQ} = -\Pi_{OPQ}^{\text{maj}} + \Pi_{OP}^{\text{maj}}\Pi_{OQ}^{\text{maj}} \tag{2.120}$$

$$\zeta_{OP} = -\Pi_{OP}^{\text{min}}, \qquad \zeta_{OPQ} = -\Pi_{OPQ}^{\text{min}} + \Pi_{OP}^{\text{min}}\Pi_{OQ}^{\text{min}} \tag{2.121}$$

where Π_{ijkl} and Π_{ijklmn} have been reduced in the same way as the other tensors.

2.3.4 Evaluation of the first-order coefficients

The independent coefficients of the piezoresistive and piezojunction tensors can be calculated in terms of band parameters. If the models presented are correct, their numerical values must correspond to the measured behaviour presented in literature and in the next chapter. Especially the coefficients of the first-order are interesting. They give a first approach in practice and are also the least sensitive to inaccuracies in the modelling, such as the neglect of anisotropy in the intervalley scattering [51, 50]. The coefficients can be calculated analytically for the electron piezoresistive effect. However, the coefficients for the hole piezoresistive and the piezojunction effect contain integrals that should be solved numerically.

General expressions

The first-order coefficients can be found by relating them to the conductivity, which was already expressed in microscopic quantities. This is done by replacing the piezoconductive coefficients in Equation 2.113 by the expressions of Equation 2.120 and 2.121 and by differentiating the conductivity after stress. In the reduced-index notation this yields:

$$\pi_{OP} = -\Pi_{OP}^{\text{maj}} = -\frac{1}{\overline{\sigma}^{\text{maj}}} \left. \frac{\partial \sigma_O^{\text{maj}}}{\partial X_P} \right|_{X_P=0} \tag{2.122}$$

$$\zeta_{OP} = -\Pi_{OP}^{\text{min}} = -\frac{1}{\overline{\sigma}^{\text{min}}} \left. \frac{\partial \sigma_O^{\text{min}}}{\partial X_P} \right|_{X_P=0} \tag{2.123}$$

It has already been noted that the coefficients are defined at zero field. In the following, it will also be assumed that the material is in thermodynamic equilibrium, to avoid ambiguity. In addition, it will be assumed that the material is nondegenerate. These assumptions often hold in practice, and facilitate the evaluation of the coefficients considerably.

Under these assumptions it is possible to write the piezocoefficients as functions of the mobility and the pn product. Using Equation 2.107 and 2.108, this yields:

$$\pi_{OP} = -\left(\frac{1}{\overline{\mu}_{\text{maj}}} \frac{\partial \mu_O^{\text{maj}}}{\partial X_P} \right)_{X_P=0} \tag{2.124}$$

$$\zeta_{OP} = -\left(\frac{1}{\overline{\mu}_{\text{min}}} \frac{\partial \mu_O^{\text{min}}}{\partial X_P} + \frac{1}{n_{i0}^2} \frac{\partial n_{i0}^2}{\partial X_P} \right)_{X_P=0} \tag{2.125}$$

where $\overline{\mu}$ is the trace of the mobility matrix: $\overline{\mu} = \text{Tr}(\mu_{ij})/3$. The above equations reveal some interesting features of the piezoresistive and piezojunction effects:

- The mathematical formulation is very similar;

Table 2.4: First-order piezoresistive coefficients as calculated from band parameters of different authors, determined from experiments, and found in literature, in 10^{-11} Pa^{-1}.

	This work				Literature			
	[14] [41]	[5]	[15, 16] [41]	exp.	[80] exp.	[50] exp.	[43] calc.	[78] calc.
π_{11}^{n}	−78.1	−96.8	−84.8	–	−102.2	−77	–	–
π_{12}^{n}	39.0	48.4	42.4	–	53.4	39	–	–
π_{44}^{n}	−9.4	−9.4	−9.4	–	−13.6	−14	–	–
π_{11}^{p}	4.2	3.4	3.8	1.8	6.6	−6	5.35	8
π_{12}^{p}	−6.9	−6.9	−7.7	2.5	−1.1	1	−2.68	−2
π_{44}^{p}	91.5	116.6	131.9	118.4	138.1	112	103.20	140

- The piezo*resistive* coefficients are equal to the changes in the majority mobility;
- The piezo*junction* effect depend on the changes in the minority mobility and in the *pn* product. It therefore depends on both the minority and the majority bands;
- The shear piezojunction coefficient is equal to the shear piezoresistive coefficient for both electrons and holes.

The latter feature is the result of the fact that the *pn* product is insensitive to shear stress in the first order, as shown by Figure 2.22. The equations for π_{44} and ζ_{44} thereby reduce to the same expression.

Analytical evaluation

The first-order piezoresistive coefficients for *n*-type material can be evaluated analytically from Equation 2.124, since μ_O^{n} is known in closed form from Section 2.2. This yields the following expressions for the three independent first-order piezoresistive coefficients:

$$\pi_{11}^{n} = \frac{2}{3}\frac{\Xi_u}{k_B T}\left(S_{11} - S_{12}\right)\frac{m_t - m_l}{m_t + 2m_l} \tag{2.126}$$

$$\pi_{12}^{n} = -\frac{1}{3}\frac{\Xi_u}{k_B T}\left(S_{11} - S_{12}\right)\frac{m_t - m_l}{m_t + 2m_l} \tag{2.127}$$

$$\pi_{44}^{n} = -\alpha S_{44}\frac{m_t m_l}{m_t + 2m_l} \tag{2.128}$$

It can be seen that the coefficients are functions of the electron effective masses, the deformation potentials, the compliance coefficients, and the temperature. The expressions agree with those found in literature [51, 10, 22, 23, 79]. When numerical values are inserted they yield coefficients near to those found from experiments (see Table 3.4).

Table 2.5: First-order piezojunction coefficients, in 10^{-11} Pa^{-1}, as calculated from band parameters of different authors and compared with experimental values from this thesis and from Fruett [81]-[83]. Indirectly determined values are indicated between square brackets.

Coeffic.	Calculated			Experimental	
	[14, 41]	[5]	[15, 16, 41]	Chapter 3	Fruett
$\zeta^n_{11} - \zeta^n_{44}$	−44.7	−54.0	−52.2	-41.5 ± 3.1	–
ζ^n_{12}	63.0	81.8	65.6	43.4 ± 1.5	45.5
ζ^n_{11}	−54.1	−63.4	−61.6	$[-28.4 \pm 3.0]$	–
ζ^n_{44}	−9.4	−9.4	−9.4	$[13.1 \pm 4.3]$	–
$\zeta^p_{11} - \zeta^p_{44}$	−63.4	−79.8	−104.9	-89.0 ± 3.3	[−87.4]
ζ^p_{12}	17.0	26.5	15.5	13.8 ± 1.3	14.3
ζ^p_{11}	28.1	36.8	27.0	$[30.8 \pm 2.6]$	8.9
ζ^p_{44}	91.5	116.6	131.9	$[119.8 \pm 4.2]$	96.3

Numerical evaluation

The remaining first-order piezocoefficients can be found by similar derivations followed by a numerical integration over the wave vector space. This was done with similar FORTRAN procedures as in Section 2.2.2, where the effective densities of states were calculated. The derivatives after stress were calculated analytically. This procedure is straightforward in principle, but very elaborate in practice and prone to mistakes. Fortunately, some of the calculations could be verified by a comparison with literature results. Starting from the same input parameters, the same piezoresistive coefficients for holes were obtained as Ohmura in his numerical calculations [42, 43].

For the final calculations different sets of more recent band parameters were used, given in Table B.2 and 2.1. The valence band scattering parameter s was chosen equal to -0.2. This resulted in the sets of first-order coefficients given in Table 3.4 and 2.5. It can be seen that they approach the measurement values, but also that they are very sensitive to the band parameters used. This will be further discussed in Chapter 4.

2.3.5 Interpretation of the piezocoefficients

The first-order piezocoefficients differ in sign and magnitude because they are caused by different energy band effects. These effects were discussed in Section 2.1 and are resumed in Figure 2.23. They include changes in the bandgaps, leading to population changes and redistribution of the charge carriers. They also include changes in the band shape, leading to modifications in the effective mass and the density of states. Population redistribution and effective mass changes lead to modifications in the average mobility of a charge carrier type. Changes in the effective density of states and in the population can modify the total concentration of the carriers through the pn product, but only if this

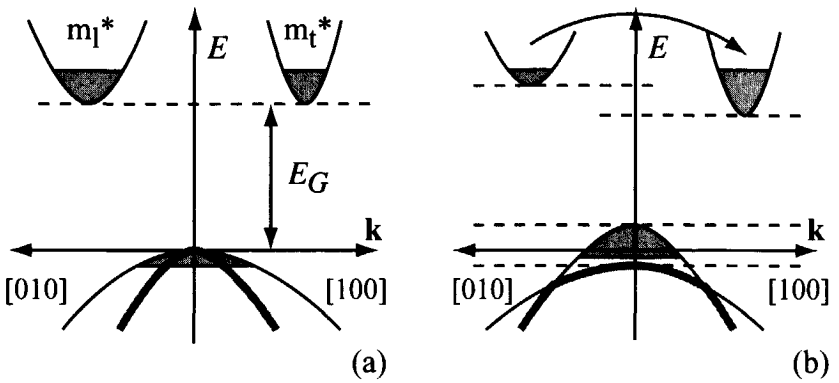

Figure 2.23: Schematic changes in the energy bands of silicon due to stress, and their effects on the charge carrier population.

concentration is not determined by the doping.

Piezoresistive effect for electrons

The piezoresistive effect for electrons is dominated by charge redistribution. This redisribution is the result of an unequal shift of the six conduction band edges. Electrons in higher bands scatter to lower-lying ones, which have another mobility effective mass in that current direction [84]. This changes the average mobility. It was seen before that the band edges only shift due to normal stresses. As a result, the largest piezoresistive coefficients for electrons are the normal coefficients π^{n}_{11} and π^{n}_{12}.

For a small part, the piezoresistive effect for electrons is also caused by band deformation. This deformation is only caused by shear stress and is not accompanied by shifts of the band edges. It leads to a shear piezoresistive coefficient π^{n}_{44} which is smaller than the normal coefficients.

Theoretical evaluations of higher-order piezoresistive coefficients have not been very successful yet. Probably, they are of limited validity because the model neglects the anisotropy of the intervalley scattering [51, 50]. Higher-order evaluations are also complicated by the fact that the effective masses m_l and m_t are independent of stress only to the first order [85, 22, 21]. Finally, they are very sensitive to minor variations in the band parameters on which they are based.

Piezoresistive effect for holes

The piezoresistive effect for holes is dominated by changes in the off-diagonal mobility elements. These changes are caused by the shifts in the valence band edges as well as by the changes in the curvature, which are closely related.

The principal cause of the hole piezoresistance has long remained unclear [84]. This probably resulted from the complicated mathematical description, which requires numerical means to be evaluated [43, 73, 44].

The hole piezoresistance is large for shear stress and small for normal stress. This reaction is opposite to the piezoresistance of electrons. The contrast is sometimes employed to design sensors which are optimally sensitive to both types of stress [77].

Piezojunction effect for electrons

The piezojunction effect for electrons is not only determined by the mobility of the conduction bands, but also by the *pn* product, which involves both the conduction and the valence bands. The hole concentration in the valence bands is set by the doping. The shape of the bands, however, changes with stress. To keep the concentration constant, the Fermi level must change. This Fermi level also determines the number of electrons which are thermally excited over the bandgap and take part in the minority conduction.

The electron piezojunction coefficients are in the same order of magnitude as the electron piezoresistive coefficients. However, the normal coefficient ζ_{11}^{n} is less pronounced than π_{11}^{n}. This means that it is attenuated by a change in the *pn* product, as can be seen in Equation 2.124 and 2.125. The shear coefficient ζ_{44}^{n} is small and near to the shear piezoresistive coefficient π_{44}^{n}. This was already expected and discussed after Equation 2.125, because the sensitivity of the *pn* product to shear stress is zero in the first order. It should be noted that the mentioned experimental value of ζ_{44}^{n} is probably not very exact and less precise than indicated, as will be discussed in Section 3.4.3 and Chapter 4.

Piezojunction effect for holes

The piezojunction effect for holes is determined by the valence band mobility as well as by the *pn* product. The hole piezojunction coefficients for normal stress are not as small as in the piezoresistive case. This is again the result of the attenuating effect of the *pn* product, in which the shifts of the conduction band edges play a role.

However, the coefficient for shear stress ζ_{44}^{p} is large and is very close to the hole piezoresistive coefficient π_{44}^{p}. Just as for the electrons, this was already expected from Equation 2.124 and 2.125. It can also be explained by the weak reaction of the conduction bands to shear stress. This stress does not shift the band edges, meaning that there is no changing band gap from the side of the conduction bands. The result is that the shear coefficient is completely determined by the valence bands, just as in the case of the piezoresistive effect. This yields similar coefficients.

2.4 Modification of device characteristics

In the previous section the piezoresistive and piezojunction effects were defined with respect to the conductivity of the semiconductor. However, these effects only become observable through their influence on a device. Not surprisingly, the piezoresistive effect dominates the stress-dependent behaviour of resistors, whereas the piezojunction effect is responsible for that of bipolar transistors. The resistor behaviour under stress is a well-known topic [84, 86]. It is briefly treated here because it has many parallels with the stress influence on the transistor.

In this section the stress dependence of the device characteristics is described in analytical models. For this purpose, use is made of the regional approach of charge transport, discussed in Section 2.2.1. This approach is usually applied to derive the transistor saturation current in the absence of stress, as shown in Appendix D. The derivation there follows the method of Gummel and Moll-Ross [87, 88]. This method is elegant from both a mathematical and a physical viewpoint. It is also suitable to include the effects of stress. It leads to the theoretical characteristics of some resistors and transistors which are experimentally determined in the following chapter.

2.4.1 General assumptions

The analytical modelling of resistors and transistors requires some general assumptions concerning the geometry, the operating conditions, and the material. They are discussed in Section 2.2.1 and Appendix D. The most important ones are:

- The regional approach. Semiconductor regions are considered to be either quasi-neutral or filled with space charge. In the quasi-neutral regions the space charge is negligible and the gradient in the quasi Fermi level is small.
- Homogeneous stress throughout the active region of the device. This is not strictly necessary, but it simplifies calculations.
- The one-dimensional approach. Gradients in the quasi Fermi level are supposed to be oriented in one direction only. The current component along this direction is constant throughout the device and determines the device behaviour.
- Nondegenerate material in the active region of the device. As a result, some important device characteristics can be calculated analytically.
- Extrinsic material. The doping of a region is large enough for one type of charge carrier being dominant.
- Low-level injection. The concentration of injected charge carriers remains inferior to the majority concentration in the region.

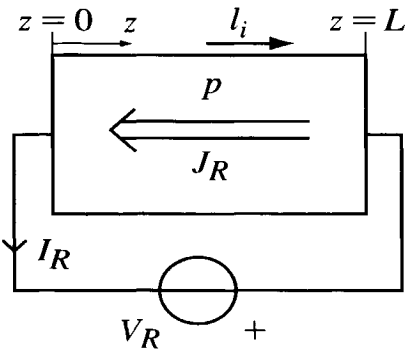

Figure 2.24: Schematic of a p-type silicon resistor, supporting a current I_R with a density J_R as a result of the applied voltage V_R.

- Normal bias conditions. The applied bias voltages and current are such that the device functions in the normal, forward mode.

2.4.2 Resistors

Most semiconductor resistors can be represented as a rectangular slab of material with contacts on both ends, as drawn in Figure 2.24. When a voltage V_R is applied over the contacts, a current I_R starts to flow with a density J_R. The I_R will be written here in terms of the resistor geometry, charge transport parameters, and mechanical stress.

Current-voltage characteristic

In the case of p-type doping the current is almost entirely supported by holes. The current density vector J_i at an arbitrary place in the resistor is therefore given by Equation 2.77:

$$J_i = \sum_u \sigma_{ij}^u \frac{\partial E_{Fp}}{q\partial x_j} = \sigma_{ij}^{\mathrm{p,maj}} \frac{\partial E_{Fp}}{q\partial x_j} \tag{2.129}$$

where the sum runs over all valence bands. The gradient in the quasi Fermi level is now supposed to be along the long side of the resistor, in the direction of the unit vector l_i, defined in the same axes system as σ_{ij}. The current density J_R is the current component along that direction. It can be expressed as the inner product:

$$J_R = l_i J_i = l_i \sigma_{ij}^{\mathrm{p,maj}} \frac{\partial E_{Fp}}{q\partial x_j} = l_i l_j \sigma_{ij}^{\mathrm{p,maj}} \frac{\partial E_{Fp}}{q\partial z} \tag{2.130}$$

where z is the variable of length between the contacts.

This relation is integrated over z up to the length L to establish the relation with the applied voltage:

$$\int_0^L J_R dz = \int_0^L l_i l_j \sigma_{ij}^{\mathrm{p,maj}} \frac{\partial E_{Fp}}{q\partial z} dz \tag{2.131}$$

The current density and the conductivity are supposed to be constant over z and can be placed outside the integral. In addition, the integral can be evaluated to:

$$J_R = \frac{1}{qL} l_i l_j \sigma_{ij}^{\mathrm{maj}} \left(E_{Fp}(L) - E_{Fp}(0) \right) \tag{2.132}$$

The difference in quasi Fermi levels is by definition equal to the voltage multiplied by the unit charge, qV_R. By integration over the contact area A (at $z = 0$ or $z = L$) an expression for the total resistor current I_R is found. Not surprisingly, it is equal to Ohm's law, but with the resistance R expressed as a tensor summation. In general, also including n-type material, it can be written as:

$$I_R = \frac{V_R}{R}, \qquad \frac{1}{R} = \frac{A}{L} l_i l_j \sigma_{ij}^{\mathrm{maj}} \tag{2.133}$$

Stress-induced changes

In the case a mechanical stress X_{kl} is applied to this resistor, the above current changes with the following factor:

$$\frac{I_R(0)}{I_R(X_{kl})} = \frac{R(X_{kl})}{R(0)} = \frac{(A/L)(0)}{(A/L)(X_{kl})} \frac{l_i l_j \sigma_{ij}^{\mathrm{maj}}(0)}{l_i l_j \sigma_{ij}^{\mathrm{maj}}(X_{kl})} \tag{2.134}$$

The piezoresistive part of this factor is dominant and can be calculated from the theory in Section 2.3.1. This has been done for a specific case which is also characterised experimentally in the following chapter. It concerns a rosette of three p-type resistors which is subjected to an in-plane uniaxial stress in the [130] crystal direction. The resistors are oriented in the [100]-, [110]-, and [010]-directions. The calculations were performed numerically with the band parameters of Hensel [41] and the deformation potentials of Friedel [14]. The calculation time was two hours. They resulted in the resistance changes in Figure 2.25 which are plotted as a function of the stress magnitude. The calculations were repeated with the parameters of Fischetti and Laux [5]. This yielded equally flat curves for the [100]- and [010]-resistors, but a curve for the [110]-resistor which was 25% steeper.

The current changes can be calculated much quicker with the piezoresistive and compliance coefficients given in the previous section. By recalling Equation 2.114 and 2.115, the piezoresistive part of this can be written to the second

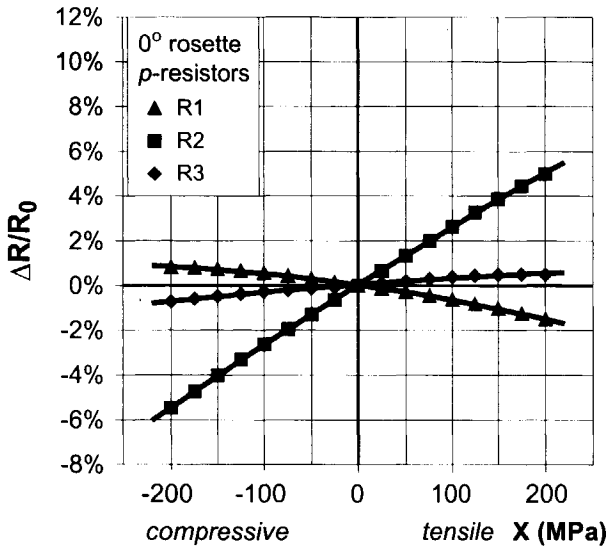

Figure 2.25: Calculated resistance changes as a function of stress of the p-type rosette resistors described in Chapter 3.

order in stress as:

$$\frac{l_i l_j \sigma_{ij}^{\mathrm{maj}}(0)}{l_i l_j \sigma_{ij}^{\mathrm{maj}}(X_{kl})} = 1 + l_i l_j \frac{\pi_{ijkl}}{\rho^{0,\mathrm{maj}}} X_{kl} + l_i l_j \frac{\pi_{ijklmn}}{\rho^{0,\mathrm{maj}}} X_{kl} X_{mn} \tag{2.135}$$

The piezoresistive coefficients in this equation, π_{ijkl} and π_{ijklmn}, can be transcribed to the reduced-index notation of Section 2.3.3. Measured values of the first-order coefficients can be looked up in e.g. Table 3.4, whereas the second-order coefficients can be found in Reference [50]. For a stress range of ± 250 MPa the second-order effect is relatively small compared to that of the piezojunction effect. This is difficult to deduce from the numerical values of the coefficients, but appears very clearly when graphs of both effects are compared, e.g. those of Figure 3.22 and 3.31.

The geometrical part of the current changes can be calculated with the stress-strain relation of Equation 2.114: $e_{ij} = S_{ijkl} X_{kl}$. To the first order in stress this yields:

$$\begin{aligned} \frac{(A/L)(0)}{(A/L)(X_{kl})} &= 1 - e'_{11} - e'_{22} + e'_{33} \\ &= 1 - (a_{1i}a_{1j} + a_{2i}a_{2j} - a_{3i}a_{3j})\, S_{ijkl} X_{kl} \end{aligned} \tag{2.136}$$

where a_{ij} is the transformation matrix relating the crystal axes x_i to the axes system with the z-axis parallel to l_i. In the stress range of ± 250 MPa the

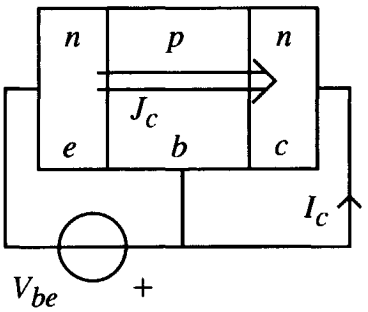

Figure 2.26: One-dimensional schematic of a bipolar *npn* transistor in forward bias. The main current I_c flows from the emitter e *via* the base b to the collector C and has a constant density J_c. It is driven by the base-emitter voltage V_{be}.

geometry changes the resistance with 0.3 percent at most. As a result, it can often be neglected compared to the piezoresistive effect. The stress also changes the product $l_i l_j$ in Equation 2.135. This yields a small additional term quadratic in stress which can often be neglected.

Anisotropy

The piezoresistive effect is highly anisotropic. This means that the stress-induced resistance change strongly depends on the orientation of the current and the stress with respect to the crystal axes. The only exception is the first-order resistance change due to a current parallel to a uniaxial stress in one of the {111} planes.

The anisotropy is often represented graphically with the aid of polar plots introduced by Kanda [79, 89, 90, 84, 70]. The principle of those plots is discussed in the next section for the case of the piezojunction effect.

2.4.3 Bipolar transistors

The main behaviour of most bipolar transistors can be derived from the one-dimensional device of Figure 2.26. This is done in detail in Appendix D. If this *npn* device is biased in forward, a collector current I_c is generated by the base-emitter voltage V_{be}. The relation between I_c and V_{be} will be developed here in terms of geometry, charge transport parameters, and mechanical stress.

Current-voltage characteristic

It was seen in Appendix D that the collector current I_c is given by the equation:

$$I_c = I_S \exp\left(\frac{qV_{be}}{k_B T}\right) \tag{2.137}$$

in which I_S is the saturation current. This saturation current depends on the properties of the base region only. More exactly, it depends on the base properties experienced by the minority charge carriers, since it is the minorities in the base which determine the main current in the transistor. Under the assumptions of Section 2.4.1, I_S can be written as:

$$I_S = k_B T \frac{A}{W N_A} \sum_m l_i l_j \mu_{ij}^m \left(n_{i0}^m\right)^2 \tag{2.138}$$

where the sum runs over the energy bands m of the minorities. The symbol A is the area of the emitter-base junction, W is the effective base width, N_A is the doping concentration of the majorities in the base, and l_i is the direction vector of the current. In addition, μ_{ij}^m is the mobility tensor of the minorities in band m, while n_{i0}^m is the pn-product in thermodynamic equilibrium of band m with respect to all opposite bands (see Section 2.2.4).

The saturation current can be expressed in terms of minority conductivity, defined in Section 2.3.3. For this purpose the pn-product n_{i0}^m is rewritten using the assumption of extrinsic material. The total majority concentration in that case is equal to the doping concentration N_A:

$$n_{i0}^m = n_0^m \sum_u p_0^u = n_0^m N_A \tag{2.139}$$

In addition, the mobility and the concentration of band m are combined to the minority conductivity tensor in thermodynamic equilibrium $\sigma_{ij}^{\min}$:

$$q \sum_m \left(n^m \mu_{ij}^m\right)^{\min} = \sigma_{ij}^{\min} \tag{2.140}$$

This means that the I_S of Equation 2.138 is equal to:

$$I_S = \frac{k_B T}{q} \frac{A}{W} l_i l_j \sigma_{ij}^{\min} \tag{2.141}$$

This expression is quite similar to that of the inverse resistance of a resistor, described by Equation 2.133. The only differences are the factor $k_B T/q$ and the use of the minority conduction instead of the majority conduction.

Stress-induced changes

The collector current of the bipolar transistor changes when a mechanical stress is applied. The change is schematically drawn in Figure 2.27, where it is indicated to be relatively small. It is shown in the following that this is the case for the stress range studied in this work (up to 200 MPa). It is also shown there that the change can be both positive and negative, depending on the current orientation, the stress orientation, and the sign of the stress.

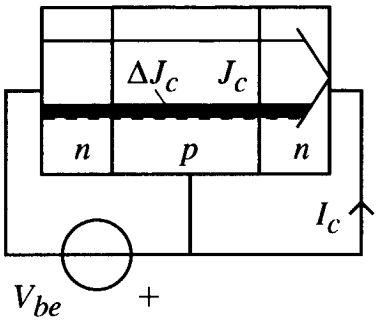

Figure 2.27: Schematic of the change ΔJ_c in the main electron current density J_c of an *npn* transistor as a result of stress.

The collector current change is the result of the influence of the mechanical stress tensor X_{kl} on the saturation current I_S. It changes the saturation current current with the following factor:

$$\frac{I_S(0)}{I_S(X_{kl})} = \frac{(A/W)(0)}{(A/W)(X_{kl})} \frac{l_i l_j \sigma_{ij}^{\min}(0)}{l_i l_j \sigma_{ij}^{\min}(X_{kl})} \tag{2.142}$$

This is very similar to the stress-induced changes in a resistor current, as can be seen from a comparison with Equation 2.134. The differences are that the resistor length is replaced by the base width, and the majority conductivity by the minority conductivity.

As a result, the collector current changes can be calculated from the theory in Section 2.3.1. In that section the minority conductivity is related to the *pn* product and the minority mobility, of which the stress dependences were calculated in terms of the energy bands in Section 2.2. The calculations have been made for eight specific cases which are characterised experimentally in the following chapter. They concern four *npn* and four *pnp* transistors with the collector currents in different directions and subjected to uniaxial stresses in various orientations. The calculations were again performed numerically with the band parameters of Hensel [41] and the deformation potentials of Friedel [14]. They resulted in the plots of Figure 2.28 to 2.31. The calculations were repeated with the parameters of Fischetti and Laux [5]. This resulted in plots of similar shapes, but 30% larger in scale.

The plots show a strongly anisotropic behaviour, depending on both the stress and the current orientation. In addition, they show large differences between *npn* and *pnp* transistors. The current changes are in the same order of magnitude as the piezoresistive effect, on the stress range considered. However, they are considerably less linear. The nonlinearity is mainly second-order in this stress range, as shown by the parabolic least-squares fits.

The collector current changes can also be expanded into a series with the aid

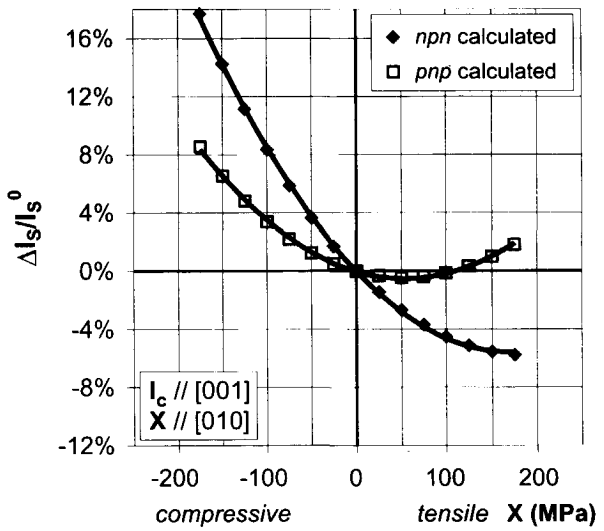

Figure 2.28: Calculated saturation current changes of an *npn* and a *pnp* transistor as a function of uniaxial stress, for the current and stress orientation indicated. The points are connected with a second-order least-squares fit.

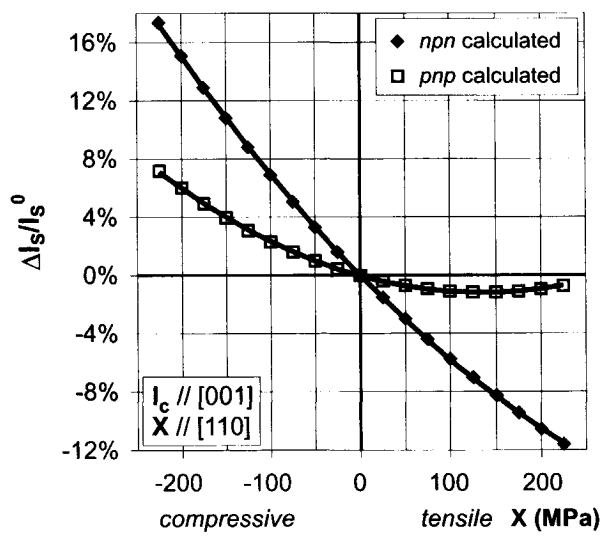

Figure 2.29: Calculated saturation current changes of an *npn* and a *pnp* transistor as a function of uniaxial stress, for the current and stress orientation indicated. The points are connected with a second-order least-squares fit.

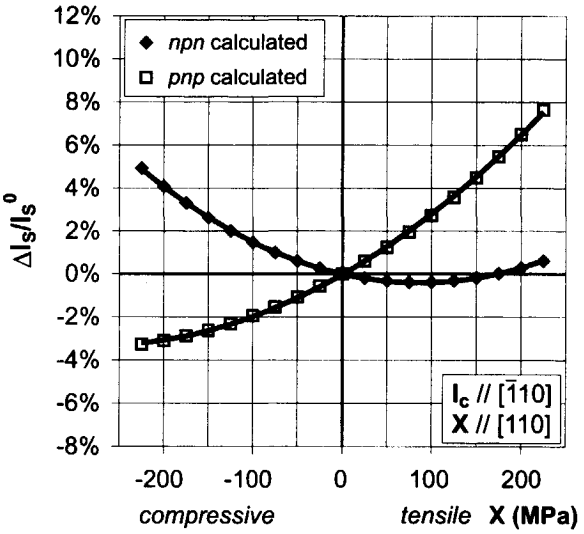

Figure 2.30: Calculated saturation current changes of an *npn* and a *pnp* transistor as a function of uniaxial stress, for the current and stress orientation indicated. The points are connected with a second-order least-squares fit.

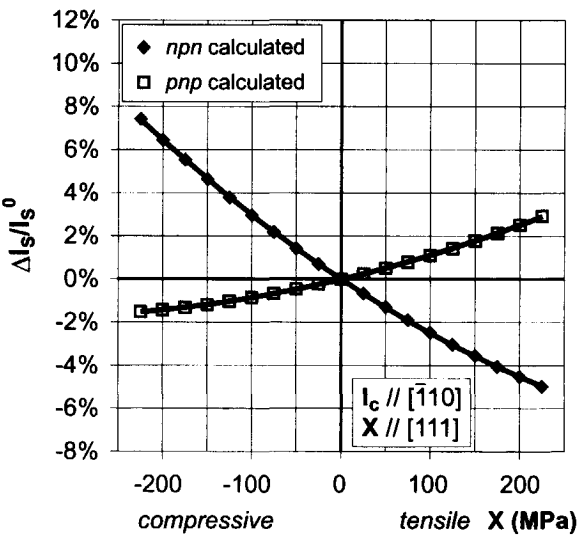

Figure 2.31: Calculated saturation current changes of an *npn* and a *pnp* transistor as a function of uniaxial stress, for the current and stress orientation indicated. The points are connected with a second-order least-squares fit.

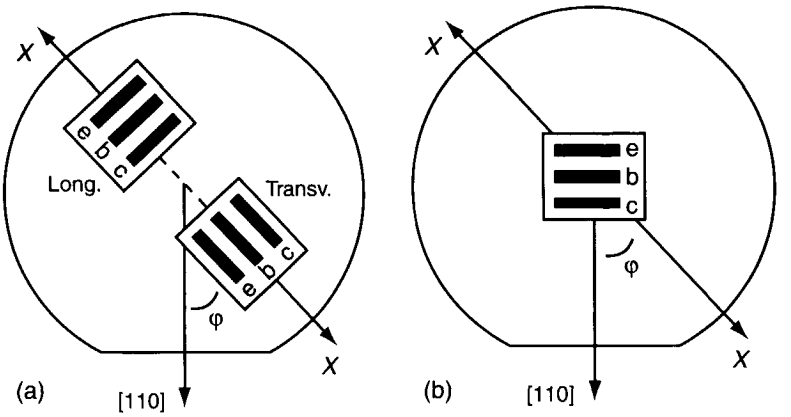

Figure 2.32: Lateral transistors under uniaxial stress. (a): the stress is longitudinal or transverse to the current directions. (b): the current direction is fixed to the [110] direction.

of the piezojunction coefficients. Using Equation 2.114 and 2.116 this yields:

$$\frac{l_i l_j \sigma_{ij}^{\min}(0)}{l_i l_j \sigma_{ij}^{\min}(X_{kl})} = 1 + l_i l_j \frac{\zeta_{ijkl}}{\rho^{0,\min}} X_{kl} + l_i l_j \frac{\zeta_{ijklmn}}{\rho^{0,\min}} X_{kl} X_{mn} \qquad (2.143)$$

The coefficients in this equation are known from theory and experiment and can be found in Table 2.5 and 3.3. The geometrical part of the collector current changes is exactly equal to that of the resistor current. It is given by Equation 2.136, where L should be replaced by W. In most cases it is negligible.

Anisotropy

In analogy to the piezoresistive effect, the piezojunction effect is very anisotropic. Its magnitude and sign depend strongly on the orientation of both the current and the stress with respect to the crystal axes of silicon. If these orientations are free to be modified they are an excellent way to optimize the stress sensitivity for a specific application.

The optimization of the stress sensitivity is facilitated by using a graphical representation. Such representations can be applied in the design of piezoresistors [90, 86, 84]. Generally, they consist of a polar plot indicating the first-order sensitivity of a resistor to uniaxial stress as a function of stress and current direction in a specific crystal plane. Such polar plots are given here for the case of a bipolar transistor. They are based on Equation 2.143 and the first-order piezojunction coefficients in Table 2.5, determined from the experiments described in the next chapter.

The graphical representation of anisotropy has the inherent limitation that it can only describe a very small section of all possible current and stress orien-

Table 2.6: Sensitivity to in-plane uniaxial stress of the saturation current of a transistor for orientations in which it is isotropic, in 10^{-11} Pa^{-1}.

Wafer plane	Current or.	Stress or.	npn	pnp
{100}	Vertical	Transverse	-43.4	-13.8
{111}	Vertical	Transverse	-15.1	20.5
{111}	Lateral	Longitudinal	-17.6	-82.9
{111}	Lateral	Transverse	-28.2	4.0

tations. This need not be a large problem because the orientations are already limited in practice for technology reasons. Stress, for instance, generally occurs in the plane of the wafer only.

Configuration of the graphs In the following the anisotropy graphs are shown for the most common transistors. These are vertically or laterally fabricated in either {100} or {111} wafers, as explained in Appendix D. They are subject to a uniaxial stress in the wafer plane; see Figure 2.32. If the stress is more complex, it can usually be considered as a superposition of uniaxial stresses. Their first-order electrical effects can be read from the graphical representation, and can be added to find the total effect.

The graphs only describe the directions between 0 and 180 degrees because of the crystal symmetry. The stress sensitivities are positive in the upper halves, while they are negative sensitivities in the lower halves. The graphs are plotted for both *npn* and *pnp* transistors.

Vertical transistors The most important transistors in circuit design are vertical. In vertical transistors the main current is flowing through the base in a direction normal to the wafer surface. The base of such transistors can be very thin, making them much faster than lateral transistors.

The stress sensitivity of vertical transistors is quite special. If they are fabricated on {100} or {111} wafers, the first-order sensitivity to in-plane stress is *isotropic*. This means that the sensitivity is independent of the angle of the uniaxial stress in the wafer plane. This is mainly the result of the cubic symmetry of the unstressed silicon. After some manipulation of Equation 2.143 the sensitivity to a uniaxial stress X can be written as:

$$\frac{\partial}{\partial X}\left(\frac{I_S(X)}{I_S(0)}\right)(\varphi) = -\zeta_{12}, \text{ stress in a } \{100\} \text{ plane} \qquad (2.144)$$

$$= -\frac{1}{3}\left(\zeta_{11} + 2\zeta_{12} - \zeta_{44}\right), \text{ stress in a } \{111\} \text{ plane}$$

These expressions can be evaluated by inserting the values of Table 2.5. The results are shown in Table 2.6.

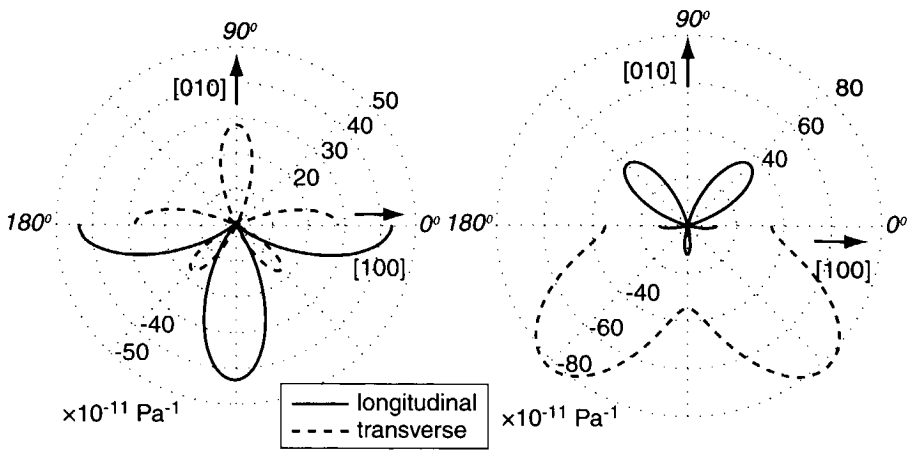

Figure 2.33: Sensitivity to longitudinal and transverse stress of lateral bipolar transistors in a {100} crystal plane, when the plane is rotated. Left: *npn*; Right: *pnp*.

Lateral transistors; longitudinal and transverse sensitivity Lateral transistors are not as fast as vertical ones, but are available in many IC processes. The main current in those transistors is flowing through the base in a direction parallel to the wafer surface. This means that during the IC design the current direction is free to be chosen within the wafer plane.

The stress sensitivity of lateral transistors is very anisotropic and can therefore be tuned for specific needs. A traditional way of defining the anisotropy is to consider the sensitivity to stress parallel and perpendicular to the current direction as a function of that direction [90]. This configuration is sketched in Figure 2.32.a. It yields the longitudinal and transverse sensitivities, respectively. For the (001) plane they are plotted in Figure 2.33. Among other things, the figure shows that the lateral sensitivity of *pnp* transistors is very high in <110>directions, while their transverse sensitivity is zero near the <100>directions.

A special case arises for the {111} planes, where the sensitivities are again isotropic. Manipulation of Equation 2.143 learns that they are given by:

$$\frac{\partial}{\partial X}\left(\frac{I_S(X)}{I_S(0)}\right)(\varphi) = -\frac{1}{2}(\zeta_{11}+\zeta_{12}+\zeta_{44}) \quad \text{longitudinal} \tag{2.145}$$

$$= -\frac{1}{6}(\zeta_{11}+5\zeta_{12}-\zeta_{44}) \quad \text{transverse} \tag{2.146}$$

Their numerical values are given in Table 2.6.

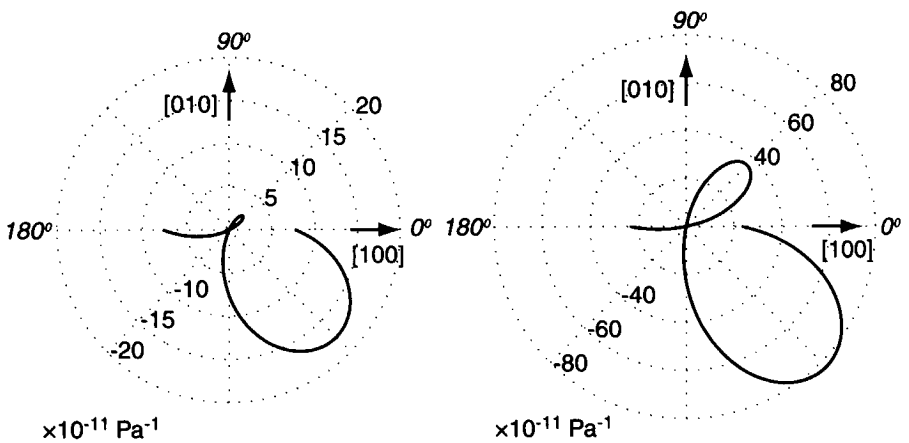

Figure 2.34: Stress sensitivity of lateral bipolar transistors in a standard [110] orientation as a function of the uniaxial stress direction in the (001) crystal plane. The stress is rotated while the transistor is fixed on the wafer. Left: *npn;* Right: *pnp.*

Lateral transistors; sensitivity in standard orientation It often occurs that the stresses have an arbitrary orientation in the wafer plane, but that the transistor orientation is fixed. This situation is sketched in Figure 2.32.b. In many IC processes the transistors must be oriented along either the horizontal or the vertical die edge. This corresponds to a direction parallel or perpendicular to the wafer flat, which is equivalent to a <110>direction in the crystal axes system.

The stress sensitivities of such lateral transistors are plotted in Figure 2.34 and 2.35. They appear to be largest for *pnp* transistors. For both transistor types they are very anisotropic, and often reduce to zero for a specific stress direction.

2.5 Conclusions

In this chapter the effects of mechanical stress on the characteristics of transistors and resistors were calculated from first principles. These calculations were performed in several steps, which were implemented in a computer program. In order to give a brief overview of the calculations the main steps are given in Figure 2.36.

Stress deforms the silicon crystal and thereby the period of the electron wave functions. This shifts and deforms the band edges and lifts their degeneracy. These processes can be described with the aid of the deformation potential theory. They modify the bandgap and change the effective mass of the charge

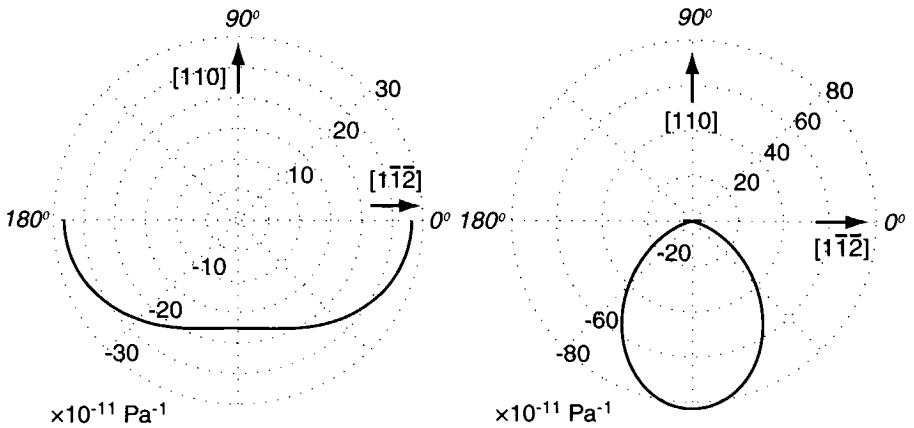

Figure 2.35: Stress sensitivity of lateral bipolar transistors in a standard [110] orientation as a function of the uniaxial stress direction in the $(1\bar{1}1)$ crystal plane. The stress is rotated while the transistor is fixed on the wafer. Left: *npn;* Right: *pnp.*

carriers.

The band modifications change the concentration, the mobility, and therefore the electrical conductivity of the charge carriers. These changes are generally different for each band. The change of the total, macroscopic conductivity is therefore a weighted sum of the contributions of the individual bands. It can be divided into changes in the average mobility and changes in the *pn* product. Both changes are of comparable magnitude on the stress range.

If the stress is sufficiently small, the conductivity change can be described by a series expansion. This expansion can be evaluated rapidly and gives a good description of the experimental observations. The coefficients of the expansion are called the piezoresistive coefficients when they describe the conductivity of majority charge carriers. For minority charge carriers they are called the piezojunction coefficients. All coefficients are tensor properties of the material, and must therefore have the symmetry of the crystal of which the material consists. They can be calculated from the band changes, but can also be determined from experiments.

The conductivity of majorities determines the resistance of a resistor, whereas the minority conductivity sets the saturation current of a bipolar transistor. The stress effects on those characteristics can therefore be described with the piezoresistive and piezojunction coefficients, respectively. Theoretically, the effects are tightly coupled. The resistance changes depend on the stress-induced changes in the bands of the majority charge carriers, whereas the saturation current changes depend on the changes in the conduction and the valence bands together. For specific orientations, they can be calculated numerically. The effects

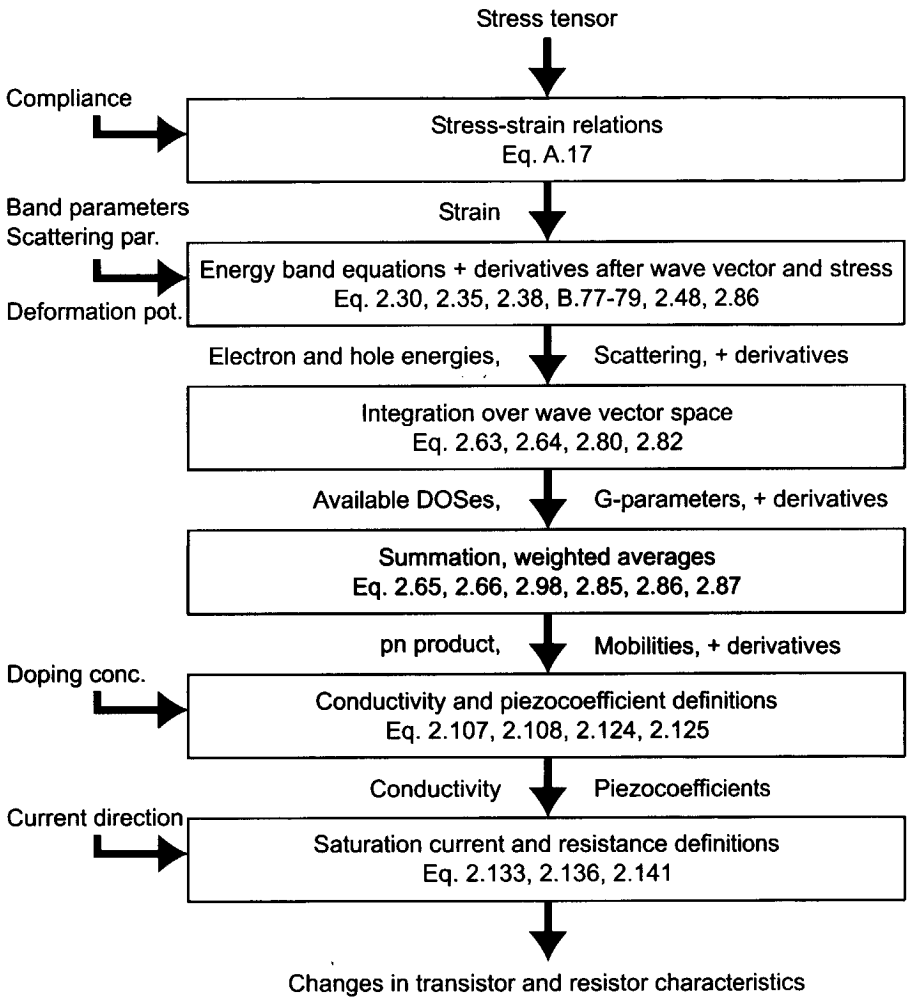

Figure 2.36: Flow chart of the computer program used to calculate the stress induced changes in the characteristics of transistors and resistors. The piezojunction and piezoresistive coefficients were obtained at an input stress of zero.

are very anisotropic with the stress and current orientations. The anisotropy of the first-order coefficients can be represented in polar plots.

In the following chapter measurements are presented to verify the theory and determine the piezocoefficients experimentally.

Bibliography

[1] M. G. Middelhoek, *The identification of analytical device models*, Ph. D. Thesis of the Delft University of Technology, Delft University Press, Delft, The Netherlands, 1992.

[2] N. M. Ashcroft and N. D. Mermin, *Solid state physics*, Saunders College, Philadelphia, 1976.

[3] I. Ipatova and V. Mitin, *Solid-state electronics*, Addison-Wesley, Reading, Mass., 1996.

[4] I. M. Tsidilkovski, *Band structure of semiconductors, Int. Ser. Scie. Sol. St., 19,* Pergamon Press, Oxford, 1982.

[5] M. V. Fischetti, and S. E. Laux, Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys, *J. Appl. Phys., 80* (1996) 2234-52.

[6] G. I. Pikus and G. L. Bir, Effect of deformation on the hole energy spectrum of germanium and silicon, *Sov. Phys.-Solid State, 1* (1960) 1502-1517.

[7] W. H. Kleiner and L. M. Roth, Deformation potential in germanium from optical absorption lines for exciton formation, *Phys. Rev. Lett., 2* (1959) 334-336.

[8] G. L. Bir and G. E. Pikus, *Symmetry and strain-induced effects in semiconductors,* Wiley, New York, 1974.

[9] J. Bardeen and W. Shockley, Deformation potentials and mobilities in non-polar crystals, *Phys. Rev., 80* (1950) 72-80.

[10] C. Herring and E. Vogt, Transport and deformation-potential theory for many-valley-semiconductors with anisotropic scattering, *Phys. Rev., 101* (1956) 944-961.

[11] E. N. Adams, Elastoresistance in p-type Ge and Si, *Phys. Rev., 96* (1954) 803-804.

[12] O. Madelung (ed.), *Semiconductors - Basic data*, Springer, Berlin, 1996.

[13] A. K. Ramdas and S. Rodriguez, Piezospectroscopy of semiconductors, *Semiconductors and semimetals, 36* (D. G. Seiler and C. L. Littler, eds.), Wiley, New York, 1992.

[14] P. Friedel, M. S. Hybertsen, and M. Schluter, Local empirical pseudopotential approach to the optical properties of Si/Ge superlattices, *Phys. Rev. B, 39* (1989) 7974-7977.

[15] C. G. van de Walle and R. M. Martin, Theoretical calculations of disicontinuities in the Si/Ge system, *Phys. Rev. B, 34* (1986) 5621-5634.

[16] C. G. van de Walle, Band lineups and deformation potentials in the model-solid theory, *Phys. Rev. B, 39* (1989) 1871-1883.

[17] J. D. Wiley, Valence-band deformation potentials for the III-V compounds, *Solid-St. Comm., 8* (1970), 1865-1868.

[18] J.-C. Merle, M. Capizzi, and P. Fiorini, Uniaxially stressed silicon: fine structure of the exciton and deformation potentials, *Phys. Rev. B, 17* (1978) 4821-4834.

[19] L. D. Laude, F. H. Pollak, and M. Cardona, Effects of uniaxial stress on the indirect exciton spectrum of silicon, *Phys. Rev. B, 3* (1971) 2623-2636.

[20] J. C. Hensel and G. Feher, Cyclotron resonance experiments in uniaxially stressed silicon: valence band inverse mass parameters and deformation potentials, *Phys. Rev., 129* (1963) 1041-1062.

[21] J. C. Hensel, H. Hasegawa, and M. Nakayama, Cyclotron resonance in uniaxially stressed silicon, II. Nature of the covalent bond, *Phys. Rev., 138* (1965) 225-238.

[22] K. Suzuki, Distorsion of energy surfaces by a strain field, *Progr. Theor. Phys. Supplement, 101* (1990) 215-20.

[23] Y. Kanda and K. Suzuki, Origin of the shear piezoresistance coefficient p44 of n-type silicon, *Phys. Rev. B, 43* (1991-I) 6754-6755.

[24] E. O. Kane, Strain effects on opticial critical-point structure in diamond-type crystals, *Phys. Rev., 178* (1969) 1368-1398.

[25] H. Brooks, *Advances in electronics and electron physics* (L. Marton ed.), *Vol. 7, p.85,* Academic Press, New York, 1955.

[26] J. E. Dijkstra and W. Th. Wenckenbach, Hole transport in strained Si, *J. Appl. Phys., 81* (1997) 1259-1263.

[27] C. K. Kim, M. Cardona, and S. Rodriguez, Effect of free carriers on the elastic constants of p-type silicon and germanium, *Phys. Rev. B, 13* (1976) 5429-5441.

[28] J. T. Lenkkeri, Nonlinear effects in the piezoresistivity of p-type resistivity, *Phys. Status Solidi B, 136* (1986) 373-385.

[29] W. Greiner, *Quantum mechanics, an introduction,* Ch. 11, Springer, Berlin, 1989.

[30] H. Hasegawa, Theory of the cyclotron resonance in strained silicon crystals, Phys. Rev., 129 (1963) 1029-1040.

[31] A. H. Marshak and C. M. van Vliet, Electrical current and carrier density in degenerate materials with nonuniform band structure, *Proc. IEEE, 72* (1984) 148-164.

[32] A. van der Ziel, *Solid State Physical Electronics, 3rd ed.,* Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

[33] M. Lundstrom, *Fundamentals of carrier transport, Modular series on solid state devices, X,* Addison-Wesley, Reading, Mass., 1992.

[34] P. A. Gough, *Fundamental equations, in: Bipolar and bipolar-MOS integration,* P. A. H. Hart (ed.), Elsevier, Amsterdam, 1994.

[35] S. J. Chapman, *Introduction to FORTRAN 90/95,* McGraw-Hill, New York, 1998.

[36] J. Stoer and R. Bulirsch, *Introduction to numerical analysis, 2nd ed.,* Springer, New York, 1993.

[37] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in FORTRAN 77, The art of scientific computing, 2nd ed.,* Cambridge University Press, Cambridge, 1996.

[38] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in FORTRAN 90, The art of parallel scientific computing, 2nd ed.,* Cambridge University Press, Cambridge, 1996.

[39] J. J. Stickler, H. J. Zeiger, and G. S. Lax, Quantum effects in Ge and Si. I, *Phys. Rev., 127* (1962) 1077-1084.

[40] R. N. Dexter, H. J. Zeiger, and B. Lax, Cyclotron resonance experiments in silicon and germanium, *Phys. Rev., 104* (1956) 637-644.

[41] J. C. Hensel in: R. G. Humphreys, Valence band averages in silicon: anisotropy and nonparabolicity, *J. Phys. C: Solid State Phys., 14* (1981) 2935-2942.

[42] Y. Ohmura, Piezoresistance effect in p-type Si, *Phys. Rev. B, 42* (1990) 9178-9181.

[43] Y. Ohmura, Numerical study of the piezoresistance effect in p-type Si, *J. Phys. Soc. of Jpn., 61* (1992) 217-26.

[44] P. Kleinmann, B. Semmache, M. Le Berre, and D. Barbier, Stress-dependent hole effective mass and piezoresistive properties of p-type monocrystalline and polycrystalline silicon, *Phys. Rev. B, 57* (1998-I) 8966-8971.

[45] M. A. Green, Intrinsic concentration, effective densities of states, and effective mass in silicon, *J. Appl. Phys., 67* (1990) 2944-2954.

[46] J. W. Brown and R. V. Churchill, *Fourier series and boundary value problems, 5th ed.,* McGraw-Hill, New York, 1993.

[47] W. E. Boyce and R. C. DiPrima, *Elementary differential equations and boundary value problems, 4th ed.,* Wiley, New York, 1986.

[48] J. F. Nye, *Physical properties of crystals, 2nd ed.,* Clarendon Press, Oxford, 1985.

[49] J. E. Dijkstra, *Monte Carlo simulation of hole transport in Si, Ge and* $Si_{1-x}Ge_x$, Ph. D. thesis of the Delft University of Technology, Delft, Netherlands, 1997.

[50] K. Matsuda, K. Suzuki, K. Yamamura, and Y. Kanda, Nonlinear piezoresistive coeficients in silicon, *J. Appl. Phys., 73* (1993) 1838-1847.

[51] C. Herring, Transport properties of a many-valley semiconductor, *Bell Syst. Techn. J., 34* (1955) 237-290.

[52] P. Kramer and L. J. van Ruyven, Position of the band edges of silicon under uniaxial stress, *Appl. Phys. Lett., 20* (1972) 420-422.

[53] R. Rueda, J. Slinkman, D. Chidambarrao, L. Moszkowicz, P. Kaszuba, and M. Law, Mechanical stress characterization of shallow trench isolation by Kelvin probe force microscopy, *Proc. Mater. Res. Soc. Sympos. Front-end Processing, Physics and technology of dopant-dependent interactions,* USA, Spring 1999, pp.245-250.

[54] L. Egley and D. Chidambarrao, Strain effects on device characteristics: implementation in drift-diffusion simulators, *Solid-St. Electron., 36* (1993) 1653-64.

[55] A. H. Marshak and R. Shrivastava, Law of the junction for degenerate material with position-dependent band gap and electron affinity, *Solid-St. Electron., 22* (1979) 567-571.

[56] G. C. M. Meijer, Thermal sensors based on transistors, *Sensors and Actuators, 10* (1986) 103-125.

[57] A. Bakker, *High-accuracy CMOS smart temperature sensors,* Ph. D. Thesis of the Delft University of Technology, Delft, The Netherlands, 2000.

[58] D. K. Schroder, *Semiconductor material and device characterization,* Wiley, New York, 1990.

[59] B. J. Sloan and J. R. Hauser, Effects of uniaxial compressive stress on minority-carrier lifetime in silicon and germanium, *J. Appl. Phys., 41* (1970) 3504-3508.

[60] A. A. Patrin, M. I. Tarasik, V. D. Tkachev, and A. M. Yanchenko, Influence of uniaxial compression on the nonequilibrium carrier lifetime in silicon, *Sov. Phys. Semicond., 8* (1975) 1326-1327.

[61] W. Wlodarski and B. Moeschke, The effect of hydrostatic pressure on the characteristics of the forward biased p-n junctions, *Electron Technol., 13* (3) (1980) 3-42.

[62] Y. Matukura, Minority carrier lifetime in uniaxially stressed germanium, *Japan. J. Appl. Phys., 4* (1965) 72-73.

[63] T. Otaredian, *Contactless microwave lifetime measurement,* Ph. D. thesis of Delft University of Technology, Delft, Netherlands, 1992.

[64] M. Cardona, W. Paul, and H. Brooks, Dielectric constant measurements in germanium and silicon at radio frequencies as a function of temperature and pressure, *Proc. Int. Conf. Solid State Phys. in Electron. and Telecomm.,* Brussels, Belgium, June 2-7, 1958, Academic Press, London, 1960, pp. 206-214.

[65] D. E. Aspnes, Optical functions, in: *Properties of silicon, EMIS Datareview series 4,* INSPEC, London, 1988, pp. 59-80.

[66] Z. H. Levine, H. Zhong, S. Wei, D. C. Allen, and J. W. Wilkins, Strained silicon: A dielectric-response calculation, *Phys. Rev. B, 45* (1992) 4131-4140.

[67] P. Etchegoin, J. Kircher, and M. Cardona, Elasto-optical constants of Si, *Phys. Rev. B, 47* (1993) 10,292-10,303.

[68] B. E. A. Saleh and M. C. Teich, *Fundamentals of photonics,* Wiley, New York, 1991.

[69] S. M. Sze, *Physics of semiconductor devices,* Wiley, New York, 1981.

[70] S. Durand and C. R. Tellier, Linear and non-linear piezoresistance coefficients in cubic semiconductors. I. Theoretical formulations, *J. Phys. III France, 2* (1996) 237-266.

[71] W. P. Mason and R. N. Thurston, Use of piezoresistive materials of displacement, force and torque, *J. Acoust. Soc. Am., 29* (1957) 1096-1101.

[72] C. S. Smith, Macroscopic symmetry and properties of crystals, in: *Solid State Physics, Vol. 6,* F. Seitz and D. Turnbull (eds.), Academic Press, New York, 1958, pp. 175-249.

[73] Y. Ohmura, Role of the valence band density of states in the piezoresistance of p-type semiconductor Si and Ge, *J. Phys. Soc. of Jpn., 62* (1993) 3615-20.

[74] Y. Ohmura and J. Nishimura, Second-order piezoresistance coefficients of p-type Si, *Phys. Status Solidi A, 149* (1995) 659-667.

[75] Y. Ohmura and W. Morinaga, Fourth-order poezoresistance coefficients in cubic semiconductors, *Jpn. J. Appl. Phys., 39* (2000) 3483-3487.

[76] Y. Ohmura, Third-order piezoresistance coefficients in p-type Si, *Jpn. J. Appl. Phys., 33* (1994) 3314-3318.

[77] D. A. Bittle, J. C. Suhling, R. E. Beaty, R. C. Jaeger, and R. W. Johnson, Piezoresistive stress sensors for structural analysis of electronic packages, *J. Electron. Pack., 113* (1991) 203-214.

[78] A. Nathan and T. Manku, Piezoresistance and drift-diffusion model in strained silicon, *Proc. 6th Int. Conf. Simul. Semicond. Dev. Processes (SISDEP '95),* Erlangen, Germany, September 1995, Vol. 6, pp. 94-97.

[79] Y. Kanda, Piezoresistance effect of silicon, *Sensors and Actuators A, 28* (1991) 83-91.

[80] C. S. Smith, Piezoresistance effect in germanium and silicon, *Phys. Rev., 94* (1954) 42-49.

[81] F. Fruett and G. C. M. Meijer, Measurement and compensation of piezoresistive coefficient π_{44} for minority-carrier concentration, *Electron. Letters, 36* (2000) 173-175.

[82] F. Fruett and G. C. M. Meijer, Exploration of the piezojunction effect using PNP lateral transistors on [100] silicon, *Proc. 14th Eur. Conf. Solid-State Transducers (Eurosensors XIV),* Copenhagen, Denmark, 27-30 August 2000, pp.287-288.

[83] F. Fruett, Private communication, November 27, 2000.

[84] S. Middelhoek and S. A. Audet, *Silicon Sensors, Ch. 3,* Academic Press, London, 1989.

[85] Y. Ohmura and W. Morinaga, Comment on "Second-order piezoresistance coefficients of n-type silicon", *Jpn. J. Appl. Phys., 35* (1996) L 280-L 281.

[86] A. Belu-Marian, E. Candet and A. Devenyi, *Piezoresistive sensors, in: Thin film resistive sensors,* C. Ciureanu and S. Middelhoek (eds.), IOP Publishing, Bristol, 1992.

[87] H. K. Gummel, A charge control relation for bipolar transistors, *Bell Syst. Techn. J.*, *49* (1970) 115-120.

[88] J. L. Moll and I. M. Ross, The dependence of transistor parameters on the distribution of base layer resistivity, *Proc. IRE*, *44* (1956) 72-78.

[89] Y. Kanda, Graphical representation of the piezoresistance coefficients in silicon-Shear coefficients in plane, *Jpn. J. Appl. Phys.*, *26* (1987) 1031-1033.

[90] Y. Kanda, A graphical representation of the piezoresistance coefficients in silicon, *IEEE Trans. El. Dev.*, *ED-29* (1982) 64-70.

Chapter 3

Measurements of Stress-induced Changes

The macroscopic model of the piezojunction effect is of practical use only if it is supported by measurements. Measurements can validate the structure of the model and yield numerical values of its parameters. These parameters can subsequently be compared with those derived from the more fundamental microscopic model.

The measurements in this chapter describe the influence of mechanical stress on the relation between the base-emitter voltage and the collector current. In addition, they describe the stress-induced changes in resistors. These resistance changes have also been described in literature and can therefore validate the transistor measurements. All devices lie in different crystal planes. In addition, they are subjected to uniaxial stress which ranges from -200 to +200 MPa and has different orientations with respect to the crystal axes. Finally, the devices are normally biased.

3.1 Principle of the measurement set-up

The piezojunction and piezoresistive effects can be characterized experimentally when the following basic elements are present:

- Devices, preferably accompanied by a model structure,
- A generator of mechanical stress, and
- Electrical measurement equipment.

The principles of these elements are discussed in this section, whereas their implementation is treated in the following section.

3.1.1 Devices and their stress dependence

The experimental devices should satisfy the assumptions made in the theory (see Section 2.4.1). These assumptions are not very restrictive: they can be met quite well with standard technology and measurement equipment. In this case the devices can be modelled with the following model structure.

Transistors

In Section 2.4.3 it was shown that at a constant base-emitter voltage, the collector current of a transistor I_c is particularly influenced by a stress X_{kl} through the minority conductivity $\sigma_{ij}^{\min}$:

$$\frac{I_c(0)}{I_c(X_{kl})} \approx \frac{l_i l_j \sigma_{ij}^{\min}(0)}{l_i l_j \sigma_{ij}^{\min}(X_{kl})} \tag{3.1}$$

where l_i is the direction vector of the current. This influence is relatively small in the stress domain considered and is sufficiently described by a second-order power series:

$$\frac{l_i l_j \sigma_{ij}^{\min}(0)}{l_i l_j \sigma_{ij}^{\min}(X_{kl})} = 1 + l_i l_j \frac{\zeta_{ijkl}}{\rho^{0,\min}} X_{kl} + l_i l_j \frac{\zeta_{ijklmn}}{\rho^{0,\min}} X_{kl} X_{mn} \tag{3.2}$$

The ζ_{ijkl} and ζ_{ijklmn} here are the first- and second-order piezojunction tensors. These tensors contain 810 coefficients in total. In Section 2.3.3 it was shown, however, that this number is reduced to only 12 by knowledge about the device operation and the material symmetry.

The first aim of the measurements is to verify whether this drastic reduction of coefficients is indeed allowed. The second aim is to extract them. For this purpose it is useful to rewrite Equation 3.2 as an explicit function of the three independent first-order piezojunction coefficients ζ_i^{I}, the nine second-order coefficients ζ_j^{II}, and the stress *magnitude* X:

$$\frac{I_c(0)}{I_c(X)} \approx 1 + \sum_{i=1}^{3} a_i \zeta_i^{\mathrm{I}} X + \sum_{j=1}^{9} b_j \zeta_j^{\mathrm{II}} X^2 \tag{3.3}$$

The factors a_i and b_i are vector constants which contain all information about the *orientation* of the stress and the current. For the following experiments they are known in advance. The stress magnitude X is varied during the measurements. Finally, the ratio of the collector currents with and without stress is measured. Since Equation 3.3 contains nine unknown ζ_j^{II}, it requires at least nine different experiments to find them all. This can be done by varying the orientations of the stress and the current.

Our experiments were performed for six different orientations. This gives a good indication about the model structure, but does not enable the determination of all coefficients. The current in the experiments was in either the [001], the $[\bar{1}10]$, or the $[1\bar{1}1]$ direction of the crystal. This was done by making

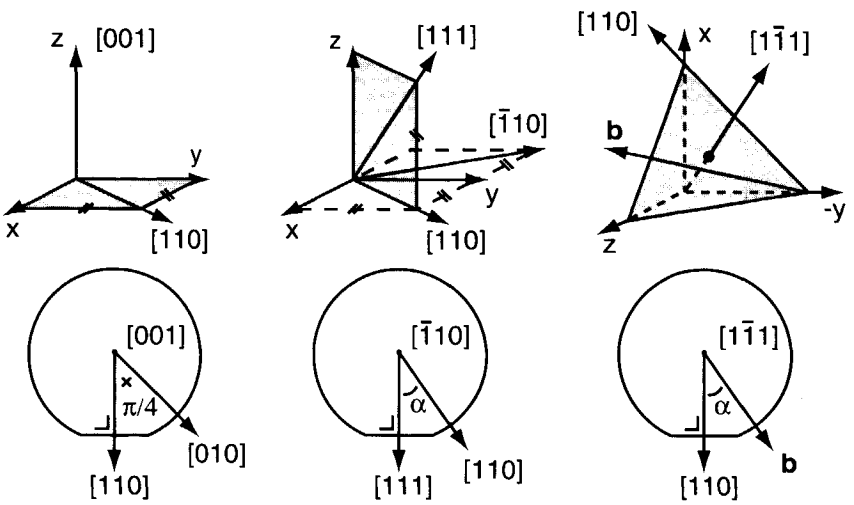

Figure 3.1: Orientations of the current and the uniaxial stress used in the experiments. *Above:* with respect to the crystal axes. *Below:* with respect to the employed silicon wafers. The currents are directed out of the wafer planes, which are drawn in grey in the upper figures. The vector **b** points in the direction $\left[\left(\sqrt{6}-1\right),\left(\sqrt{6}+1\right),2\right]$.

vertical transistors in different silicon wafers, as illustrated by Figure 3.1. The stress was chosen to be uniaxial and in the plane of the wafer. In each wafer it could be directed along two vectors q_i, as shown by Figure 3.1. The tensor of this stress is equal to $X_{kl} = q_k q_l X$. The six experiments yield a system of six equations in the form of Equation 3.3. This system can be written in matrix form as:

$$I_i = 1 + \sum_{j=1}^{3} A_{ij}\zeta_j^{\mathrm{I}} X + \sum_{k=1}^{9} B_{ik}\zeta_k^{\mathrm{II}} X^2, \quad i \in \{1,2,..,6\} \tag{3.4}$$

where I_i contains the six ratios of $I_c(0)/I_c(X)$, and the matrices A and B contain the different a_i and b_i, respectively.

Since it is impossible to resolve all twelve piezocoefficients from this system, it is better to combine them into linear combinations that can be found. These combinations are denoted by a quote: $\zeta_i^{\mathrm{I}\prime}$ and $\zeta_j^{\mathrm{II}\prime}$. They are given by the vectors of the row space of A and B [1]. For the orientations given in Figure 3.1 they are:

$$\zeta_i^{\mathrm{I}\prime} = \begin{bmatrix} \zeta_{11} - \zeta_{44} \\ \zeta_{12} \end{bmatrix}, \quad \zeta_j^{\mathrm{II}\prime} = \begin{bmatrix} \zeta_{111} - 4\zeta_{616} + 2\zeta_{414} + 2\zeta_{456} \\ \zeta_{112} + \frac{1}{2}\zeta_{166} - \zeta_{414} - \zeta_{456} \\ \zeta_{122} \\ \zeta_{123} + \frac{1}{2}\zeta_{144} \end{bmatrix} \tag{3.5}$$

As a result, Equation 3.4 reduces to:

$$I_i = 1 + \sum_{j=1}^{2} A'_{ij} \zeta_j^{\mathrm{I}\prime} X + \sum_{k=1}^{4} B'_{ik} \zeta_k^{\mathrm{II}\prime} X^2, \quad i \in \{1..6\} \tag{3.6}$$

where the matrix A' consists of the first two columns of A, and B' of the first four columns of B. They are equal to:

$$A' = \frac{1}{6} \begin{bmatrix} 0 & 6 \\ 0 & 6 \\ 2 & 4 \\ 3 & 3 \\ 2 & 4 \\ 2 & 4 \end{bmatrix}; \qquad B' = \frac{1}{36} \begin{bmatrix} 0 & 0 & 36 & 0 \\ 0 & 0 & 18 & 18 \\ 4 & 16 & 8 & 8 \\ 9 & 18 & 9 & 0 \\ 6 & 12 & 12 & 6 \\ 6 & 12 & 12 & 6 \end{bmatrix} \tag{3.7}$$

The unknown piezocoefficients can be extracted if measurements yield a series of N different data points $(X, A'_{ij}, B'_{ik}, I_i)$. In that case Equation 3.6 expands to a system of N linear equations in six unknowns. This system can be solved by a least-squares method, such as the one described in Section 3.3.3.

The extraction can be made more accurate by correcting the data points for the geometrical effect. This effect concerns the stress-induced changes in the base width and the junction area, as discussed in Section 2.4.3. The changes can be calculated with the aid of Equation 2.136, the orientation of the current and the stress, and the values of the compliance coefficients (see Appendix A). Their effect on the current changes is generally small: in the order of 0.3% for the stress domain considered.

Resistors

Although the experiments are mainly concerned with transistors, also a set of resistors was characterized under stress. This allows the extraction of the first-order piezoresistive coefficients of p-type silicon. These coefficients can be compared with those found in literature, which enables a benchmarking of the measurement set-up.

Like the transistor, the stress-induced changes in the resistor can be described by the independent piezoresistive coefficients. These changes, however, are much more linear than those in the transistor collector current. When geometrical effects are neglected and the voltage difference is fixed, the changes in the resistance are given by:

$$\frac{R(X)}{R(0)} = 1 + \sum_{i=1}^{3} c_i \pi_i X \tag{3.8}$$

where π is the vector with the piezoresistive coefficients π_{11}, π_{12}, and π_{44}. The c_i are again constants depending on the orientation of the stress and the current.

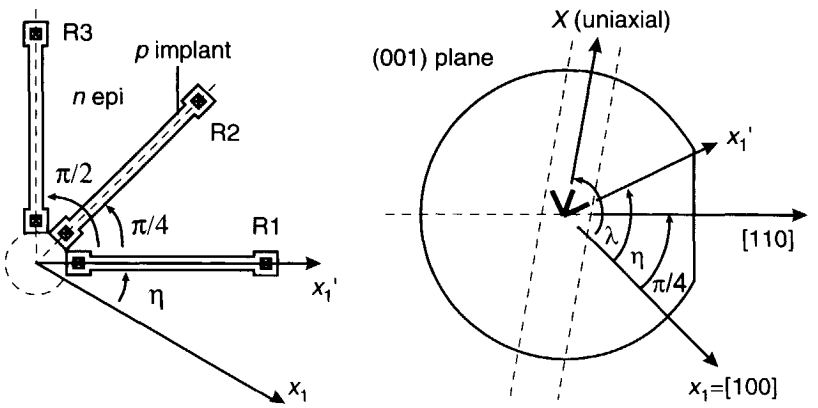

Figure 3.2: Lay-out of the three-element rosette. *Left:* orientation of the resistors with respect to each other. *Right:* Orientation of the rosette with respect to the wafer and the crystal axes.

The piezocoefficients can be found from measurements if at least three different orientations of stress and current are considered. This is possible by making a rosette of three resistors in the (001) crystal plane and subjecting them to one uniaxial stress in that plane, as shown in Figure 3.2 [2, 3, 4]. The advantage of such a rosette is that three current directions can be characterized at the same moment, with closely matched resistors, and at almost the same temperature. If only first-order coefficients are required, measurements at only one stress level are in principle sufficient.

The measured normalised resistances R_i are linear combinations of the π_j. Equation 3.8 can therefore be extended to the system:

$$R_i = 1 + C_{ij}\pi_j X, \qquad i,j \in \{1,2,3\} \tag{3.9}$$

This system can be solved again by collecting a series of measured points $(X, C_{ij}, R_i(X))$ and applying a least-squares method as in Section 3.3.3. It can also be corrected for geometrical changes. If only one stress value is available, it can also be solved by inverting the system:

$$\pi_j = C_{ij}^{-1}(R_i - 1)X^{-1} \tag{3.10}$$

This has the disadvantage that measurement errors are not averaged and that nonlinearities pass unnoticed. However, it calculates the π_j very quickly and is easily programmed.

The system can only be inverted, of course, if C is invertible. It can be proved that for the rosette of Figure 3.2 this is always possible, unless the uniaxial stress is directed along a $\langle 100 \rangle$ or a $\langle 110 \rangle$ direction [5]. This is valid for any orientation of the rosette in the (001) plane. For this reason, the stress in the

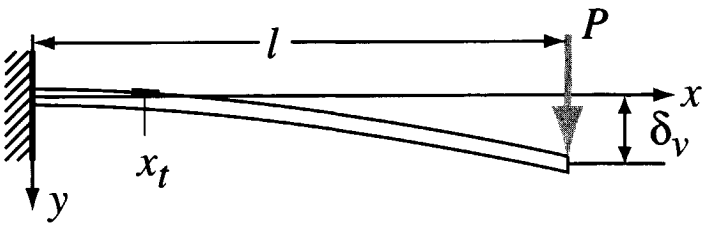

Figure 3.3: Principle of bending a cantilever beam. The beam is clamped on the left-hand side and deflected at the right-hand side over a distance δ_v. This requires a force P. The transistors and resistors are located at the surface of the beam at the point x_t.

resistor experiments is chosen along the [130] direction. The rosette orientation is chosen along the [100] direction (thus $\eta = 0$) to minimise the propagation of measurement errors [5]. In that case, C and C^{-1} are given by:

$$C = \frac{1}{10}\begin{bmatrix} 1 & 9 & 0 \\ 5 & 5 & 3 \\ 9 & 1 & 0 \end{bmatrix}, \qquad C^{-1} = \frac{1}{24}\begin{bmatrix} -3 & 0 & 27 \\ 27 & 0 & -3 \\ -40 & 80 & -40 \end{bmatrix} \tag{3.11}$$

Rosettes along $\eta = 63°$ and $\eta = 27°$ are added to verify the principle.

3.1.2 Stress generation

The transistors were subject to mechanical stress by integrating them into a beam and bending the beam as a cantilever. This method is illustrated in Figure 3.3. It is well-known and represents a number of advantages with respect to other methods, such as four-point bending, needle pressure, uniaxial pressure, or hydrostatic pressure:

- The stress domain is interesting: it ranges from -250 MPa to +250 MPa, and can be compressive as well as tensile by changing the direction of deflection;
- The stress orientation can be varied by cutting the beam in different ways from the silicon crystal. In this manner both normal and shear stresses can be generated (see Figure 3.4);
- The stress in the transistors is uniquely determined by the bending. This is the result of the cantilever being a determinate structure [6]. In addition, it is possible because the clamp is at some distance from the transistors. This means that stresses from the clamp itself are negligible by Saint-Venant's principle [6];

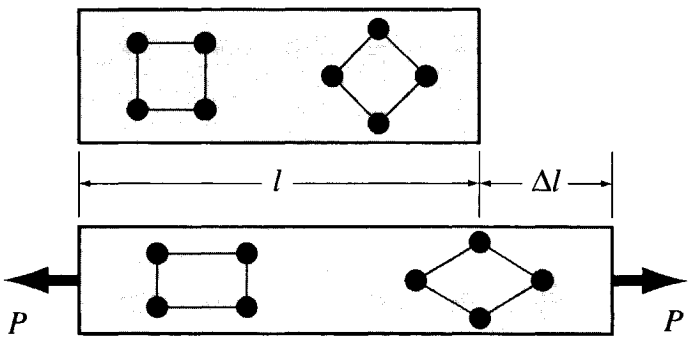

Figure 3.4: Variation of the stress orientation by cutting the beam from a crystal under different angles with respect to the crystal axes. *Above:* Situations before stressing. The long sides of the beam are either cut along the shortest vector between two atoms, or under an oblique angle. *Below:* Situation with uniaxial stress along the beam. The stress changes the interatomic distances (normal stress), but can also distort the angles (shear stress).

- The stress is high where the deflection is small. This means that the stressed transistors can be connected electrically with bonding wires, which only support small deflections.

The bending configuration has the disadvantage that it can only generate stresses in the plane of the beam. Another disadvantage of bending is that high, compressive stresses are impossible. The reason is that bending simultanuously generates compressive and tensile stresses of the same magnitude. This limits the magnitude because tensile stress in silicon lead much faster to fracture than compressive stress.

The stress in a thin cantilever beam is principally caused by the bending moment M resulting from a deflection δ_v with a force P at the tip. It is directed along the length of the beam. As sketched in the cross section of Figure 3.5, it is highest at the surface, zero inside, and opposite of sign at the opposite surface. It is zero at the tip, but increases linearly along the beam and reaches a maximum at the clamp. This can be expressed in the equation [6]:

$$X = \frac{3}{2}\frac{Ew(l - x_t)}{l^3}\delta_v \tag{3.12}$$

in which X is the stress magnitude at the upper surface of the beam at a distance x from the clamp, E is Young's modulus along the stress direction, w the beam thickness, and l the length. E is a function of the compliance coefficients S_{ijkl} and the direction vector of the beam with respect to the crystal axes q_i; see Appendix A. This can be written as the tensor sum:

$$\frac{1}{E} = q_i q_j q_k q_l S_{ijkl} \tag{3.13}$$

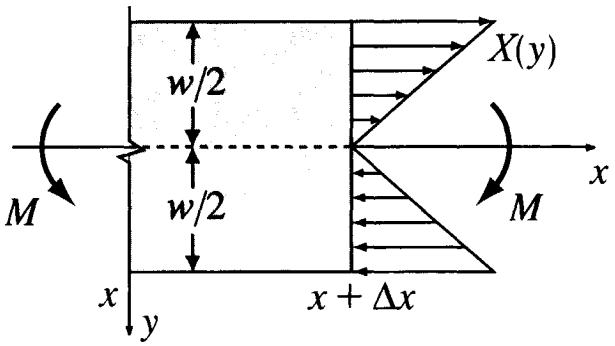

Figure 3.5: Distribution of stress $X(y)$ through a cross section of the cantilever beam for an infenitesimal beam segment between x and $x + \Delta x$. The stress is generated by the bending moment M.

The above beam equation is inaccurate in two respects. Firstly, it is the linear approach of a more general expression which cannot be written in a closed form. It can be quite accurate, however, if the deflection is small with respect to the length. In the experiments described here the deflection/length ratio did not exceed 7%, resulting in an error of less than 1% [6]. Secondly, the equation neglects shear stresses. In the present case, however, this modifies the results with less than 0.1% because the beams are long [6, 7].

3.1.3 Electrical characterization

The electrical measurements should determine the stress-induced variations in the current-voltage characteristics of the devices. These characteristics should be measured in forward bias, which, in the case of bipolar transistors, covers a current domain of many decades. They should also be measured in DC conditions. For transistors the main characteristic is determined by the saturation current I_S. This parameter should therefore be extracted from the measurements with sufficient accuracy. Other parameters are less important and include the DC current gain β and the Early voltage V_A. They should be considered, but only qualitatively. For resistors, the main characteristic is, of course, the resistance.

Bipolar transistors

The transistors were characterised according to the following strategy. First, it was determined where the exponential model of the collector current I_c as a function of the base-emitter voltage V_{be} was the most valid. Then V_{be} was fixed and I_c was measured as a function of stress.

For this purpose the transistors were connected to source-measurement units (SMUs) according to the scheme of Figure 3.6. The V_{be} was swept between 0.3

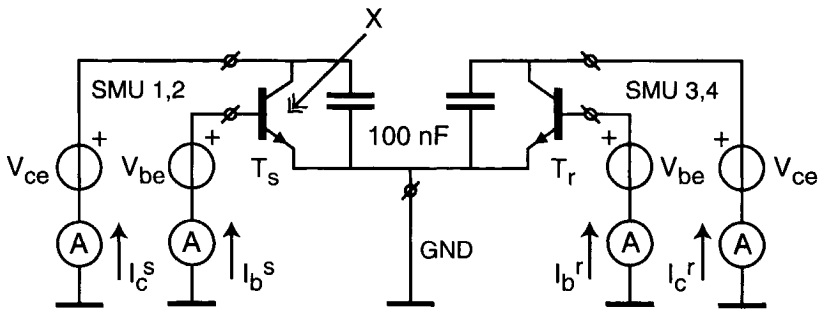

Figure 3.6: Electrical connections to measure the bipolar transistor T_s under a stress X. The transistor T_r serves as a reference. The voltages of the base and the collector are set, while the currents are measured by the source-measurement units (SMUs). The voltage of the substrates is set with a separate voltage source which is not shown.

to 0.8 V to measure I_c and the base current I_b from the lowest possible value up to the high-injection region. This was repeated for different collector-emitter voltages V_{ce} and also for different stress magnitudes X.

The advantage of biasing through V_{be} is that both V_{be} and V_{ce} can be varied with the SMUs, and that both I_c and I_b can be measured. This would be complicated if the transistors are biased through I_c because the SMUs have always one terminal connected to ground internally. Biasing through V_{be} has the disadvantage that it is relatively inaccurate. Any error in V_{be} is amplified forty times when it reaches I_c. This error, however, can be neglected for a number of reasons. Firstly, it does not influence the stress sensitivity, which is independent of I_c over many decades. Secondly, it is partially suppressed by making differential measurements. Thirdly, it results in an offset which is smaller than the stress-induced changes, and which can be removed mathematically. Finally, its noise and drift components appear to be small compared to the measurement resolution.

The differential measurement of Figure 3.6 does not only reduce noise, but is also very effective in suppressing the effects of thermal drift. During the measurements the temperature of the transistor changes by variations in the room temperature and self-heating. Even if those changes are a fraction of a degree they influence the collector current by the same amount as the mechanical stress [8, 9]. This can be compensated by simultaneous measurements on an unstressed reference transistor which is thermally coupled to the stressed transistor. Differences between the transistors result in an offset, but this can be filtered out mathematically.

The capacitors in Figure 3.6 serve to avoid oscillations caused by parasitic impedances in the connecting wires. In addition, they suppress noise injected on the high-ohmic collector terminals.

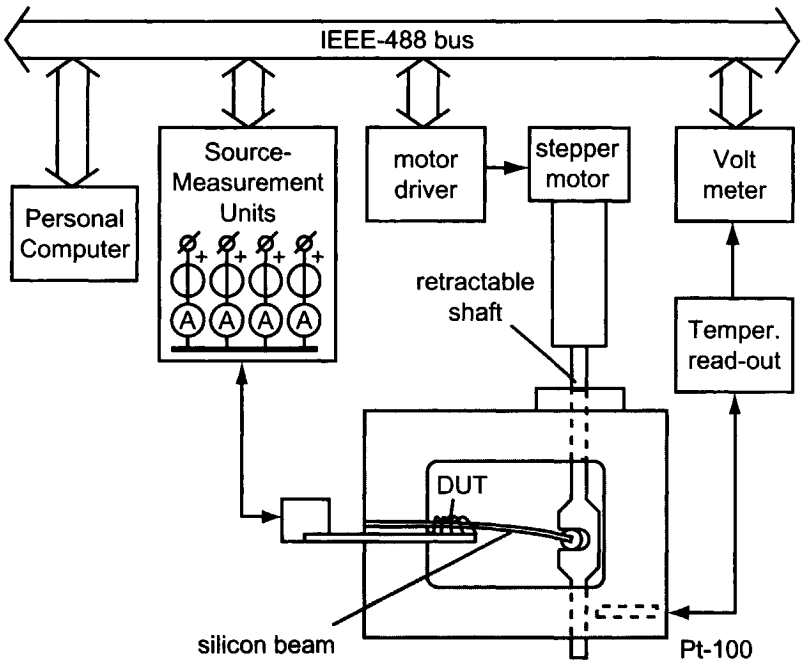

Figure 3.7: Schematic of the total measurement set-up for characterising transistors and resistors while bending them.

Resistors

The resistors were characterised by following a similar strategy. First, they were measured over a range of bias voltages. Then the voltage was kept constant and the current variations were measured as a function of stress.

These measurements had a much lower sensitivity to voltage and temperature fluctuations than the transistor measurements. Nevertheless, they were also performed with unstressed reference devices to improve the accuracy.

3.2 Implementation of the measurement set-up

As shown in the previous section, the measurement set-up requires various components. It is discussed here how they can be implemented in practice, which are their limitations, and how they are connected in the total system.

3.2.1 Overview

The total measurement set-up is is schematically drawn in Figure 3.7, and shown in Figure 3.8. Its core is formed by a device under test (DUT), which is inte-

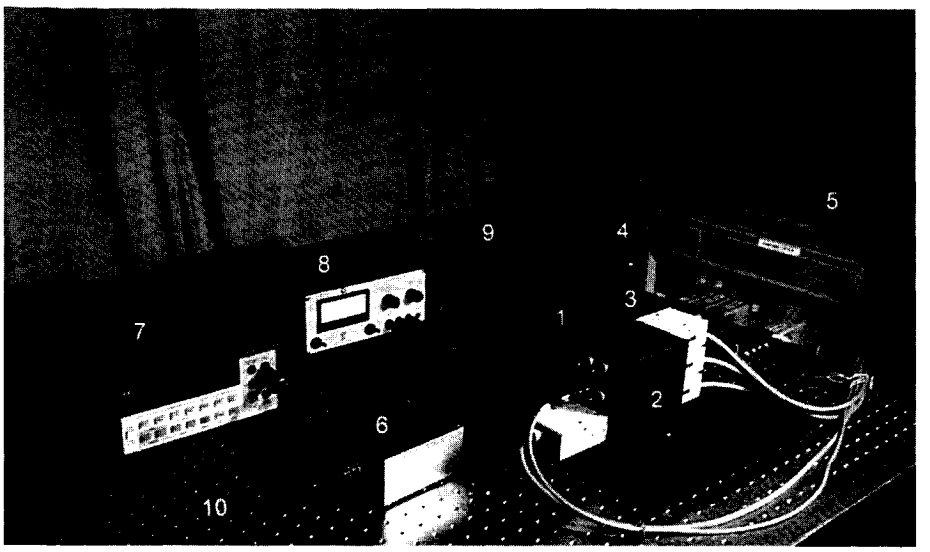

Figure 3.8: Total measurement set-up. 1) Block with beams inside, 2) Box with electrical connections, 3) Stepper motor, 4) Motor driver, 5) SMUs, 6) Temperature read-out, 7) Voltmeter, 8) Power supply of temperature read-out, 9) Four storage boxes with devices, 10) Optical table for mechanical stability.

grated in a silicon beam. One end of the beam is clamped in a hollow aluminium block. The other end is deflected by a shaft, which can be retracted by a stepper motor. The devices are biased and measured electrically by source-measurement units. The temperature of the block is measured by a platinum Pt-100 resistor. This resistor is biased and linearised by a read-out circuit, which transfers its signal to a voltmeter.

All parts of the set-up are controlled by a personal computer through an IEEE-488 bus. The control software has been written HP-VEE.

As mentioned above, the measurements were differential. This means that the stressed transistor was characterised in the same time as an unstressed reference transistor. In this way it was possible to suppress noise and compensate for thermal drift, while introducing only a minor offset. As a result, it was not necessary to apply techniques such as synchronous detection. It should be noted that the reference device is not shown in Figure 3.7.

3.2.2 Devices under test

The silicon devices were designed according to standard layouts. It was sometimes necessary, however, to position them under an oblique angle with respect to the wafer flat. In addition, they had to be accompanied by special marks to saw the wafers into beams along other oblique angles.

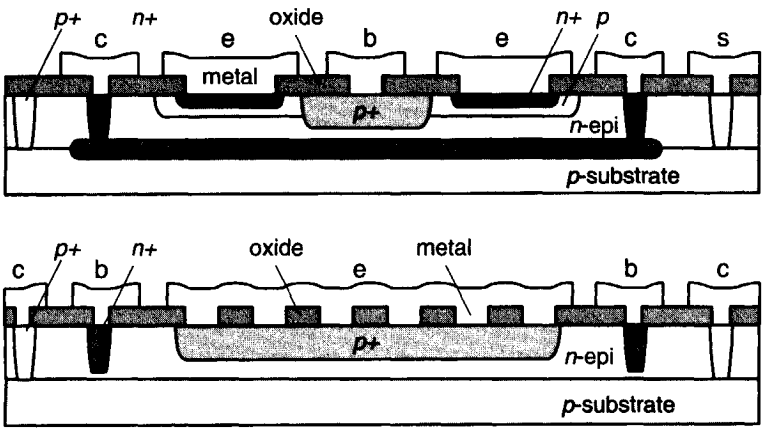

Figure 3.9: Schematic cross sections of the characterized transistors. *Above:* vertical *npn*. *Below:* vertical substrate *pnp*.

Lay-out of the transistors

The transistors used are both of the *npn* and the *pnp* types. They were designed with the masks available in the DIMES01 process. This is a standard bipolar process of the Delft University of Technology using 4-inch wafers with a thickness of 525 μm. (see Reference [10] and Appendix D). The cross sections of the devices are shown in Figure 3.9. The *npn* transistor is based on the standard BW type, which has a base with a width of 0.1 μm and a doping of 3×10^{17} cm^{-3}. The *pnp* transistor is a new design. It uses the substrate as the collector, the $p+$ WP implantation as the emitter and the 3.5 μm epitaxial layer as the base. This results in a base width of 2 μm and a base doping of 6×10^{15} cm^{-3}. In both transistor types the doping of the base is low enough for assuming that the material is nondegenerate.

The collector current should be driven through the base in a well-defined direction. This is achieved by making the transistors vertical, and avoiding edge effects and current spreading as much as possible. The dimensions of the emitter were therefore chosen large compared to the base width.

The emitter dimensions are restrained, however, by the resistance of the base layer. This parasitic series resistance causes current crowding: a voltage drop between the base region at the periphery of the emitter and the regions further away, right beneath the emitter implantation. As a result, the periphery experiences a higher base-emitter voltage and therefore generates a larger collector current density. At large base currents the influence of the periphery becomes dominant and deteriorates the transistor performance.

A compromise was found by choosing emitter strips of 5×100 μm for the *npns* and 20×100 μm for the *pnps*. In that case the base resistance reduces

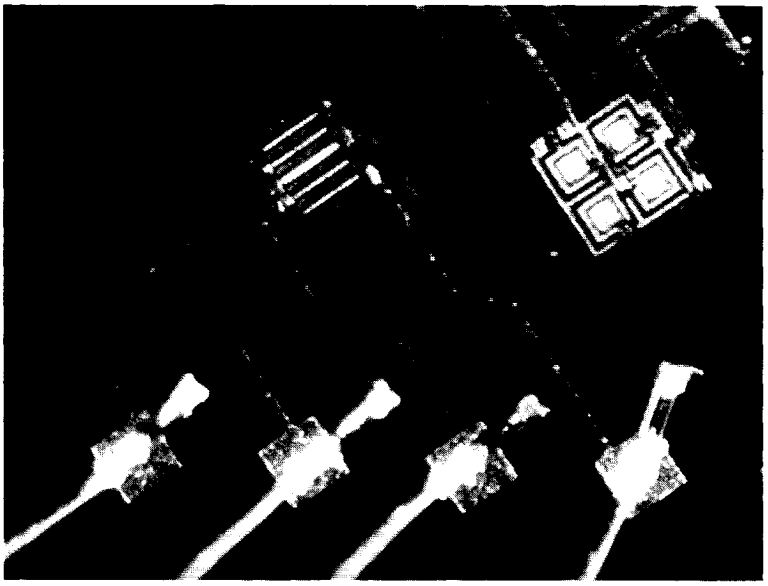

Figure 3.10: Photograph of a vertical *npn* on an oblique beam. It is connected with metal lines and bond pads in standard orientations.

the current density with 50% at maximum when the total current reaches the maximum value allowed in the bond wires (10 mA). The current density in that case is equal to 500 A cm^{-2}, which is five times below the density where high injection starts in the standard BW transistor.

The chosen emitter dimensions lead to a ratio between the junction width and the base width of 50 for the *npn* transistors, and 10 for the *pnps*. The latter ratio is quite low. Unfortunately, the resulting edge effects appear to be the major source of uncertainty in the extraction of the piezojunction coefficients from the *pnps*.

The transistors are assumed to be free of process-induced stress. This stress certainly occurs on the interface of the oxide and the silicon. It has been shown, however, that it is much lower in the silicon than in the oxide, and small compared to the stress applied through the bending [7].

Lay-out of the resistors

The resistors were given a standard lay-out, as shown in Figure 3.2. They were designed as a *p*-type implantation in the epitaxial layer, which is the same BW implantation as used for the intrinsic base of the *npn* transistors. It was expected, however, that the peak concentration of this implantation would make the material slightly degenerate near the surface.

Therefore some pinch resistors were added. They were designed with the same implantation, but with a shallow $n+$ emitter layer on top. In this way only the tail of the p-type implantation is conducting and has the same well-defined doping level as the base of the npn transistors. It was estimated that the cross effects between pinching and the piezoresistance are small. These effects could consist of a stress-induced increase in the deplation layer width, which would change the cross section of the resistor and influence its resistance.

The resistors were given such dimensions that they could be modelled conveniently. To be sure of the current direction, the resistors were made rectangular, without curves. They were made long enough to neglect the resistance of the clubheads and the connections. The length was kept smaller than 100 μm for reasons of available chip area. The width was maximised within those limits to improve the matching between the resistors. As a result, the BW resistors were designed to have a value of 1.6 kΩ, and the pinch resistors 55 kΩ.

The resistors were grouped in three-element rosettes, as shown in Figure 3.2. They were positioned on $\{100\}$ wafers with the first resistor directed along the [100] direction ($\eta = 0$ rad).

Wiring and bond pads

Although some devices were positioned in nonstandard orientations, this had to be avoided as much as possible. This was necessary for the pattern generator, which severely slows down on oblique mask patterns. It was therefore decided to use standard orientations for metal lines and bond pads. This resulted in many zig-zag patterns, such as shown in Figure 3.10.

Beam design

The silicon devices had to be part of a cantilever beam, which posed two problems to the mask design.

Firstly, the beams had to be much longer than the standard reticle of 1×1 cm. This follows from the requirements in Section 3.1.2 that the deflection should be small compared to the beam length, but that it should be accurately known at the same time. The beam length was therefore set at 66.5 mm, with a free end of 37.5 mm. This end was deflected up to 2.50 mm. The beam width was chosen between 1.4 mm and 1.5 mm and adjusted to fit a whole number of beams into one reticle. The thickness was equal to thickness of the wafer: 0.525 mm. The devices were positioned on the beam at a distance between 4 mm and 6 mm from the clamp.

It was not possible to design such beams with full-wafer masks. Instead, the beams were composed by repeating a reticle with the standard size of 1×1 cm. On this reticle, the boundaries of the beams were indicated with metal markers. These markers served to align the sawing machine, after which the length, width, and repeat count of the beams could be programmed. Other markers were added to position the beam with respect to the clamp.

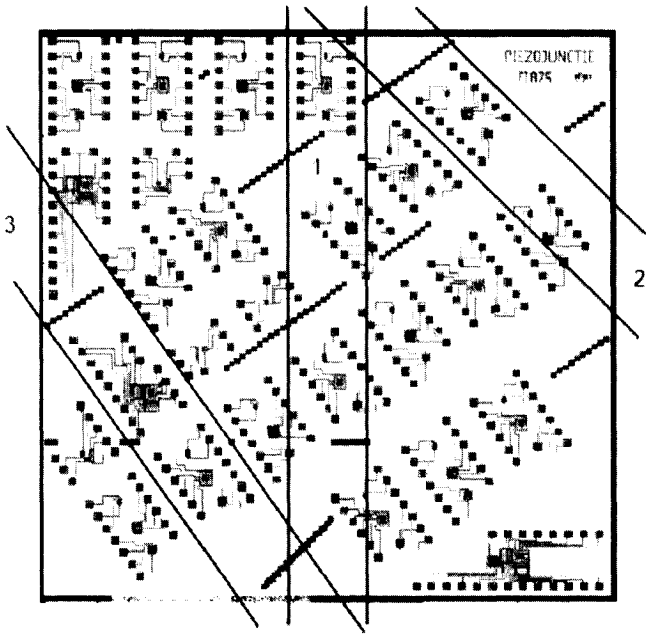

Figure 3.11: Mask set of the reticle with transistors. The reticle can be sawn into beams in three ways, indicated by the lines and the numbers. Compare with Figure 3.1.

Secondly, the beams had to be directed in other orientations than the standard sawing lanes. These directions were shown already in Figure 3.1. It was impossible, however, to rotate the masks with respect to the wafer flat. As a solution, the mask set was made suitable for three different beam orientations. In this way all wafers could be processed in the same way, whereas the sawing direction could be chosen for each wafer separately. The result is shown in Figure 3.11.

The orientation of the beams in the crystal is known only within certain limits, which are determined by the tolerance of the wafers. These should be specified by the supplier. For the case of {100} and {111} wafers they are prescribed by the SEMI M1 standard. This standard allows for a misorientation of $\pm 1°$. This is far too small to be a significant source of error in the present measurements.

Processing

The transistors and resistors were fabricated in the DIMES01 process, which is a standard bipolar process of the Delft University of Technology [10]. The fabrication was slightly complicated because {110} and {111} wafers were used

in combination with standard {100} wafers.

In the first place this led to different oxidation rates and therefore to modified oxide thicknesses. As a result, the alignment marks were invisible to the wafer stepper and had to be refabricated. In addition, the oxide after the anneal/oxidation step was too thick and had to be etched back by 30%.

In the second place the different wafer orientations led to variations in the thickness of the epitaxial layer. This was corrected by adjusting the deposition time.

Finally, the nonstandard orientations led to sheet resistances which were slightly out of the specifications. The sheet resistance of some pinch resistors, for instance was 40% too low. Probably, this was caused by the dependence of the diffusion speed on the crystal orientation.

All wafers, nevertheless, yielded a large number of working devices. These included test transistors with gains between 57 and 92 for the *npn* transistors, and gains between 63 and 103 for the *pnp* transistors. The base doping of the *npn* devices was determined by C-V measurements. It yielded concentrations of $(3 \pm 1) \times 10^{17}$ cm^{-3} and a quite flat doping profile compared to the general transistor profile depicted in Figure D.3. It also yielded base widths between 85 and 130 nm. The base doping of the *npn* transistors was calculated from the epilayer sheet resistance and thickness and yielded a concentration of 6×10^{15} cm^{-3}.

3.2.3 Beam carrier

The silicon beams had to be placed on a carrier in which they can be clamped mechanically and connected electrically. In addition, they had to be protected from accidental touching. These requirements led to the beam carrier design shown in Figure 3.12 and 3.13.

Clamping

The beams were clamped between a stack of epoxy-resin plates reinforced with glass fibre, the same material as used for printed-circuit boards. The beams were aligned in long slots that were milled to a depth of 0.50 mm. This is slightly below the wafer thickness. As a result, the beams slightly emerge from the surface of the plate and can be clamped more strongly by external pressure on the stack.

The stack was glued together and screwed to an aluminium protection cap. This cap covers the fragile area where the beams are wire bonded. In addition, it protects the beam ends when they are posed on a table by its two protuding fingers.

On the free tip of the beams a small PVC cylinder was placed of 3.0 mm in diameter. This cylinder fits into the notch within the retractable shaft used to deflect the beam, as shown in Figure 3.7. It allows the tip to be deflected while

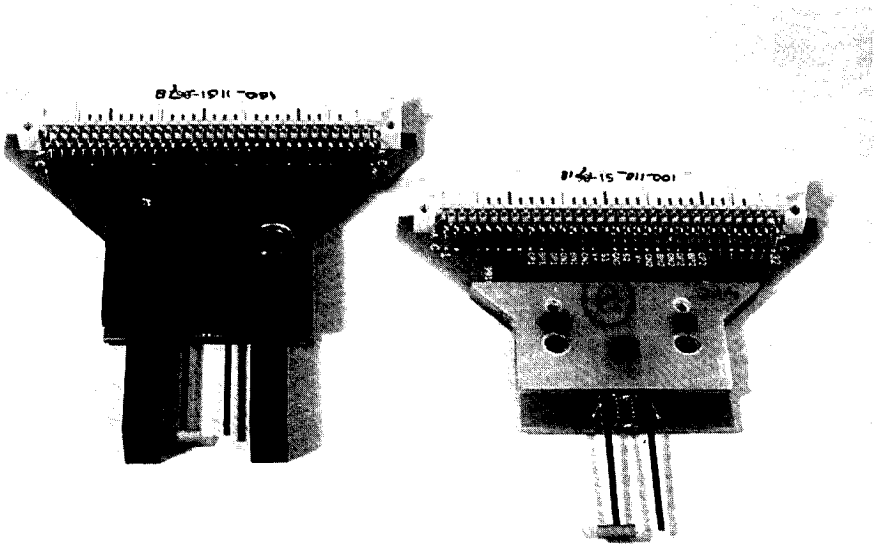

Figure 3.12: Assembled beam carriers with and without black aluminium protection cap. The beams with the cylinder on the tip are bent whereas the others serve as a reference.

its vertical distance to the shaft remains constant. It may move freely, however, in the horizontal direction.

For an optimum accuracy the thickness of the beams was determined by measuring spare beams of each wafer with a micrometre screw. The resulting values have an error interval of about 5 μm, which does not lead to a dominant error in the measurement results. In addition, the free length of each beam was controlled by aligning the marks on the beam with the edge of the epoxy plates. If this is done under a microscope, it is possible to obtain an error of less than 0.1 mm on a total length of 37.50 mm.

Although the above method of clamping worked, a number of problems remained. The most important was that the beams could be fixed in a slightly tilted position with respect to their longitudinal axis. As a result, the deflection of the tip also caused torsion. This torsion could be estimated by measuring the play between the cylinder and the notch. It amounted to up to 21 MPa for a tilt angle of 1.1° and was thereby a significant source of error. Another problem was that one epoxy plate is covered with an anti-soldering mask. This prevents a good adhesion of many types of glue. It was solved by using Loctite 4105. This is a toughened cyanoacrilate-based adhesive which is designed to bond large components on epoxy printed-circuit boards. It sometimes occured

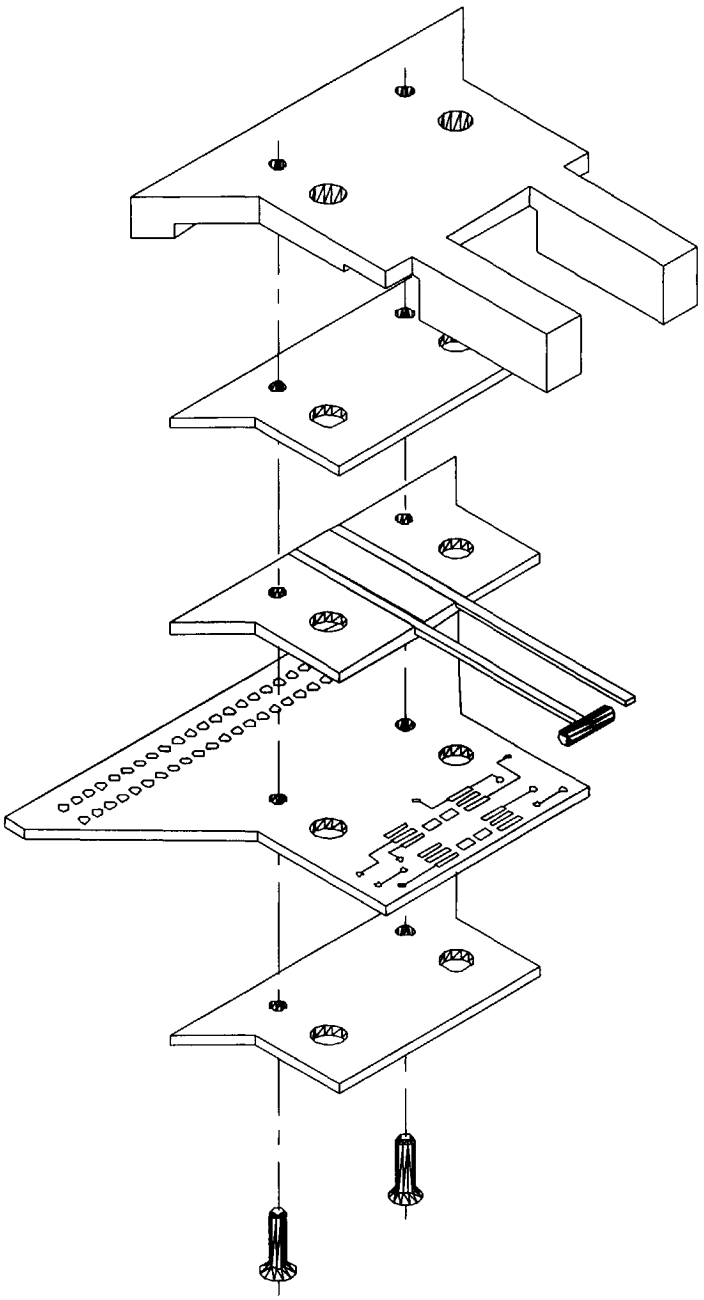

Figure 3.13: Exploded view of the beam carrier. The two silicon beams are clamped in a stack of fiber-reinforced epoxy plates. One of the plates is a printed-circuit board containing the electrical connections and some electronic parts. The beams are protected by an aluminium cap with two fingers.

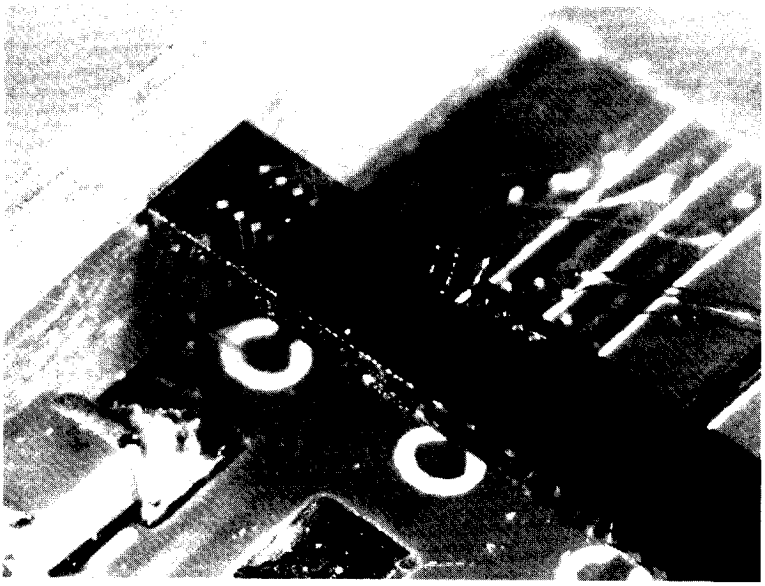

Figure 3.14: Clamping of a beam and wire-bonding to the PCB. The beam floats 1 mm above the PCB surface. The latter contains gilded bonding pads (top, right) and decoupling capacitors (bottom, left).

that some slots were too deep, meaning that the top of the beams were not supported. This problem revealed itself rapidly through the strange shape of the measured curves. Finally, the clamping between the epoxy plates could not be completely rigid because the elastic modulus of this material is about ten times inferior to that of silicon. However, the influence of this effect could not be observed when the clamping was prestressed by tightening the screws.

Printed circuit board

The stack clamping the beams also contains a printed-circuit board (PCB) which connects the devices electrically. For this purpose the board protrudes from the stack one millimetre below the beams. This protruding end is 1 centimetre long and contains pads to wire-bond the devices. It is shown if Figure 3.12 and 3.14. The end also hosts the decoupling capacitors of Figure 3.6. The opposite end of the PCB contains a DIN 41612 connector with 64 gilded pins.

The PCBs were fabricated by the printed-circuits group of the Delft University of Technology. They consist of a 1.6 mm epoxy core, supporting copper traces on both sides with a thickness of 18 μm. The traces were gilded on the place of the wire-bonding to obtain a good adhesion. They were plated with a lead-tin alloy where components should be soldered. On the remaining places they were covered with an anti-soldering mask.

Initially, the PCBs suffered from some problems, which were solved in a new design. The most important was that they were too difficult to wire-bond, which severely decreased the yield. This was partially due to the large number of the bonding pads, 88. It was also caused by the small width of the pads of 0.2 mm, with a spacing of another 0.2 mm. As a result, the loose ends of the bonding wires easily caused short circuits. The problem was solved by reducing the number of the pads to 14, and by doubling the dimensions of the pads.

Another problem was oscillation of the transistor circuit due to parasitic impedances of the connecting wires. For suppressing them a number of measures was taken on the PCB:

- The width of the copper traces was doubled, and the distance between them was increased.
- Decoupling capacitors were added on high-impedance points close to the bonding pads.
- Placeholders were created to add pre-amplifyers and more capacitors. This, however, appeared to be unnecessary in the end.

Finally, it was a problem that the measurement set-up evolved, whereas the connections on PCB could hardly be changed. Their flexibility was greatly enhanced, however, by wiring each bonding pad directly to the DIN connector. In this way each pad could be addressed separately. Moreover, these wires were connected to soldering pads to add new links and electronic parts if necessary.

Wire bonding

The silicon devices were connected to the PCB by wire-bonding. This connection is flexible enough to follow the movements of the beam, whereas it does not introduce additional stress. The disadvantage of the wire-bonding was its difficult realisation, which made it the decisive step for the yield.

The first difficulty was that the devices had to be supported during the bonding. This was required to counterbalance the force of the wedge bonding machine and to avoid that its ultrasonic vibration led to resonance. The beams were therefore supported by a specially designed fork, which fills the gap between the beams and the PCB. Often, however, this was not enough because the width of the gap was slightly different for each device. This was solved by increasing the thickness of the fork teeth with pieces of adhesive tape.

The second difficulty were short circuits caused by bonding wires which were touching the edge of the silicon beam. This resulted from the position of the silicon with respect to the bonding pads of the package. In a normal package, the silicon lies below the pads, but in this case it lied 1.5 mm above the pads. The wires thereby arrived at the silicon surface in an almost horizontal direction and often touched the edge. Once this was recognised, it could be solved by inverting the bonding order. This means that now, the wire was firstly attached to the silicon and then to the PCB. In this way the wire could be curved around the edge of the silicon, as shown in Figure 3.14.

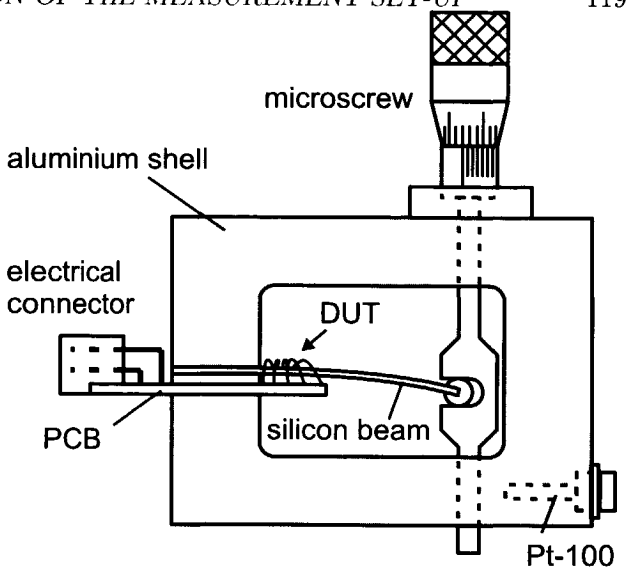

Figure 3.15: Schematic view of the bending apparatus.

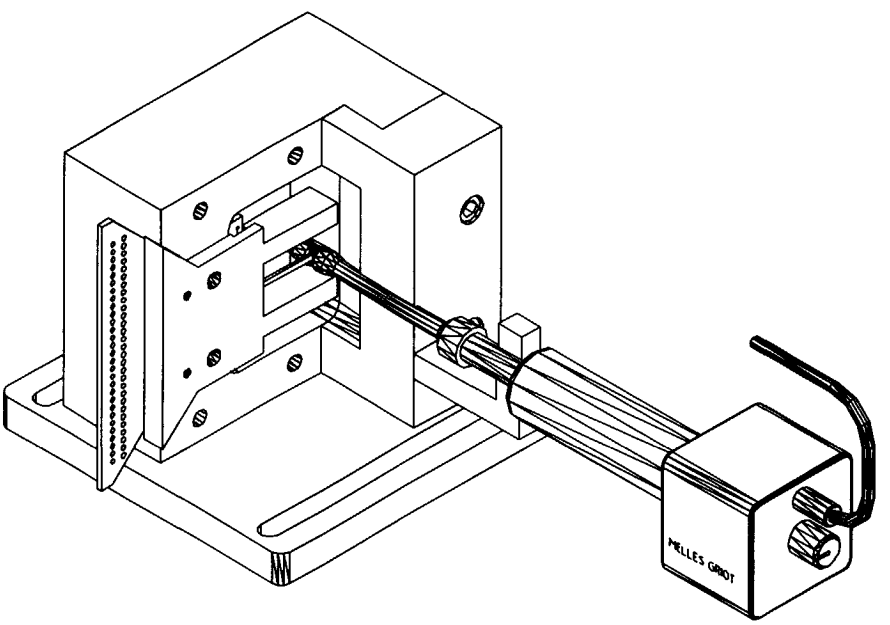

Figure 3.16: Cross section of the bending apparatus showing the position of the beam carrier and the automatic microscrew.

3.2.4 Bending apparatus

The beam carriers were placed in a bending apparatus which was specially designed to deflect the beams in a very controlled manner. This apparatus should therefore be stiff. In addition, it should protect the devices from temperature variations, light, and electric and magnetic interference.

For those reasons the apparatus was based on a thick aluminium block with a cavity inside. Its principle is sketched again in Figure 3.15, whereas its actual construction is shown in Figure 3.16 and 3.17. The block measures $11 \times 10 \times 9$ cm. Inside its wall the beam carriers are clamped. The tips of the beams are deflected by a microscrew.

Clamping of the beam carriers

A beam carrier can be mounted in the block by removing the clamp and inserting the carrier into a special cavity. The shape of the cavity is tapered, which prevents the beams from breaking by being shifted too far into the block. The carrier is attached with two M4-screws. It is covered by the clamp, which is constructed in such way that the force of its M6-screws is transferred to the clamped ends of the silicon beams. It was verified that this force did not influence the measurements.

Retractable shaft

The tip of a silicon beam can be deflected by the retractable shaft. This shaft has a cavity of 3 mm in diameter to grip the small cylinder fixed on the beam tip (see Figure 3.15 and the previous section). The grip has a play of, typically, 0.11 mm. It was determined for each beam separately by moving the shaft around the zero-deflection point and measuring the current changes in the silicon devices. This yields curves with a 'dead move', such as shown in Figure 3.18. The move could be determined with an accuracy of 20 μm.

The movement of the shaft is controlled by a microscrew and ranges from -2.5 mm to +2.5 mm. In a first approach, the screw was a hand-driven Mitutoyo micrometre with a resolution of 1 μm. Later, it was replaced by an automatic Melles-Griot Nanomover II (see Figure 3.7). This system contains a microscrew driven by a stepper motor. It has an accuracy of 1 μm and a repeatability of 0.1 μm, which is largely sufficient in view of the other sources of uncertainty.

The retractable shaft had to be connected to the shaft of the micrometre screw, but was not allowed to rotate with the latter. This was solved by separating the shafts with a steel ball. They were pressed against each other by a spring.

Thermal stability

The transistor measurements had to be performed in an environment with a high thermal stability. This became clear after some measurements in which

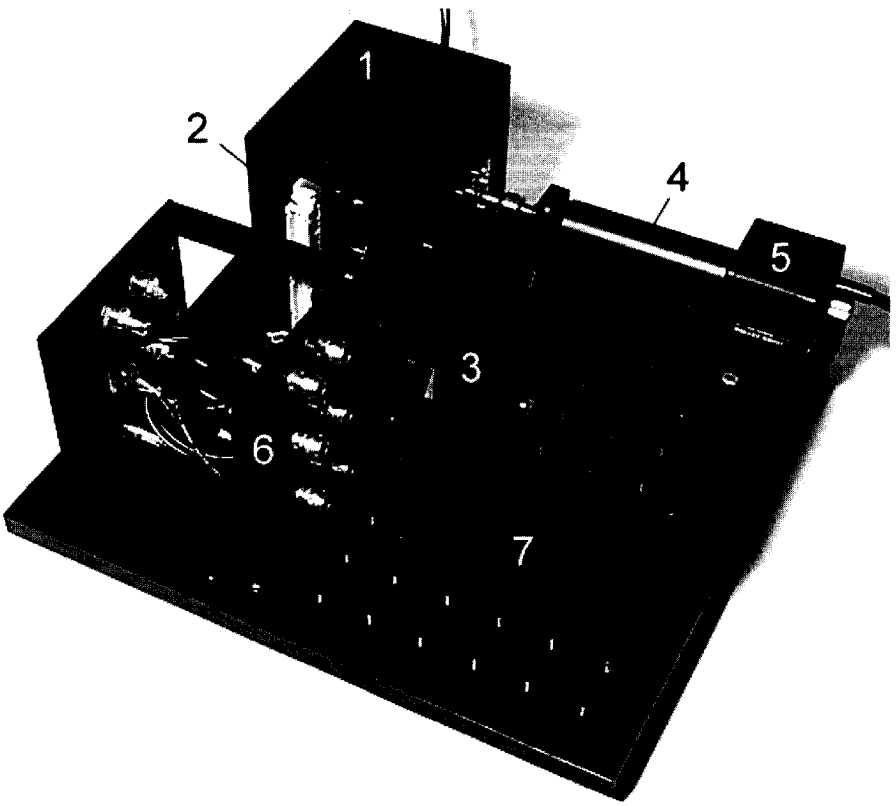

Figure 3.17: Bending apparatus with several parts opened. 1) Aluminium block, 2) Beam carrier, 3) Clamp, 4) Microscrew, 5) Stepper motor, 6) Box with triaxial connectors, 7) Lid of the connector box.

the collector currents were as much changed by the ambient temperature as by the mechanical stress. Such a measurement is shown in Figure 3.19.

For this reason the principle of differential measurement was applied, which was already discussed in Section 3.1.3. In this way the stress sensitivity can be separated from the temperature sensitivity. The temperature variations should nevertheless be slow and remain within certain limits to avoid the effects of cross-sensitivity. Recent work by Fabiano Fruett has shown that the stress sensitivity slightly decreases with a temperature increase ΔT, with a factor $\exp(\gamma \Delta T)$ in the range from -10°C to 110°C [11, 12]. The constant γ was found to be $-1.8 \times 10^{-3}\ °C^{-1}$ for the orientation considered. It therefore changes the results with less than 1% even when the temperature changes 5°C.

The differential measurement requires a tight thermal coupling between the stressed and the reference devices, whereas the slow temperature variation requires a large heat capacity. These properties were indeed obtained by placing

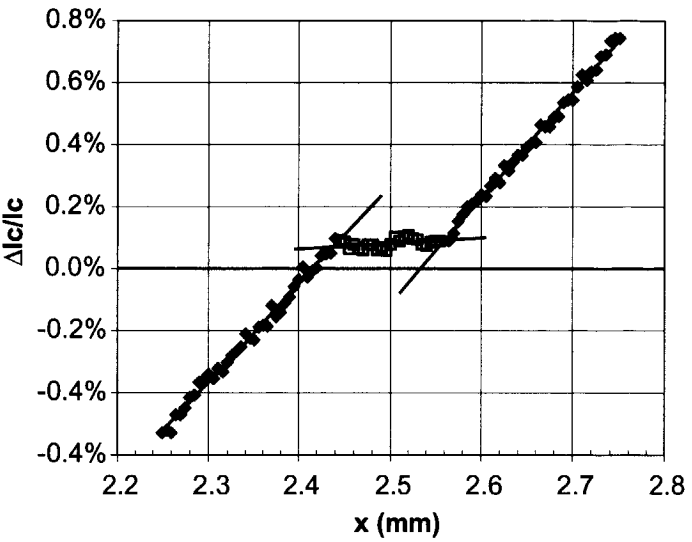

Figure 3.18: Measurement of a typical play between the beam tip and the retractable shaft. The shaft is moved around the zero-bending point of the beam. At the same time, the changes are measured of the collector current of a transistor on the beam.

the devices in the cavity of the aluminium block. This block has a large heat conduction and thereby avoids temperature gradients. It limits heat convection by protecting the devices from outside air streams. Finally, it emits heat radiation which is isotropic inside the cavity due to the resemblance to a black body [13]. This black-body action is enhanced by blackening the cavity walls.

The temperature of the block was constantly measured by the Pt-100 platinum resistor inside its wall. This resistor was of a cylindrical, wire-wound type with a length of 25 mm and corresponds to the DIN43760 norm. It was biased with 2.1 mA by an RS Platinum resistance instrumentation lineariser, delivering 1 mV $^{\circ}C^{-1}$ at its output. This voltage was measured with an HP 34401A multimeter. The absolute accuracy of these temperature measurements was estimated at 0.4 °C.

The thermal stability of this set-up was tested in two steps. Firstly, the absolute temperature was monitored as a function of time after switching on all equipment and the climate control of the room. This created a step response with a maximum after 2.5 hours, an overshoot of about 0.3 °C, and a thermal stability within 1.0 °C afterwards.

Secondly, the collector currents of a transistor pair were monitored together with the Pt-100 resistance during a complete measurement cycle of 90 minutes. In this period the stepper motor was driven, but without actually stressing the

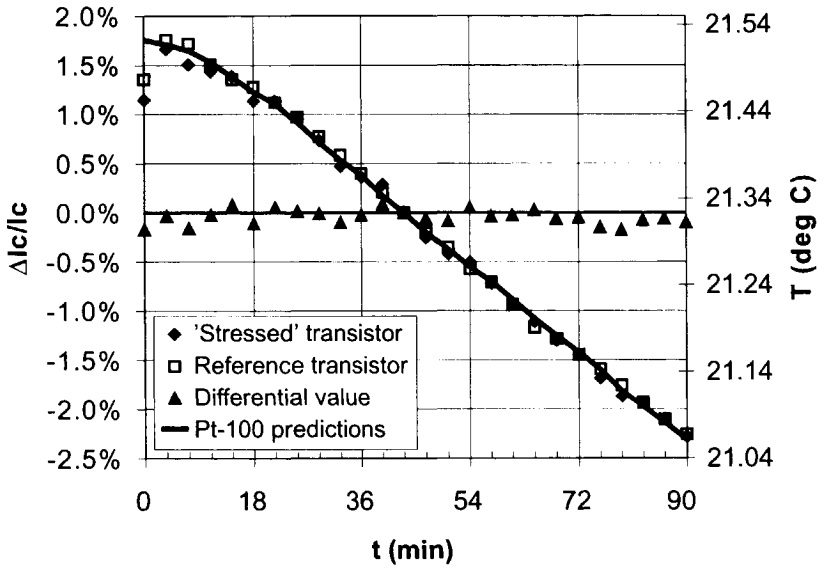

Figure 3.19: Change in the collector currents due to ambient temperature variations in one measurement cycle. Both transistors remain unstressed here while the microscrew is moving. The drawn line is a prediction based on temperature measurements by the Pt-100 and the bandgap equation of Ref. [8].

beams. The results are shown in Figure 3.19. It indeed appears that the currents change with a considerable amount, about 4%, due to a variation of the ambient temperature of 0.47 °C. The difference between the two currents, however, remains very constant. Moreover, it appears that the current changes are closely correlated to the temperature measured by the Pt-100. Those temperatures can also be used to predict the current changes theoretically. For this purpose the following equation can be used [8, 9, 14]:

$$\frac{\Delta I_c}{I_{c0}} = \left(\frac{T}{T_0}\right)^{\eta} \exp\left[\frac{q\,(V_{g0} - V_{be})}{k_B T}\,\frac{T - T_0}{T}\right] \tag{3.14}$$

where the subscript '0' indicates the situation at the time $t = t_0$, and where $\eta = 3.72$ and $V_{g0} = 1.166$ V. These predictions even lead to a curve which follows the measurements closely.

It can therefore be concluded that the temperature variations are slow, and that the temperature differences are very small between the stressed transistor, the reference transistor, and the Pt-100.

Electrical connections and shielding

The electrical connections of the bending apparatus should fulfil at least four requirements. They should:

- Offer a mechanical attachment to the wires of the source-measurement units which is robust and reliable;
- Keep the wires to the beam carrier as short as possible to avoid oscillations;
- Enable a flexible addressing of the connector pins of the beam carrier;
- Shield the internal wires from electrical interference.

This was obtained by designing an aluminium connection box, which is screwed to the bending block. It is shown in Figure 3.8 and 3.17. On the outside, it contains triaxial sockets and screwholes to directly attach the quadraxial wires of the source-measurement units. On the inside, it has a number of short coaxial wires leading to small crimp terminals with which the pins of the beam carrier can be addressed.

The electrical shielding and guarding of the SMUs is continued inside the box as far as possible. The box and the block are grounded through the shielding of the SMU cables. They are anodised with an isolating layer, but electrically re-connected to each other by blanking some screw holes. The force and the sense line of each SMU are kept separated down to the crimp terminals. They are also guarded down to this point.

Measurements showed that the connections were effective. They did not reveal any oscillations, in contrast to earlier measurements with longer cables. In addition, they did not show noise, current leakage or thermal voltage outside the specifications of the SMUs.

Possible extensions

The construction of the measurement set-up allows a number of extensions. Its core is mounted on an optical base plate, offering space and screwholes for extra equipment, as shown in Figure 3.17.

This could include a dynamic excitation of the beam. In combination with coherent detection this is an efficient way to suppress noise, drift, and offset. The excitation could be realised, for instance, by forcing the beams into vibration by a pulsed air stream, a magnetic coil, or the rotation of an asymmetric wheel.

The core of the set-up could also be placed in a climate chamber. In this way the ambient temperature can be controlled more precisely, or varied over a wide range to measure its influence on the piezojunction effect. For this reason the core has been kept rather compact ($50 \times 40 \times 16$ cm).

3.2.5 Source-measurement units

The source-measurement units (SMUs) in the set-up are made by Hewlett-Packard and united in the HP 4142B Modular DC Source/Monitor. They were

used to force a voltage and sense the resulting current. The voltages applied to the base-emitter junctions varied between -0.8 and +0.8 V and fell in the 2 V-range of the SMUs. The collector-emitter voltages were varied between between -0.8 to 4.2 V and used the 20 V range. The resulting currents varied roughly between 1 pA and 10 mA. They were measured at the measurement ranges between 1 nA and 10 mA, which have a spacing of a decade.

The accuracy of the SMUs depends on the measurement range, but is typically 0.05%. This is close to the full-scale output resolution of 0.02%, which determines the quantisation noise. This noise source is the dominant one, especially if the measurements are averaged. To ensure the accuracy of the measurements, the SMUs were calibrated shortly before.

3.3 Operation of the measurement set-up

The measurement set-up was operated in three steps, discussed in this section. Firstly, the devices were excited and their response recorded. Secondly, the data from this procedure were processed to calculate intermediate results and add the error intervals. Thirdly, the data and the errors were used to extract information: the validity of the device model and its unknown parameters.

3.3.1 Measurement procedure

Before the measurements, the set-up was given time to warm up. It was usually switched on at the same moment as the climate control of the room. Subsequently, the signal of the Pt-100 was monitored until a stable temperature was reached, after three to four hours (see Section 3.2.4).

After the temperature stabilisation, the Gummel plots of the transistors were measured to reveal if the set-up was working correctly. This was done by stepping the base-emitter voltage from 0.3 to 0.8 V and measuring the collector and base currents. This voltage range was limited on the one side by the leakage currents, and on the other side by the maximum allowed current in the bonding wires.

Next, a measurement was performed to determine the 'dead move' of the device under test. A typical result of such a measurement was shown in Figure 3.18. It was done by stepping the microscrew over a small range in which the dead zone was expected and by measuring the changes in the device currents.

Finally, the bending measurements were performed. These measurements consisted of three nested cycles. For a specific bending and collector voltage the Gummel plots of the transistors were measured. This cycle was performed for collector voltages between 0.2 and 4.2 V and for different bendings with deflections of the beam tips between -2.5 and 2.5 mm. For the *pnp* transistors a slightly different procedure was followed, because they showed a high leakage between the collector and the base. Their collector voltage was therefore kept equal to the base voltage. The entire procedure took about 90 minutes for *npn*

transistors, and 20 minutes for the *pnp*s. During this time, 6318 and 2091 data points were collected, respectively.

3.3.2 Data processing

The data from the measurement set-up had to be processed before their entrance into the parameter extraction program. For this processing a program was written in FORTRAN 90. This program performed different operations:

1. It converted the deflections of the beam tips into a mechanical stress with the aid of Equation 3.12. For this purpose it was supplied with a file containing information on the specific beam: the dimensions, the dead move, and the crystal orientation.

2. It selected the measurement points with the same base-emitter voltage V_{be}. This was necessary because the parameter extraction program could handle only a limited amount of points. The absolute value of V_{be} was chosen around 0.6 V because at this point it obeyed the best to the ideal model of the collector current: $I_c = I_S \exp(qV_{be}/k_BT)$.

3. It calculated the ratio I_i between the I_c of the reference and the stressed transistors. The ratio was corrected for offset by using the information of the zero-stress points. It was needed in combination with Equation 3.6 to calculate the piezojunction coefficients.

4. It calculated the influence of the geometrical effect on the current changes.

5. It added to each measurement point i the factors A'_{ij} and B'_{ik}. These factors also appear in Equation 3.6 and contain the information about the orientation of the stress and the current in the transistor considered.

6. It added an error interval to each value in a measurement point. This interval was based on the *a priori* knowledge about the accuracy of the measurement set-up. The accuracy depends on many different factors, discussed in the preceding sections.

3.3.3 Model identification

After the processing, the data points were used to identify the model developed in Section 3.1.1. This was done with the use of the program ParX, developed by M. G. Middelhoek [15]. It relies on the following principles.

Relation between the measurements and the device models

Measurements can only give a limited view of the entire behaviour of a device. This behaviour is observed by measuring data points, which by nature cover only a finite number of situations. In addition, the observations are limited to the interface of the device, which means that the internal physical processes

remain hidden. It is possible, for example, to measure the collector current of a transistor, but not to observe the distribution of electrons in the base.

The internal mechanisms in a device can be described, however, by using a model. Such a model consists of a set of equations and a collection of parameters. The equations are often based on *a priori* knowledge about the device, which makes the model analytical in the sense of Section 1.6. The equations of the piezojunction effect, for instance, are based on the symmetry of silicon crystal and the construction of the transistor. In fact, the equations form a hypothesis about the device behaviour.

The parameters of the model may be known *a priori*. However, they may also be extracted from the observations if the model equations are sufficiently accurate. This parameter extraction is referred to as the *identification* of the model, and if it succeeds, it confirms the model hypothesis.

Method of the identification

Models are generally identified by using the least-squares method [1, 15, 16, 17]. This method chooses a parameter set which minimizes the distances between the measured points and the modelling curve. The remaining distance is a measure of the success of the identification.

This distance will never be zero, however, because of the errors. On the one hand, these errors are caused by imperfections in the measurements. On the other hand, they arise from the limitations of the model, which is necessarily a simplification of the reality. All errors result in an uncertainty of the extracted parameters.

For the calculation of the uncertainty it is necessary to specify the errors of all variables in the model equations. Ideally, an error is specified by its probability density function, or at least by its mean value and standard deviation. In practice, however, this information is rarely available. The program ParX avoids this problem by only requiring an error interval. This interval is a closed domain, centred at the observation, in which the true value is expected to lie.

Definition of the accuracy

Although ParX is based on the least-squares method, it differs from conventional programs in the way it obtains the required accuracy. It uses the fact that analytical models have a limited domain of validity. The piezojunction model, for instance, is only valid when the transistor is biased in forward. It also uses the fact that, in general, the domain of validity is not sharply limited. The forward-bias region, for instance, changes only gradually into the high-injection region if the collector current is increased. Analytical models therefore fit the measurements better if the domain is restricted.

As a result, ParX can improve the accuracy of the identification by discarding measurements at the boundaries of the validity domain. The number of discarded measurements can be chosen by the user in the form of an overall

accuracy limit. The program then yields the best parameter set, the error intervals of the parameters, and the domain in which these numbers are valid. This combination is quite valuable and unique.

ParX may abort the identification if too many measurements are discarded, or if the error intervals become too large. The abortion indicates that either the model equations or the measurements are too inaccurate.

3.4 Measurement results

The measurements have led to three important types of results. Firstly, they yielded the DC parameters at zero stress, which are an indication of the correct operation of the devices. Secondly, they provided the stress-induced changes in the saturation current and the resistance for different current and stress orientations. Finally, they yielded the piezojunction and piezoresistive coefficients employed in the model of Section 3.1.1.

3.4.1 Zero-stress operation

The transistors in forward bias showed the DC behaviour which was expected from the analysis in Section 3.2.2. This can be concluded from the Gummel plots and the current gain shown in Figure 3.20 and 3.21, which were measured at zero stress. The plots are quite similar for wafers of different orientation, and very similar for transistors on adjacent beams. It should be noted that the pair of *npn* transistors match so closely that the data points of the separate transistors can hardly be distinguished. It should also be noted that the deviating points at the right of those plots are artefacts of the current limitation in the measurement set-up.

The plots could be identified using ParX and a simple transistor model [15, 18]. The parameters of this model are the saturation current I_S, the transistor temperature T_1, the maximum current gain β_0, the forward Early voltage V_{AF}, and the reverse Early voltage V_{AR}. Their values are given in Table 3.1.

The values in this table are close to the standard values of the DIMES01 process [10]. The transistor temperatures T_1, however, are a few degrees higher than the T_2 measured by the Pt-100. They also show more spread. This is probably due to the approximative nature of the model.

npn transistors

The plots are especially successful for the *npn* transistors. Their collector current I_c varies exponentially with the base-emitter voltage V_{be} over almost eight decades of current. This range is limited on the lower side by leakage currents due to the finite insulating resistance of the set-up (around 100 GΩ). On the upper side it is limited by the maximum current allowed in the bonding wires. Just before the maximum current is reached, the effect of current crowding becomes visible.

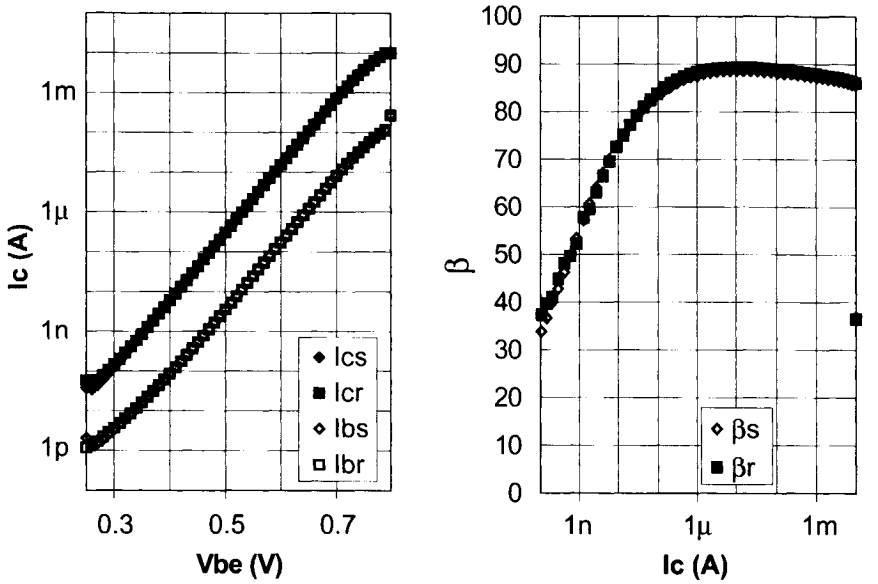

Figure 3.20: Gummel plots *(left)* and DC current gain as a function of the collector current *(right)* of a pair of *npn* transistors.

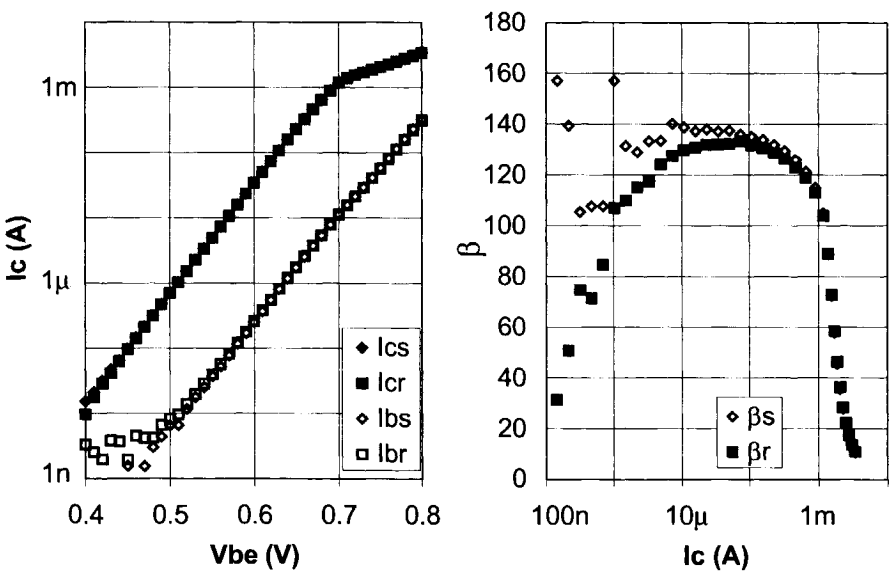

Figure 3.21: Gummel plots *(left)* and DC current gain as a function of the collector current *(right)* of a pair of *pnp* transistors.

Table 3.1: Extracted transistor parameters. The transistors were different between the boundaries indicated. The parameters of each transistor separately could be extracted with the typical accuracy in the fourth and seventh column. This is valid for the domain $0.54 \leq V_{be} \leq 0.66$ V for *npn*s and $-0.65 \leq V_{be} \leq -0.57$ V for *pnp*s. T_1 is the extracted transistor temperature, and T_2 is the temperature measured by the Pt-100.

Parameter	*npn*			*pnp*		
(unit)	**min.**	**max.**	**typ. acc.**	**min.**	**max.**	**typ. acc.**
I_S (fA)	1	3	0.2%	2	6	0.4%
T_1 (K)	295.7	303.5	0.1%	300.7	308.2	0.1%
T_2 (K)	294.5	297.6	0.2%	294.4	296.8	0.2%
β_0	80	117	1%	63	148	0.5%
V_{AF} (V)	44	66	10%	200	-	-
V_{AR} (V)	3.1	5.7	20%	10	-	-

pnp transistors

The plots for the vertical *pnp* transistors are somewhat less successful, but are still exponential over five decades of current. Their I_S and β_0 have the values expected. The current range is limited on the lower side by a parasitic shunt resistance between the base and the emitter of about 100 kΩ. This resistance was caused by a mistake in the mask design, by which the separation diffusion between the base and the surrounding epitaxial layer was left out. As a result, the voltage between the collector and the base had to be kept as small as possible. In addition, the Early voltage could not be determined. On the upper side of the range, I_c is limited by the base resistance, which happens close to the current predicted.

p-type resistors

The resistors also showed the zero-stress behaviour expected. The BW resistors measured 1.4 kΩ and the pinch resistors 47 kΩ. The matching between the resistors was within 2%. The resistance of the BW resistors does not show a dependence on the bias voltage, in contrast to the resistance of the pinch resistors. This pinching is relatively small (5% V^{-1}) at the bias voltage of 0.5 V.

3.4.2 Stress-induced transistor changes

When mechanical stress was applied to the transistors, their saturation current changed with a non-negligible amount. This piezojunction effect, however, was invisible on the scale of the zero-stress Gummel plots. It is therefore plotted differently in Figure 3.22 to 3.27. The plots show data of one base-emitter voltage of 0.6 V, and eigth different collector-emitter voltages between 0.2 and

4.2 V in the case of the *npns*. These data are compared with the model curves discussed in the following section. The plots have some remarkable features:

1. The points in these plots can be fit very well by parabolic curves.
2. The curves are asymmetric around the origin, and sometimes have a minimum away from this point.
3. The changes are very dependent on the orientation of both the current and the stress.
4. The changes depend on the transistor type: *npn* or *pnp*.
5. The changes are independent of the collector-emitter voltage V_{ce}.
6. The changes are quite repeatable between different samples.

In addition, the piezojunction effect shows to be independent from the bias voltage V_{be}, as long as this bias is in the forward, low-injection region. This point is illustrated by Figure 3.28. The independence from V_{ce} appears from Figure 3.22 to 3.27, where the points of different V_{ce} cannot be distinguished from each other.

The base currents show similar changes with stress, as shown in Figure 3.29. The changes are less pronounced, however, and show a slight decrease with decreasing bias currents.

3.4.3 Piezojunction coefficients

The identification of the model was quite successful. ParX could match Equation 3.4 to the measured data points in Figure 3.22 to 3.27 in an iterative process. The matching process yielded the theoretical curves shown in the same figures. It also yielded the unknown set of piezojunction coefficients as well as their accuracy, as shown in Table 3.2.

In some figures the matching curve follow one set of data points rather than another. This is the result of the data identification method of ParX, which considers the data of all six figures *at once*. A particular data set can therefore deviate from the overall 'average' indicated by the curves, and defined by the data of all transistor orientations and the system of polynomial series of Equation 3.4. If this deviation is too large, data points of the set are recognised as wrongly measured or out of the domain of validity of the model. They are therefore discarded. The percentage of discarded data points can be varied by the choice of the overall accuracy limit (see Section 3.3.3).

Although most coefficients are only obtained within a linear combination, it is possible to split them a little further. For this purpose literature results were used describing saturation current changes as the result of hydrostatic pressure [19, 20, 21]. In these references the junction voltage was measured as a function of the pressure, for constant current and temperature. This function was found

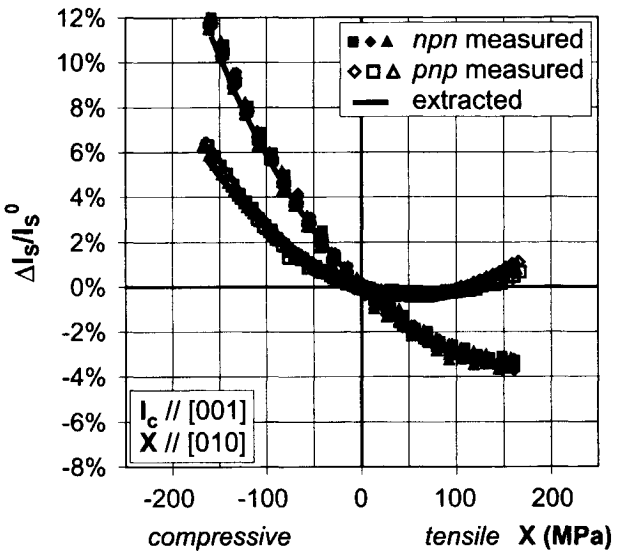

Figure 3.22: Saturation current changes as a function of uniaxial stress, for different *npn* and *pnp* transistors and for the orientations indicated. The plot shows both the measured values and the model curves corresponding to the extracted piezojunction coefficients.

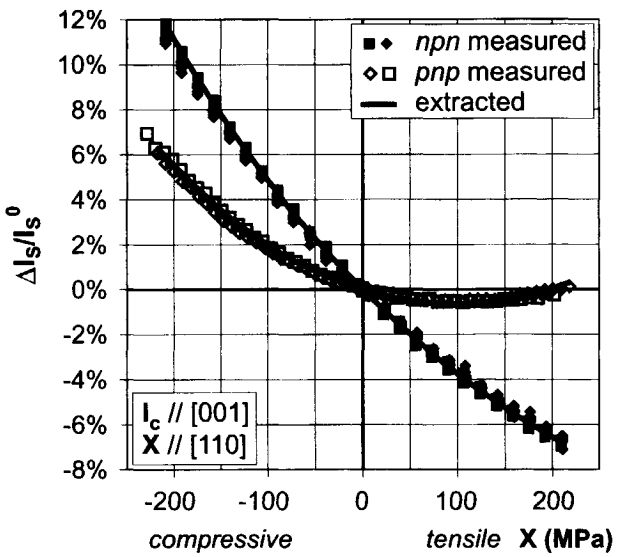

Figure 3.23: As Figure 3.22, but for another stress orientation.

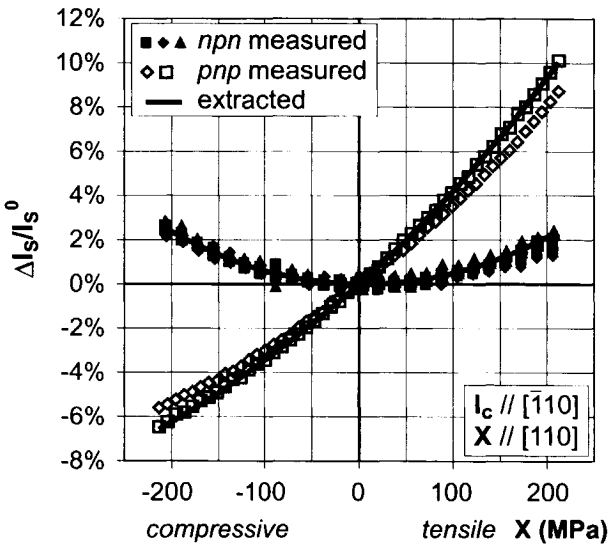

Figure 3.24: Saturation current changes as a function of uniaxial stress, for different *npn* and *pnp* transistors and for the orientations indicated. The plot shows both the measured values and the model curves corresponding to the extracted piezojunction coefficients.

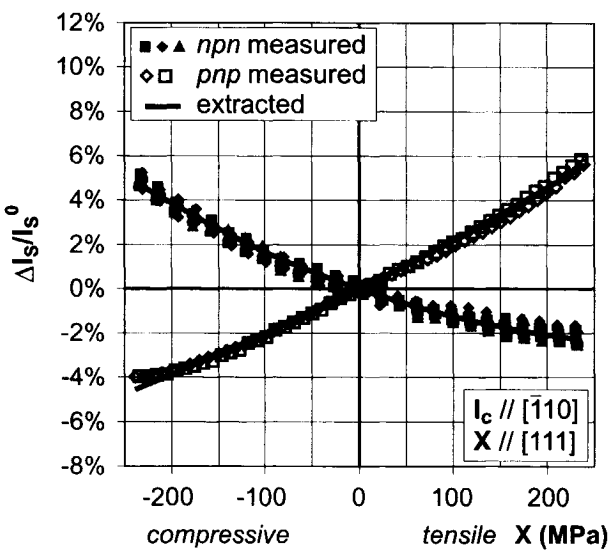

Figure 3.25: As Figure 3.24, but for another stress orientation.

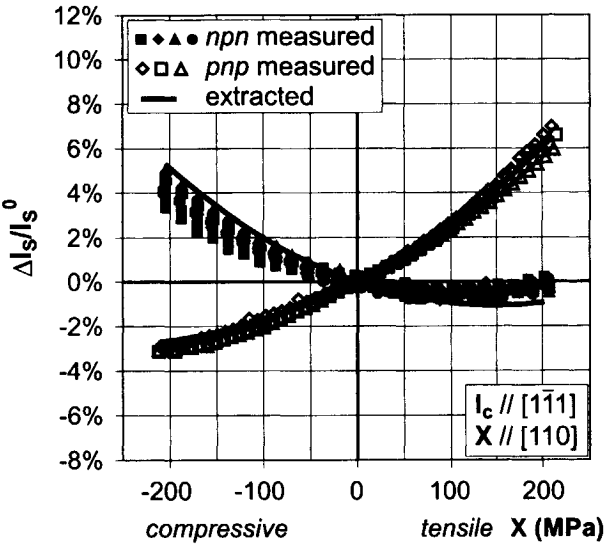

Figure 3.26: Saturation current changes as a function of uniaxial stress, for different *npn* and *pnp* transistors and for the orientations indicated. The plot shows both the measured values and the model curves corresponding to the extracted piezojunction coefficients.

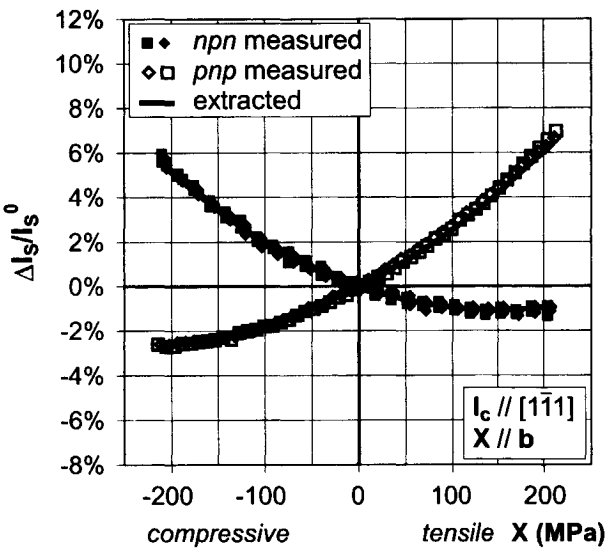

Figure 3.27: As Figure 3.26, but for stress along the vector **b** (see Figure 3.1).

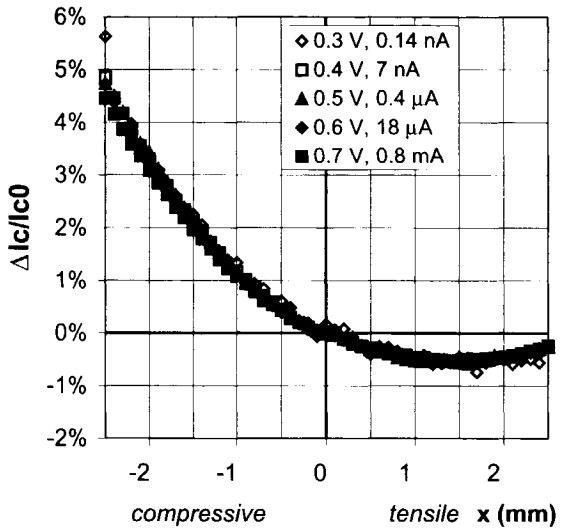

Figure 3.28: Relative changes in the collector current of an *npn* transistor as a function of the deflection of the beam end. The changes have been plotted for different V_{bc}. The zero-stress currents are noted in the legend.

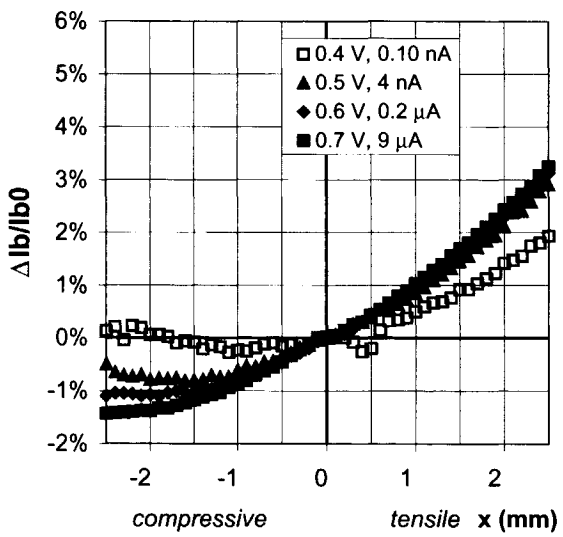

Figure 3.29: Relative changes in the base current of an *npn* transistor as a function of the deflection of the beam end. The changes have been plotted for different V_{bc}. The zero-stress currents are noted in the legend.

Table 3.2: First- and second-order piezojunction coefficients as extracted from the measurements. R.c. stands for 'Reduced coefficients'.

R.c.	Linear combination	Electrons	Holes
	First order ($\times 10^{-11}$ Pa^{-1})		
$\zeta_1^{I\prime}$	$\zeta_{11} - \zeta_{44}$	-41.5 ± 3.1	-89.0 ± 3.3
$\zeta_2^{I\prime}$	ζ_{12}	43.4 ± 1.5	13.8 ± 1.3
	Second order ($\times 10^{-18}$ Pa^{-2})		
$\zeta_1^{II\prime}$	$\zeta_{111} - 4\zeta_{616} + 2\zeta_{414} + 2\zeta_{456}$	-1.5 ± 1.0	-1.7 ± 1.3
$\zeta_2^{II\prime}$	$\zeta_{112} + \frac{1}{2}\zeta_{166} - \zeta_{414} - \zeta_{456}$	0.30 ± 0.43	0.99 ± 0.53
$\zeta_3^{II\prime}$	ζ_{122}	-1.29 ± 0.21	-1.21 ± 0.23
$\zeta_4^{II\prime}$	$\zeta_{123} + \frac{1}{2}\zeta_{144}$	0.54 ± 0.30	-0.11 ± 0.29

to be linear up to 3.7 GPa (!) pressure. It means that the saturation current of the devices increases exponentially with this type of stress:

$$\frac{I_S(X)}{I_S(0)} = \exp\left(-\frac{q\alpha_h X}{k_B T}\right) \tag{3.15}$$

where X is the stress magnitude of the hydrostatic pressure and α_h the hydrostatic sensitivity. This sensitivity α_h is $(1.500 \pm 0.005) \times 10^{-11}$ VPa^{-1} according to Jayaraman *e.a.* [19], 1.50×10^{-11} VPa^{-1} according to Loaëc *e.a.* [20], and 1.44 to 1.65×10^{-11} VPa^{-1} according to Włodarski and Moeschke [21]. The latter two values were obtained at pressure ranges of 60 and 100 MPa, respectively.

The above equation can be compared to the power expansion of Section 3.1.1 for the case of hydrostatic pressure. This yields an extra equation to the system of six in Equation 3.4:

$$I_7 = 1 + (\zeta_{11} + 2\zeta_{12})\,X + (\zeta_{111} + 4\zeta_{112} + 2\zeta_{122} + 2\zeta_{123})\,X^2 \tag{3.16}$$

As a result, $\zeta_1^{I\prime}$ and $\zeta_4^{II\prime}$ in Table 3.2 can be split up in the quantities of Table 3.3.

The literature on hydrostatic pressure seems to be quite reliable for holes, but far less reliable for electrons. This point will be discussed further in Section 4.3.

3.4.4 Stress-induced resistor changes

The application of stress to the p-type resistors resulted in resistance changes which were more or less expected from literature. Some of them are shown in Figure 3.30 and 3.31. They have the following characteristics:

1. The resistance changes in the considered stress range are quite linear.
2. The changes are very dependent on the current orientation.
3. The changes are repeatable between different samples.

Table 3.3: First- and second-order piezojunction coefficients as extracted from both measurements and literature. R.c. stands for 'Reduced coefficients'.

R.c.	Linear combination	Electrons	Holes
	First order ($\times 10^{-11}$ Pa^{-1})		
$\zeta_1^{\mathrm{I}\prime\prime}$	ζ_{11}	-28.4 ± 3.0	30.8 ± 2.6
$\zeta_2^{\mathrm{I}\prime\prime}$	ζ_{12}	43.4 ± 1.5	13.8 ± 1.3
$\zeta_3^{\mathrm{I}\prime\prime}$	ζ_{44}	13.1 ± 4.3	119.8 ± 4.2
	Second order ($\times 10^{-18}$ Pa^{-2})		
$\zeta_1^{\mathrm{II}\prime\prime}$	$\zeta_{111} - 4\zeta_{616} + 2\zeta_{414} + 2\zeta_{456}$	-1.5 ± 1.0	-1.7 ± 1.3
$\zeta_2^{\mathrm{II}\prime\prime}$	$\zeta_{112} + \frac{1}{2}\zeta_{166} - \zeta_{414} - \zeta_{456}$	0.30 ± 0.43	0.99 ± 0.53
$\zeta_3^{\mathrm{II}\prime\prime}$	ζ_{122}	-1.29 ± 0.21	-1.21 ± 0.23
$\zeta_4^{\mathrm{II}\prime\prime}$	$\zeta_{123} - \zeta_{166} + 2\zeta_{616} + \zeta_{414} + \zeta_{456}$	1.6 ± 1.0	0.3 ± 1.3
$\zeta_5^{\mathrm{II}\prime\prime}$	$\zeta_{144} + 2\zeta_{166} - 4\zeta_{616} - 2\zeta_{414} - 2\zeta_{456}$	-2.1 ± 2.1	-0.7 ± 2.6

4. The resistance changes of the BW resistors are somewhat lower than those of the pinch resistors in the same orientation.

3.4.5 Piezoresistive coefficients

The first-order stress-sensitivity of the resistors was used to calculate the piezoresistive coefficients in the way described in Section 3.1.1. The results are shown in Table 3.4. They are close to the coefficients found by other authors. The largest coefficient is π_{44} and is higher for the pinch resistors than for the BW-resistors. This can be explained by the higher doping concentration of the latter (see Section 3.2.2).

Some remarks can be made about the accuracy of the calculated coefficients. Firstly, the accuracy cannot be determined in a consistent way because the coefficients were extracted with the inversion method of Equation 3.10. This method is simpler than the one employed for the piezo*junction* coefficients. Secondly, the coefficient π_{44}^{hl} varies slightly between the samples, and between rosettes of different orientation (at most 7%). It is also quite consistent with literature values. However, the other coefficients seem to vary wildly. This can be explained from the difference in magnitude between the coefficients, and the fact that the coefficients are linearly related to the measurements and each other and the measurements (see Equation 3.10). A small error in a measurement value will therefore result in errors in π_{44}^{hl} and π_{11}^{hl} which are approximately equal in an absolute sense, but very different in a relative sense. A further theoretical analysis of the propagation of errors is given in Reference [5].

In the following chapter it will be discussed how these coefficients are related to the findings on the piezojunction effect.

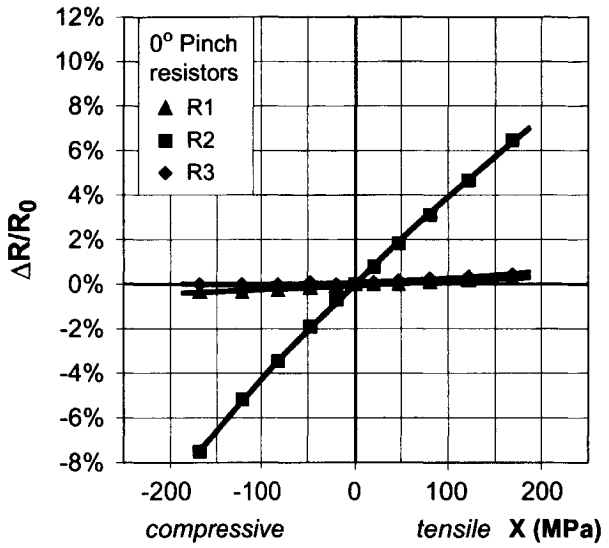

Figure 3.30: Relative change in the resistance of three p-type pinch resistors in the (001)-plane as a result of uniaxial stress along the [130]-direction. The resistors are oriented as in Figure 3.2, with $\eta = 0$.

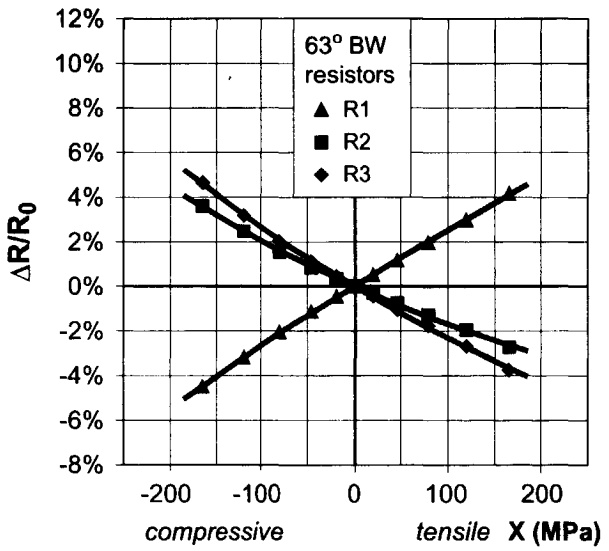

Figure 3.31: Relative change in the resistance of three p-type pinch resistors in the (001)-plane as a result of uniaxial stress along the [130]-direction. The resistors are oriented as in Figure 3.2, with $\eta = 63°$.

Table 3.4: First-order piezoresistive coefficients for holes as determined from the experiments, and compared with some experimental literature values. See also Table 3.4 for theoretical predictions.

Coefficient	Experiments		Literature	
$(10^{-11}\ \mathrm{Pa}^{-1})$	Pinch	BW	[23]	[24]
π^{p}_{11}	1.8	−1.1	6.6	−6
π^{p}_{12}	2.5	1.2	−1.1	1
π^{p}_{44}	118.4	99.5	138.1	112

3.5 Conclusions

Several experiments were set-up to validate the models of the piezojunction effect developed in Chapter 2. For this purpose the saturation currents of *npn* and *pnp* transistors were determined for different values of uniaxial stress. The transistors as well as the stress had different orientations. In addition, the resistance changes of some resistors were determined to benchmark the experiments.

The stress was generated by sawing the wafers with the transistors in long, thin strips, and bending these as a cantilever. For this purpose a special apparatus was developed. While the strips were being stressed, source-measurement units measured the Gummel plots of the transistors and the resistances of the resistors.

The measurements showed saturation currents which varied up to 12% for the stress range considered. The variations were almost parabolical with the stress magnitude, and strongly depended on the transistor type and the stress and current orientation.

All measurement data were succesfully identified with the model equations. This yielded a set of piezojunction coefficients with acceptable error intervals. Unfortunately, some coefficients could not be determined because of the limited set of orientations considered. The resistor measurements yielded more linear curves and piezoresistive coefficients which correspond well with literature values.

Bibliography

[1] G. Strang, *Linear algebra and its applications*, Harcourt Brace Jovanovich, San Diego, 1988.

[2] R. E. Beaty, R. C. Jaeger, J. C. Suhling, and R. W. Johnson, Evaluation of piezoresistive coefficient variation in silicon stress sensors using a four-point bending test fixture, *IEEE Trans. Comp., Hybrids, Manuf. Techn., 15* (1992) 904-914.

[3] D. A. Bittle, J. C. Suhling, R. E. Beaty, R. C. Jaeger, and R. W. Johnson, Piezoresistive stress sensors for structural analysis of electronic packages, *J. Electron. Pack., 113* (1991) 203-214.

[4] R. C. Jaeger, J. C. Suhling, M. T. Carey, and R. W. Johnson, Off-axis sensor rosettes for measurement of the piezoresistive coefficient in silicon, *IEEE Trans. Comp., Hybrids, Manuf. Techn., 16* (1993) 925-931.

[5] J. F. Creemer and P. J. French, Reduction of uncertainty in the measurement of the piezoresistive coefficients of silicon with a three-element rosette, *Proc. SPIE, Smart structures and materials, Smart Electronics and MEMS,* San Diego, CA, USA, March 2-4, 1998, Vol. 3328, pp. 392-402.

[6] J. M. Gere and S. P. Timoshenko, *Mechanics of materials,* Chapman and Hall, London, 1991.

[7] K. Yamada, M. Nishihara, S. Shimada, M. Tanabe, M. Shimazoe and Y. Matsuoka, Nonlinearity of the piezoresistance effect of p-type silicon diffused layers, *IEEE Trans. El. Dev., ED-29* (1982) 71-77.

[8] G. C. M. Meijer, Thermal sensors based on transistors, *Sensors and Actuators, 10* (1986) 103-125.

[9] G. C. M. Meijer, *Integrated circuits and components for bandgap references and temperature transducers,* Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 1982.

[10] L. K. Nanver, E. J. G. Goudena, and H. W. van Zeijl, DIMES-01, a baseline BIFET process for smart sensor experimentation, *Sensors and Actuators A, 36* (1993) 139-147.

[11] F. Fruett and G. C. M. Meijer, Exploration of the piezojunction effect using PNP lateral transistors on [100] silicon, *Proc. 14th Eur. Conf. Solid-State Transducers (Eurosensors XIV)*, 27-30 August 2000, Copenhagen, Denmark, pp. 287-288.

[12] F. Fruett and G. C. M. Meijer, A new sensor structure using the piezojunction effect in PNP lateral structures, *Sensors and Actuators A, in press* (2001).

[13] W. Greiner, *Quantum mechanics, an introduction, Ch. 2*, Springer, Berlin, 1989.

[14] R. Amador, A. Polanco, and A. Nagy, The spread of η and V_{go} and its influence on the sensitivity of a bipolar IC Celsius sensor, *Sensors and Actuators A, 77* (1999) 9-13.

[15] M. G. Middelhoek, *The identification of analytical device models*, Ph. D. Thesis of the Delft University of Technology, Delft University Press, Delft, The Netherlands, 1992.

[16] R. W. Hamming, *Numerical methods for scientists and engineers*, 2nd ed., Dover, New York, 1973.

[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in FORTRAN 77, The art of scientific computing*, 2nd ed., Cambridge University Press, Cambridge, 1996.

[18] S. M. Sze, *Physics of semiconductor devices*, Wiley, New York, 1981.

[19] A. Jayaraman, M. E. Sikorski, J. C. Irvin, and G. H. Yates, Effect of hydrostatic pressure on *p-n* junction characteristics and the pressure variation of the band gap, *J. Appl. Phys., 38* (1967) 4454-4459.

[20] J. Loaëc, A. Pochat, et P. Johannin, Effets de la pression hydrostatique et de la température sur la caractérestique directe de diodes jonctions au silicium, *C. R. Acad. Sc. Paris, Série B, 267* (1968) 856-859.

[21] W. Wlodarski and B. Moeschke, The effect of hydrostatic pressure on the characteristics of the forward biased *p-n* junctions, *Electron Technol., 13* (3) (1980) 3-42.

[22] J. J. Wortman, J. R. Hauser, and R. M. Burger, Effect of mechanical stress on *p-n* junction device characteristics, *J. Appl. Phys., 35* (1964) 2122-2131.

[23] C. S. Smith, Piezoresistance effect in germanium and silicon, *Phys. Rev., 94* (1954) 42-49.

[24] K. Matsuda, K. Suzuki, K. Yamamura, and Y. Kanda, Nonlinear piezoresistive coeficients in silicon, *J. Appl. Phys., 73* (1993) 1838-1847.

Chapter 4

Discussion and Conclusions

In Chapter 2 and 3 the central question of this thesis is discussed: how can the saturation current of a silicon bipolar transistor be modelled as a function of mechanical stress if this stress is lower than 200 MPa and may have any orientation?

The answer is a model which depends on one material parameter in particular: *the conductivity of minority charge carriers in thermodynamic equilibrium.* The stress dependence of this minority conductivity was shown to be the main factor contributing to the piezojunction effect. It was modelled in three stages:

- By enhancing a microscopic model that calculates the changes in the conductivity from the deformation of the crystal and the variations in the energy band structure;
- By developing a new, macroscopic model which is based on a polynomial series and which is restricted by the symmetry properties of silicon. This model is considerably faster to evaluate than the microscopic model;
- By performing bending experiments on transistors, in which the collector current is measured as a function of the current direction, the stress magnitude, the stress orientation, and the base-emitter voltage.

In all stages, the piezojunction effect was found to resemble the *piezoresistive* effect. This effect, describing the conductivity of *majority* charge carriers, has been described for many years and served as a reference point for the results.

A comparison between the results of the modelling stages shows a good agreement. Data points from the bending experiments are accurately followed by curves from the macroscopic model and enable the extraction of the model parameters. The curves could also be predicted from the microscopic model. Finally, the parameters of the macroscopic model can analytically be expressed in terms of the microscopic model.

The following sections discuss the validity of the models and the experimental observations. They also sketch the new opportunities for the design of circuits

and sensors in which stress plays a role. Finally, some open questions and ideas for future research are presented.

4.1 Validity of the microscopic model

The microscopic model relates the piezojunction effect to stress-induced events at the atomic level. The model predicts conductivity changes which can be compared with the experimental results. It thereby makes a large number of assumptions which limit the domain of validity. Compared to previous models, however, it has a number of advantages.

Structure of the model

The microscopic model is based on the configuration of a single conduction electron forming a standing wave with the periodicity of the crystal lattice. This lattice deforms when a mechanical stress is applied. As a result, the periodicity of the wave function is changed and in general also its symmetry. This modifies the energy eigenvalues of the electrons, corresponding to the energy bands. In particular, it changes the levels and curvatures of the band edges around the forbidden gap. These edges are responsible for the electrical conduction. Any symmetry changes result in a lifting of the degeneracy of the edges. The lifting of the degeneracy requires that the classical concept of the bandgap is redefined. In addition, it requires a modern definition of the band edge diagram used in the calculation of charge transport.

The changes in the band structure can be translated into changes of the conductivity and the electron concentration. For this purpose electrons are considered as semi-classical particles and fields as first-order perturbations. They can therefore be described by the semiclassical transport model, from which the drift-diffusion equations are rederived. The resulting conductivity is anisotropic in the presence of stress and should be written as a tensor. In addition, it should be formulated separately for each band. The conductivity tensor relates the current density to the gradient in the quasi Fermi level. Eventually, this gradient can be split up into different terms, yielding the classical drift-diffusion equations extended with gradients in the electron affinity, the effective density of states, and the bandgap.

Limitations of the model

The proposed microscopic model depends on a whole series of simplifications and assumptions which limit its working range and decrease its accuracy. It is striking, for example, on how many occasions the electrons are supposed to be independent:

- In a Schrödinger equation where interaction with other electrons only occurs through some average field;

- In Fermi-Dirac statistics, applicable when electrons form a dilute gas;
- In Maxwell-Boltzmann statistics, applicable when the electron concentration is small compared to the concentration of the states and the Pauli exclusion principle becomes insignificant;
- In the relaxation time approximation, where the scattering of an electron does not perturb the distribution function.

Inevitably, these assumptions were violated to some degree in the experiments, where the doping concentration in the base was 3×10^{17} cm^{-3} and the concentration in some resistors was even higher.

In addition, there are other causes for differences between the model predictions and measured curves:

- The relaxation time approximation, in which the information of an electron in nonequilibrium is lost by just one collision.
- Uncertainties in the literature values for the band parameters and the deformation potentials, which were often only measured at temperatures of a few Kelvin;
- The high sensitivity of the model to variations in the input parameters;
- The uncertainties in the measurements themselves.

Predictions of the experimental effects

The microscopic model predicts many aspects of the piezojunction and the piezoresistive effects. The conductivity shows the same parabolic dependences on stress as the measurements, which can be seen in Figure 4.1. The curves are different for tensile stress and for compressive stress and sometimes have a minimum on the stress range considered. The curves are clearly different for electrons and holes. In addition, they show a strong anisotropic behaviour because they vary strongly with the current and stress orientations with respect to the crystal axes. Finally, they are more linear for majority charge carriers than for minority charge carriers, which is shown by Figure 4.2.

There are also some clear differences between the microscopic model and the measurements. The scale of the predicted curves is 50 to 100% larger, depending on which set of input parameters is used. In addition, the nonlinearities are larger than measured. Finally, the predicted piezocoefficients deviate from the measured coefficients, although they remain in the same order of magnitude (see Table 2.5 and 3.4). A similar phenomenon may be found in the literature on piezoresistive coefficients, where many different values occur.

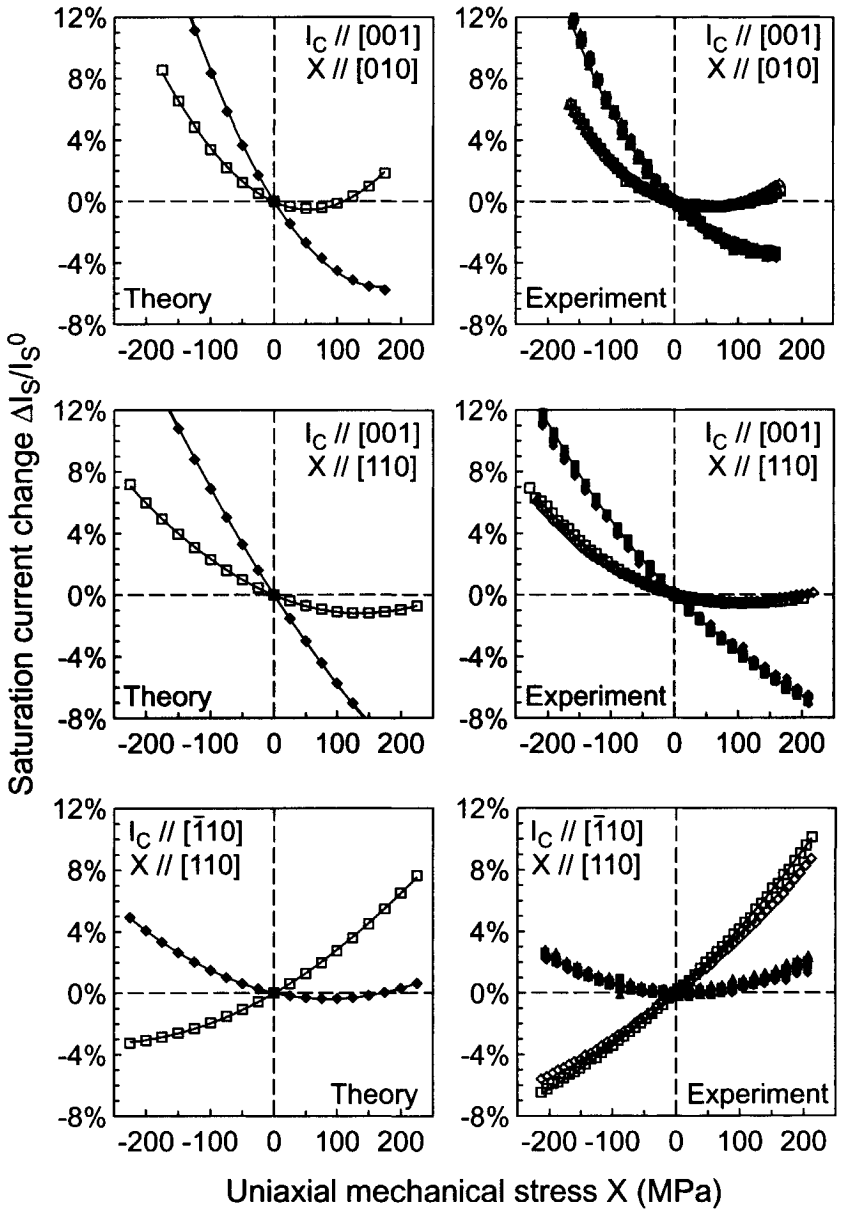

Figure 4.1: Comparison between predicted and measured saturation current changes of *npn* and *pnp* transistors as a function of a uniaxial sress, for the current and stress orientations indicated. The figure resumes Figure 2.28 to 2.31, and 3.22 to 3.25.

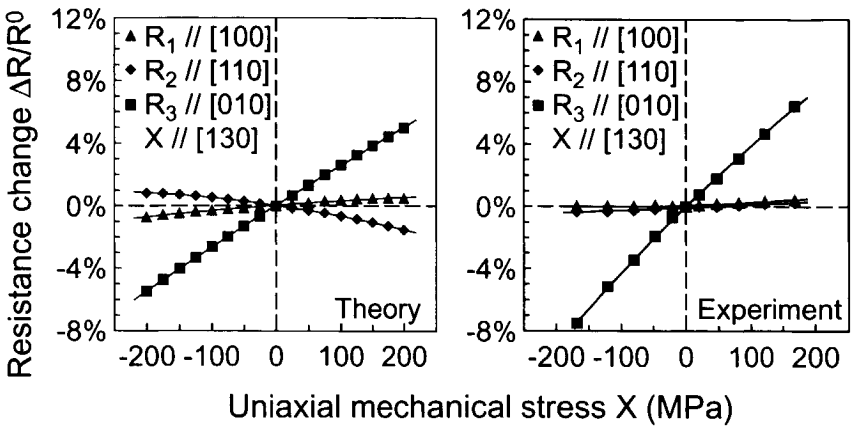

Figure 4.2: Comparison between predicted and measured resistance changes as a function of a uniaxial stress, for three p-type resistors of different orientations. The figure resumes Figure 2.25 and 3.30.

Comparison with previous models

The microscopic model presented in this thesis is the first to describe the piezojunction effect for stress of any orientation, with a compressive or tensile nature, and with a magnitude below 200 MPa. This is why it differs from the models of Wortman, Wlodarski, and Kanda [1]-[4].

The model of *Wortman, Hauser, and Burger* [1] attributes the changes in the saturation current entirely to the stress-induced shifts of the band edges. The effective masses and the effective densities of states are supposed to be constant. Therefore, it supposes that only the intrinsic carrier concentration varies to a certain extent, in contrast to the mobility. As a result, it predicts an equal piezojunction effect for *npn* and *pnp* transistors, and an independence of the current direction. Clearly, these predictions are contradicted by the experimental results of this work. However, they were supported by many experiments during the 1960's, cited in Chapter 1. This can be explained by the fact that these experiments were performed with a pressing stylus, introducing high stresses between 1 and 10 GPa. In this range the changes in the effective masses saturate and are also overwhelmed by the effects of the shifting band edges.

The model of *Wlodarski and Moeschke* [2] also limits itself to band edge shifts and the resulting changes in the intrinsic concentration. In this model, however, the stress is assumed to be hydrostatic. This maintains the symmetry of the semiconductor crystal and therefore hardly influences the effective masses. However, the hydrostatic assumption limits of course the applicability of the model. It was experimentally verified by the authors for pressures up to 100 MPa.

The *Kanda* models [3],[4] are more sophisticated because they include the

shifts in the band edges as well as the changes in the effective masses. The mass changes are supposed to modify the effective densities of states and the mobilities. At low stress values the mobilities are also supposed to vary with the first-order piezoresistive coefficients, available from experiments. However, the mass changes of the valence bands could only be calculated in the limiting case where the bands are completely decoupled and their description is relatively simple. This means that the validity of the model is restricted to very high stress levels of more than 1 GPa. The Kanda models were partially validated by a comparison with curves from the Wortman model and with some experiments from literature.

The microscopic model presented in *this work* is valid for stresses with a magnitude below 200 MPa, with an arbitrary orientation and sign. It describes the changes in the conductivity as a whole, taking into account the band edge shifts and the effective mass changes at the same time. Especially for the valence bands this makes a considerable difference. The model can predict the piezojunction effect numerically because of the analytic equations of the valence bands. These equations were derived after the Kanda models were published [5],[6]. The predictions have also become feasible because of the dramatic increase in computing power. However, the numerical effort needed for the microscopic model may still be a drawback in electronic design, where a truly analytical model is needed (see Section 1.6). It is therefore useful that the model can be abstracted to the macroscopic model.

4.2 Validity of the macroscopic model

The macroscopic model relates the piezojunction effect to equations suggested by measurements. Its analytic structure allows very rapid calculations but introduces some limitations to the domain of validity. The model can be connected to the microscopic model because its quantities have a physical significance. It thereby provides a valuable opportunity to verify the microscopic model against experimental observations.

Structure and limitations of the model

The macroscopic model describes the piezojunction effect through the stress-induced changes in the observable variables of a transistor: the collector current and the base-emitter voltage. These changes are small and are described with a quadratic tensor series. This allows a very rapid calculation but limits the validity to the observed stress and current range. In addition, it requires a very large number of polynomial coefficients.

To give the model a practical significance, more *a priori* knowledge has to be introduced. For this purpose the collector current was modelled according to the Gummel approach. Under the assumptions of low-level injection and forward bias, the current is proportional to the saturation current. The saturation current in turn is proportional to the *minority conductivity in thermodynamic*

equilibrium. This conductivity is supposed to be the dominant source of stress-induced changes. In that case, the changes are independent from the collector current, which was indeed observed experimentally.

The minority conductivity is a material property, just as the better-known conductivity of majority charge carriers. According to Neumann's principle, the properties of crystalline materials have all the symmetry elements of the crystal. The same principle applies to the first- and second-order derivatives of the conductivity at zero stress, which are the coefficients of the polynomial tensor series.

The symmetry elements of silicon reduce the number of independent coefficients to three for the first-order tensor, and nine for the second-order tensor. This gives the series a particularly simple shape. The coefficients of the majority conductivity are defined as the well-known *piezoresistive coefficients*. The coefficients of the minority conductivity are for the first time defined in this thesis and are called the *piezojunction coefficients*.

Correspondence with the microscopic model and the measurements

The connection of the collector current to the minority conductivity relates the macroscopic model to the microscopic model. In the microscopic model, the minority as well as the majority conductivity can be calculated under the assumption of extrinsic material. This also enables the calculation of their derivatives in stress: the piezojunction and piezoresistive coefficients, respectively.

The minority conductivity depends on both the conduction bands and the valence bands. It therefore partially depends on the forbidden band gap. In this respect it differs from the majority conductivity, which only depends on the bands of the majority charge carriers.

There are good reasons to believe that the link with the microscopic model is correct. The previous section already showed that the calculated first-order coefficients approach the measured coefficients. They are not very precise, but this can be attributed to the assumptions of the model and the uncertainties in the input variables. In addition, the model predicts that the first-order shear coefficients of the piezojunction effect are equal to those of the piezoresistive effect. The measured values are very close to this prediction.

There is another good reason to believe that the macroscopic model describes the measurements quite accurately. The model equation is given a very specific shape by the symmetry requirements. Nevertheless, the equation matches the curves of six differently oriented transistors with one and the same set of coefficients and with a reasonable accuracy, as shown by Figure 4.1, and by Figure 3.22 to 3.27. This is true for both the *npn* and the *pnp* transistors.

Predictions of the device characteristics

The polynomial of the macroscopic model can be rapidly evaluated once the piezojunction coefficients are known. This makes it ideal to predict changes in the device characteristics with various current and stress orientations. For

the first time, the orientations can be explored systematically, as done in Section 2.4.3. The resulting polar plots give a good picture of the orientational dependence of the effect. They are similar to the plots of the piezoresistive effect presented by Kanda [11]. Such plots can be used for optimising the orientation of a transistor for a particular task.

Because of its simplicity, the macroscopic model can easily be identified with the data of new measurements. This yields new values for the piezojunction coefficients. Generally, the model identification requires intensive calculations, such as those in the least-squares method. It is therefore an advantage that the model equations themselves can rapidly be evaluated.

4.3 Validity of the measurements

Compared to the models, the measurements show the piezojunction effect more directly as it occurs in the applications: sensors and circuits. At the same time the measurements can validate the predictions of the microscopic model and provide the piezocoefficients of the macroscopic model. Measurements invariably have an uncertainty interval. This interval is the result of many causes, of which the most important ones will be discussed.

Correspondence between the measurements and the device models

The measurements on the transistors yield zero-stress characteristics which can be explained well with the Gummel approach. The Gummel plots of Figure 3.20 and 3.21 show a normal behaviour over many decades of collector current, although the minimum current of the *pnp* transistors is somewhat limited by leakage. The plots were reproducible for different devices of the same run. In addition, they enable an accurate determination of the main parameters of the Gummel model: the saturation current and the temperature.

In the experiments under stress, the piezojunction effect is quite close to the predictions of the macroscopic model:

- Very different curves of transistors of different orientations can be fit with a single series expansion having a cubic symmetry and a reduced number of coefficients;
- The coefficients of the series are found with a reasonable accuracy, especially the first-order coefficients;
- The effect is found to be independent of the bias current and the collector-emitter voltage in the forward bias range;
- The effect is reproducible for different transistors of the same orientation.

In addition, the measurements correspond to predictions from the microscopic model. The data points follow curves that have the same shape and anisotropy as the model curves. They are much more parabolic than the curves

of the piezoresistive effect. Finally, they yield shear piezojunction coefficients which are close to the shear coefficients of the piezoresistive effect.

Accuracy of the piezocoefficients

The accuracy of the piezojunction coefficients is calculated from the error intervals of the input data and the match with the model equation (See Table 3.2). However, it can also be estimated from a comparison with information from the literature.

The piezojunction coefficients have not only been determined in this thesis, but also by F. Fruett [7]-[9]. His values are shown in Table 2.5 and are close to the values of this thesis. Fruett's values are based on measurements on wafers produced with the same IC process, but performed with a different measurement set-up.

The accuracy of the measurements can also be estimated from the measurements of the p-type piezoresistive effect. The latter measurements have yielded piezoresistive coefficients which are comparable with the values found in literature, as can be seen in Table 3.4. The coefficients in the pinch resistors were found to be larger than those in the BW resistors. This was expected in advance because of the lower doping concentration under the pinch diffusion. It should be noted that piezoresistive coefficients in the literature show a large spread. Partially, this spread is related to the doping level and the temperature [10, 11]. Another part of the spread may be related to the measurement accuracy, which is rarely discussed in literature.

In Section 3.4.3 the experimental values of the piezojunction coefficients have been combined with values of hydrostatic pressure found in the literature to separate the values of ζ_{11} and ζ_{44}. This separation is reliable only to a certain extent.

On the one hand, the hydrostatic values were obtained with a better accuracy than the coefficients in this thesis. In addition, they are not perturbed from possible alignment errors with respect to the crystal axes because the hydrostatic pressure maintains the isotropy of the crystal.

On the other hand, the values were only obtained for holes, after which it was assumed that electrons react equally. Only one measurement has been performed with an *npn* transistor, and the pressure steps in this measurement (250 MPa) were as large as the entire stress range in this thesis [12]. In addition, the values were obtained from p^+n diodes, not from *pnp* transistors. In fact, transistors and diodes share the same expression for the saturation current, except that the base width should be replaced by the diffusion length [13]. Stress-induced base width changes can generally be neglected. Diodes and transistors therefore have a comparable stress behaviour only if lifetime changes can also be neglected. It was seen in Section 2.2.4 that this is probable, but not at all proven.

The validity of the separation can be further estimated by the prediction of the microscopic model that ζ_{44} is equal to π_{44}. A comparison of Table 2.5

and 3.4 shows an excellent match for holes. The π_{44} is within the error interval of ζ_{44}, and deviates only 1.2% from the largest piezocoefficient. The comparison is less ideal for electrons, because the ζ_{44} deviates as much as 35% from the largest piezocoefficient. However, it was already argued that the hydrostatic data for electrons is less reliable. In addition, the n-type piezoresistive coefficients have not been measured for this work and were taken from the literature.

Sources of the measurement uncertainty

The experimental determination of the piezojunction effect is quite elaborate and requires knowledge of many parameters. Each of those parameters introduces some error, uncertainty or deviation from the model, as discussed in Section 3.2. These uncertainties have been systematically explored, and are believed to be small compared to the parameters themselves. Nevertheless, some uncertainties are dominant and limit the performance of the measurement set-up.

The main causes of uncertainty are related to the orientation of stress and current. One of those is the parasitic torsion introduced when the beams are glued on a nonideal surface. This torsion dominates the uncertainty for the piezojunction coefficients of electrons. Another uncertainty is caused by the spreading of the current around the edges of the base. This spreading is not taken into account in the analytical model of the current direction. It probably dominates the uncertainty of the piezojunction coefficients of holes. Finally, there is some uncertainty about the orientation of the wafer and the saw lanes with respect to the crystal axes.

There are more causes of uncertainty, but in Section 3.2 they are estimated to be small. For instance, there is some uncertainty in the geometry of the measurement set-up. The dimensions of the beams are not exactly known, and deflection of the beam tip also has a limited accuracy. In addition, the clamping is not perfect because the silicon is more rigid than the surrounding fiber reinforced epoxy.

There is also some uncertainty concerning the electrical measurements. The source-measurement units are not perfect. The leads to the stress-sensitive parts have some parasitic resistance. In addition, the leads may pick up electric and magnetic interference. If these errors have a random, they can rapidly be recognised as noise and repeatability problems. However, Figure 3.22 to 3.27 show that the random errors are not very large and are probably smaller than the systematic errors.

Finally, the uncertainty may be increased by unwanted temperature variations. These variations influence the magnitude of the piezojunction effect, but could also cause an imbalance between the stressed and the reference transistor. In addition, the variations could create thermoelectric voltages in the leads. However, their influence was not perceptible in practice.

4.4 Opportunities for electronic designers

Much of the work in this thesis has improved the understanding of the piezojunction effect, which can be used for the design of sensors and circuits. From a design point of view, the best applicable result is the macroscopic model. The possibilities of this model are discussed in the present section.

General use of the macroscopic model

The macroscopic model of the piezojunction effect is suitable for the most general level of design: circuit synthesis. It is based on analytical equations and has the following features:

- It reproduces the behaviour of the device over the relevant working domain;
- It is mathematically simple and is easily invertible;
- It contains unique parameters which are physically interpretable.

In addition, the structure of the model is very similar to that of the well-known piezoresistive effect.

Since the anisotropy of the piezojunction effect cannot be ignored, sensor and circuit designers should know the primary cause of the effect: the mechanical stress in their particular situation. If this information is not available, they will at least like to know the order of magnitude of the stress and its dominant orientation.

With this knowledge the magnitude of the piezojunction effect can be influenced. For this purpose designers have several degrees of freedom at their disposal. These include the position of the transistors on the chip, the choice for an *npn* or a *pnp* transistor, and the choice for a vertical or a lateral transistor. Sometimes they also include the orientation of the transistor in the wafer plane, and the crystal orientation of the wafer itself.

The piezojunction effect in the most common situations is described in Section 2.4.3. The section especially discusses the anisotropy of the first-order stress sensitivity. This learns that on {100} wafers, vertical *npn* transistors are three times more sensitive to in-plane stress than *pnp* transistors, for any stress orientation. Lateral *pnp* transistors, however, are five times more sensitive than lateral *npn* transistors under those conditions. The latter sensitivity is very anisotropic and has a maximum when the current flows in the direction of the wafer flat and the stress is perpendicular.

The strategy of choosing the optimum configuration for bipolar transistors is quite similar to the strategy used for resistors and MOS transistors. A main difference is that bipolar transistors can be vertical with respect to the wafer plane. In addition, their stress response is less linear. More information is given by Fruett [9], Belu-Marian [14], and Jaeger [15].

Circuit design

In circuit design, unwanted stress may be expected from packaging and processing. An idea of this stress is given in Section 1.3, together with some references. Packaging and processing stress is typically oriented in the plane of the wafer, and in the order of 100 MPa. Near LOCOS and trench isolations the stress can be much larger but it usually decays over a few microns. For a specific design it is useful, of course, to find more detailed information about the packaging and processing stress.

Once the stress is known, the maximum piezojunction effect can be estimated with the macroscopic model of this thesis. For 100 MPa stress the effect can amount to a 6% change in the saturation current, which is equivalent to a change in the base-emitter voltage of 1.5 mV.

If the maximum piezojunction effect is too large to be acceptable, it can be minimised with the strategies described before. It can also be compensated, for example by placing transistors of a differential pair or a current mirror on positions on the chip with an equal stress. However, the effect should not be compensated with a method in which the absolute values of the piezojunction coefficients are required. It was seen already that these values are not precisely known, and may vary with the IC process. An example of such a less-suited method is the compensation of an *npn* transistor with the stress sensitivity of a *pnp*.

The stress effects on circuits could be calculated more accurately if the macroscopic model were integrated in a circuit simulator such as SPICE. The main problem for such an integration would be that the stress tensor would need to be known for each transistor and each resistor separately. This would require a map of the stress distribution on the chip, and the position of each component. Such a map demands a good mechanical characterisation of the fabrication process, yielding the stresses in the different layers and around LOCOS islands and trench isolations. These stresses can e.g. be determined from the curvature of the wafer, the deformation of micromachined structures, the change in resistor values, or by finite element calculations [15]-[20].

Sensor design

It has already been stated in the Introduction that the piezojunction-based sensor is a good candidate that can replace the piezoresistors in pressure sensors, accelerometers, and other sensors in the mechanical domain. It has a similar sensitivity and anisotropy, but a much better suited internal resistance. In addition, a transistor can be smaller than a resistor, in particular its stress-sensitive region. Only its nonlinearity is larger, and its fabrication requires more masks.

Just like piezoresistors, transistors are preferably incorporated in a Wheatstone bridge [14]. Such a bridge, or half bridge, uses a differential measurement to compensate for temperature effects, fabrication-related offset, and electrical interference. It can be biased through either the base-emitter voltage or the

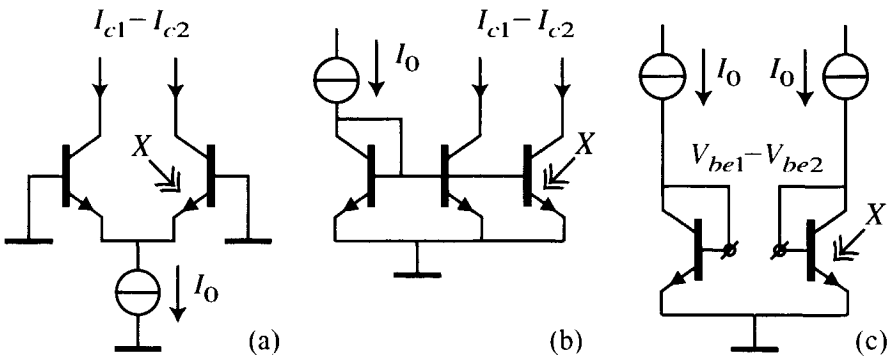

Figure 4.3: Examples of transistor pairs in half bridge circuits which can be used as stress sensors. One of the transistors is stressed, whereas the other serves as a reference. (a) Differential pair, (b) current mirror, (c) translinear circuit.

emitter current. In the first case the stress is translated to a collector current change, and in the second case to a base-emitter voltage change. Some half bridges with these features are shown in Figure 4.3.

4.5 Possible future work

A well-known feature of research is that it is never finished: one answer generates many new questions. The research presented here leaves questions which mainly fall into one of the following categories:

- How can the accuracy of the theory and the measurements be improved?
- What are the remaining values of the coefficients of the new model?
- How will the new model work in practice?

Improvement of the microscopic model

The microscopic model of the piezojunction effect in this thesis is elaborate and could be improved on several points.

The first point concerns the employed model of electron scattering. This model is based on the relaxation time approximation (see Appendix C.1) which almost certainly overestimates the efficiency of a scattering event to restore the thermodynamic equilibrium [21]. In addition, the model describes various causes of scattering with a simple power of the energy. However, any improvement of the model directly increases its complexity and requires much more computing power. Such improvements have been the subject of some dedicated studies, aimed at the description of strained-layer transistors [6]-[23]. These transistors

have become attractive because the minority conductivity in their bases is increased by strain, and therefore the cut-off frequency is increased as well [24]. The employed strain levels, however, are more than ten times higher than those in this thesis.

The second point of improvement could be found in the use of Fermi-Dirac statistics instead of Maxwell-Boltzmann statistics. This approach would have the disadvantage that the Fermi level can no longer be determined analytically. As a result, the evaluation time of the model would be multiplied with the number of steps of the newly required iteration.

A third point of improvement concerns the input parameters of the numerical predictions. These band parameters and deformation potentials have been determined by others and vary from author to author. The predictions in this thesis were seen to be very sensitive to these variations. Probably, more accurate parameters will lead to better predictions.

Improvement of the macroscopic model

Especially the improvement of the macroscopic model will require more experimental work. In the first place, measurements should be set up to find the piezojunction coefficient ζ_{11} independently from ζ_{44} and ζ_{12}. This requires the application of a stress with a component parallel to the current direction. The actual bending stress should therefore be replaced by, for instance, hydrostatic pressure.

In the second place, the piezojunction coefficients should be determined under other circumstances. Such circumstances include different IC processes, but also variations in doping and temperature. The piezo*resistive* effect was seen to be quite sensitive to those variations.

In the third place, models should be made of the influences of stress on secondary parameters of the transistor, such as the current amplification, the base resistance, the Early voltage, and the junction capacitances. These influences are probably not very significant. However, their modelling would allow a clearer picture of the influence on the primary parameter: the saturation current.

Improvement of the measurement set-up

It would be interesting to see how an improvement of the measurement set-up would reduce the uncertainty in the piezojunction coefficients. For the *npn* transistors this uncertainty is dominated by a parasitic torsion that occurs when the beam is bent. One can reduce the torsion by aligning the beam better while it is attached to the clamp. In addition, it can be reduced by replacing the cylinder on the deflected end by a small sphere.

For the *pnp* transistors the uncertainty is dominated by the edge effects in the base, through which the direction of the current is influenced. These can be reduced by making the area of the base-emitter junction larger with respect to the base width. As a penalty the exponential behaviour of the collector current will be limited to lower base-emitter voltages. This need not be a problem

because the collector currents can be measured accurately over a very wide range.

A higher accuracy could further be obtained by temperature stabilisation, biasing through the current instead of the voltage, and the use of coherent detection. The latter principle greatly reduces the influence of noise and drift. It requires, however, a dynamic deflection of the beam tip and thereby complicates the measurement set-up. In addition, the stress values could be calibrated by relating the displacement of the beam tip to the required force.

Finally, the set-up would be greatly improved if the beam carriers had a simpler construction. In the version described in Section 3.2.3, the carriers are difficult to assemble and very vulnerable during wire bonding. The yield of the assembly step was only 50%. In a new version, the bond pads on the beam carrier should lie higher than the beams. In addition, the beams should be directly attached to the printed circuit board to which they are wire bonded. Finally, they should be supported better during the bonding to avoid vibrations.

Implementation in device simulators

The last few years have seen a large development in the field of device simulators. These simulators predict the behaviour of a transistor or a sensor from its structure, external sources, and the local laws of the material behaviour. This behaviour often concerns different physical domains and may include the electrical conduction, the elasticity, and the thermal conduction. The laws are described by differential equations, which are generally solved by a finite-element method. This thesis provides equations to finally implement the stress dependence of the minority conductivity in a useful way.

Design of sensors based on the piezojunction effect

It was shown in the preceding section that the work of the piezojunction effect can be improved on different points. However, before making these improvements it seems advisable to gain more experience with the models. This will sort out the points of which the improvement which will be the most effective. More experience with the models could very well be gained in the design process of a mechanical sensor. Except for the experience, the sensor could also be interesting for its competitive properties, especially when it is made with micromachining techniques. The response of such a sensor can be predicted now, so it must be a pleasure to design it.

4.6 Conclusions

The piezojunction effect has been modelled in this thesis for moderate stress levels of any orientation with respect to the silicon crystal. This modelling starts at the atomic level and ends with the current-voltage characteristics of a transistor. Compared to previous models, it includes some refinements which are essential for the stress range considered. In addition, the model has been

abstracted to a new, macroscopic model. This macroscopic model closely describes the measurements, performed to characterise the effect experimentally. The new model is very suitable for the design of sensors and circuits, where optimum orientations of the current and stress have to be chosen.

Bibliography

[1] J. J. Wortman, J. R. Hauser, and R. M. Burger, Effect of mechanical stress on p-n junction device characteristics, *J. Appl. Phys., 35* (1964) 2122-2131.

[2] W. Wlodarski and B. Moeschke, The effect of hydrostatic pressure on the characteristics of the forward biased p-n junctions, *Electron Technol., 13 (3)* (1980) 3-42.

[3] Y. Kanda, Effect of stress on germanium and silicon p-n junctions, *Japan. J. Appl. Phys., 6* (1967) 475-486.

[4] Y. Kanda, Effect of compressive stress on silicon bipolar devices, *J. Appl. Phys., 44* (1973) 389-393.

[5] C. K. Kim, M. Cardona, and S. Rodriguez, Effect of free carriers on the elastic constants of p-type silicon and germanium, *Phys. Rev. B, 13* (1976) 5429-5441.

[6] J. E. Dijkstra and W. Th. Wenckenbach, Hole transport in strained Si, *J. Appl. Phys., 81* (1997) 1259-1263.

[7] F. Fruett and G. C. M. Meijer, Measurement and compensation of piezoresistive coefficient π_{44} for minority-carrier concentration, *Electron. Letters, 36* (2000) 173-175.

[8] F. Fruett and G. C. M. Meijer, Exploration of the piezojunction effect using PNP lateral transistors on [100] silicon, *Proc. 14th Eur. Conf. Solid-State Transducers (Eurosensors XIV),* Copenhagen, Denmark, 27-30 August 2000, pp. 287-288.

[9] F. Fruett, *The piezojunction effect in silicon, its consequences and applications for integrated circuits and sensors,* Ph. D. Thesis of the Delft University of Technology, Delft University Press, Delft, The Netherlands, 2001.

[10] K. Matsuda, K. Suzuki, K. Yamamura, and Y. Kanda, Nonlinear piezoresistive coeficients in silicon, *J. Appl. Phys., 73* (1993) 1838-1847.

[11] Y. Kanda, A graphical representation of the piezoresistance coefficients in silicon, *IEEE Trans. El. Dev., ED-29* (1982) 64-70.

[12] A. Jayaraman, M. E. Sikorski, J. C. Irvin, and G. H. Yates, Effect of hydrostatic pressure on p-n junction characteristics and the pressure variation of the band gap, *J. Appl. Phys., 38* (1967) 4454-4459.

[13] S. M. Sze, *Physics of semiconductor devices,* Wiley, New York, 1981.

[14] A. Belu-Marian, E. Candet and A. Devenyi, *Piezoresistive sensors, Ch. 3 in: Thin film resistive sensors,* C. Ciureanu and S. Middelhoek (eds.), IOP Publishing, Bristol, 1992.

[15] R. C. Jaeger, J. C. Suhling, R. Ramani, A. T. Bradley, and J. Xu, CMOS stress sensors on (100) silicon, *IEEE J. Solid-St. Circ., 35* (2000) 85-95.

[16] M. Madou, *Fundamentals of microfabrication,* CRC Press, Boca Raton, Florida, 1997.

[17] J. C. Suhling and R. C. Jaeger, Silicon piezoresistive stress sensors and their application in electronic packaging, *IEEE Sensors J., 1* (2001) 14-30.

[18] P. Smeys, P. B. Griffin, Z. U. Rek, E. de Wolf, and K. C. Saraswat, Influence of process-induced stress on device characteristics and its impact on scaled device performance, *IEEE Trans. Electron Devices, 46* (1999) 1245-1252.

[19] H. Miura and Y. Tanizaki, Effect of process-induced mechanical stress on circuit layout, *Proc. Simul. Semicond. Dev. Processes 95, Vol. 6, 1995, pp. 147-150.*

[20] L. Egley and D. Chidambarrao, Strain effects on device characteristics: implementation in drift-diffusion simulators, *Solid-St. Electron., 36* (1993) 1653-64.

[21] N. M. Ashcroft and N. D. Mermin, *Solid state physics, Ch. 16,* Saunders College, Philadelphia, 1976.

[22] J. M. Hinckley and J. Singh, Monte Carlo studies of ohmic hole mobilitiy in silicon and germanium: examination of the optical phonon deformation potential, *J. Appl. Phys., 76* (1994) 4192-4200.

[23] J. M. Hinckley and J. Singh, Hole transport theory in pseudomorphic $Si_{1-x}Ge_x$ alloys grown on Si(100) substrates, *Phys. Rev. B, 41* (1990) 2912-2926.

[24] J. Beisswanger and H. Jorke, Electron transport in bipolar transistors with biaxially strained base layers, *IEEE Trans. Electron Dev., 43* (1996) 62-69.

Appendix A

Stress, Strain, and Elasticity Tensors

The concepts of stress and strain are at the base of the models of the piezojunction and piezoresistive effects, and are therefore applied throughout this thesis. They are widely used in mechanical and civil engineering, but are not necessarily known in electronics. In addition, they require a description in tensors, especially for crystals, and this is not a very popular topic of mathematics.

This appendix therefore resumes the basic theory on stress, strain, and elasticity in crystals. It also discusses some mathematical properties of their tensor description.

A.1 Stress definitions

A body in which one part exerts a force on neighbouring parts is said to be *in a state of stress* [1]. The force acts on a surface of the body. *Stress* is defined as the force per unit area [2]. The surface as well as the force have an orientation that can be resolved in three independent Cartesian components. This is illustrated in Figure A.1.a, where an infinitesimal unit cube of the body has been drawn with all possible forces on the visible surfaces. As a result, the stress can be represented by nine numbers, united in the stress tensor X_{kl}. The subscripts k and l may take the values 1, 2, or 3, and represent the independent directions x_1, x_2, and x_3. In most cases the stress can be represented by only six numbers. This happens when the acceleration of the body is zero. The surface forces then balance each other to prevent translation and rotation and keep the body in statical equilibrium. In that case the stress tensor X_{kl} is symmetrical.

In the description of stress some more definitions are used. If the force on a surface is normal to the surface, the stress component is said to be *normal*. This is the case with X_{11}, X_{22}, and X_{33}. If it acts in the plane of the surface it is called *shear* stress, which is the case with X_{12}, X_{23}, etc. A stress component is defined as *positive* if the outer normal of the surface and the associated force

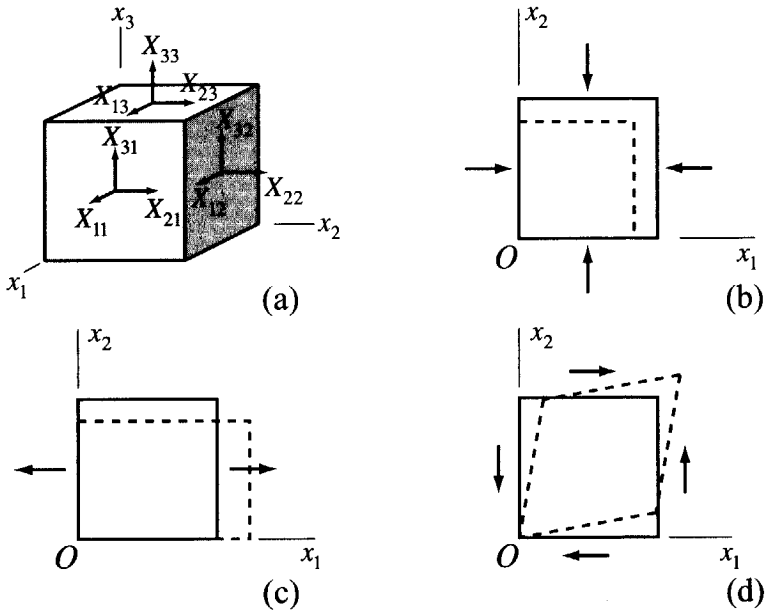

Figure A.1: (a) Components of mechanical stress X_{ij} acting on a unit cube in a stressed body; Two-dimensional view of special stress states and their effect on the shape of the cube: (b) Hydrostatic stress; (c) Uniaxial stress; (d) Pure shear stress.

point both in a positive or both in a negative direction. Normal stresses are therefore positive if they create tension in the body, and negative if they create compression.

A special state of stress arises if all normal components are equal and negative, and all shear components are zero. This is called *hydrostatic stress*, and is shown in Figure A.1.b. The state in which only one normal component is present is called *uniaxial stress*, and is shown in Figure A.1.c. Finally, the state in which only shear components are present is calles *pure shear stress*, and is shown in Figure A.1.d.

A.2 Strain definitions

Stress in a body generally leads to a displacement of a point with respect to its position in a stress-free situation [3]. If this displacement concerns the entire body it is called a rigid-body motion. However, if it is different for two points on the body it must include a deformation, which is called *strain*.

Strain can be divided, in analogy to stress, into normal and shear components. *Normal strain* is defined as the change in the length of a line segment

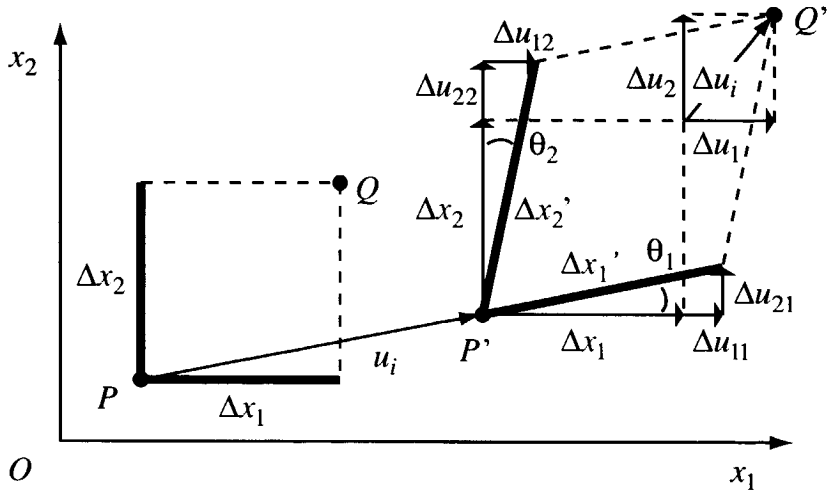

Figure A.2: Strain of a unit square in a body due to the application of stress. The point P is displaced to P^*, but with another amount than Q in its motion to Q^*.

divided by its original length, as a result of the displacement. *Shear strain* is defined as the angular change between two line segments which were originally perpendicular.

Normal and shear strain can be specified mathematically with the aid of Figure A.2. This figure represents a unit square inside a body which is displaced. The displacement moves point P with a vector u_i to the position P'. In addition, it moves point Q with a vector u_i' to position Q'. The difference between u_i' and u_i is the differential displacement Δu_i, consisting of the deformation and some rigid-body rotation. It is the sum of the deformations Δu_{i1} and Δu_{i2} of the perpendicular line segments Δx_1 and Δx_2. For a small square, the deformation may be represented by a first-order Taylor expansion:

$$\Delta u_1 = \Delta u_{11} + \Delta u_{12} = \frac{\partial u_1}{\partial x_1}\Delta x_1 + \frac{\partial u_1}{\partial x_2}\Delta x_2 \tag{A.1}$$

$$\Delta u_2 = \Delta u_{21} + \Delta u_{22} = \frac{\partial u_2}{\partial x_1}\Delta x_1 + \frac{\partial u_2}{\partial x_2}\Delta x_2 \tag{A.2}$$

If the deformation is small enough, it can be used to find an expression for the strain components e_{ij}. Starting from its definition, the small deformation of the sides of a cube, the normal strain e_{11} can be written as:

$$e_{11} = \frac{\Delta x_1' - \Delta x_1}{\Delta x_1} \approx \frac{\Delta u_{11}}{\Delta x_1} = \frac{\partial u_1}{\partial x_1} \tag{A.3}$$

The shear strain e_{12}, in addition, is the average rotation of the sides of the unit square:

$$\begin{aligned} e_{12} &= \frac{1}{2}(\theta_1 + \theta_2) \\ &= \frac{1}{2}\left[\arctan\left(\frac{\Delta u_{21}}{\Delta x_1 + \Delta u_{11}}\right) + \arctan\left(\frac{\Delta u_{12}}{\Delta x_2 + \Delta u_{22}}\right)\right] \\ &\approx \frac{1}{2}\left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2}\right) \end{aligned} \tag{A.4}$$

This can be repeated for the other orientations, which yields the complete strain tensor e_{ij}. The deformed sides of a unit cube $\Delta x_i'$ may therefore be expressed in terms of the undeformed sides as:

$$\Delta x_i' = \sum_{j=1}^{3} (\delta_{ij} + e_{ij})\, \Delta x_j \tag{A.5}$$

where δ_{ij} is the Kronecker delta.

A.3 Tensor properties

Stress and strain are both described by second-rank tensors. Before describing the relation between them, it is useful to discuss tensor notation and some of their mathematical properties.

A.3.1 Einstein summation convention

The notation of tensor equations is simplified by the use of the *Einstein summation convention*, also called the *dummy subscript notation* [1]. If two vectors $\mathbf{p}$ and $\mathbf{q}$ are related by a matrix $\mathbf{T}$, the components of $\mathbf{p}$ are generally given by the sums:

$$p_1 = \sum_{j=1}^{3} T_{1j} q_j \ , \quad p_2 = \sum_{j=1}^{3} T_{2j} q_j \ , \quad p_3 = \sum_{j=1}^{3} T_{3j} q_j \tag{A.6}$$

This can be written more compactly as:

$$p_i = \sum_{j=1}^{3} T_{ij} q_j \ , \quad i \in \{1, 2, 3\} \tag{A.7}$$

It is also possible to leave out the summation sign:

$$p_i = T_{ij} q_j \ , \quad i \in \{1, 2, 3\} \tag{A.8}$$

This is an example of the Einstein convention, which can be formulated in general as follows:

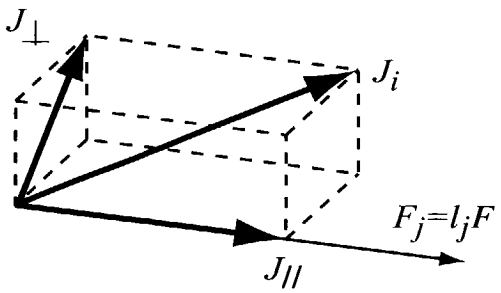

Figure A.3: Directions of the current density and the electric field in an anisotropic material. The current density vector J_i can be resolved in a parallel and a perpendicular component with respect to the field vector F_j.

> *If a letter subscript occurs twice in the same term, summation with respect to that subscript is to be automatically understood.*

Subscripts of this kind are called *dummy subscripts*. Subscripts occurring only once in a term are called *free subscripts*. The same free subscript should appear in every term of a tensor equation, and on both sides of the '=' sign. Since they may take any of the values 1, 2, or 3, they can describe a whole vector. A vector **p** is therefore frequently written as p_i.

A.3.2 Scalar description in an anisotropic case

It often occurs that a tensor quantity is spoken of as a scalar, although the material is anisotropic. In stressed silicon, for instance, one sometimes speaks about 'the' conductivity σ in a particular direction. What is usually meant is the *longitudinal conductivity*, which is the ratio of the current component along the field $J_{//}$ and the magnitude of the field F. This is illustrated in Figure A.3. If the field direction is indicated by the unit vector l_i, the field vector is equal to l_iF. Meanwhile, the current component $J_{//}$ is given by the inner product of l_i with the current vector J_i. The relation between the current and field vectors is given by the conductivity tensor σ_{ij}. This yields:

$$\sigma = \frac{j_{//}}{F} = \frac{l_i j_i}{F} = \frac{l_i \sigma_{ij} F_j}{F} = \frac{l_i \sigma_{ij} l_j F}{F} = l_i l_j \sigma_{ij} \tag{A.9}$$

which is a scalar quantity. In the same way it is possible to derive 'the' Young's modulus, discussed in Section A.4.

A.3.3 Transformation of axes

Vectors and tensors are specified with respect to particular axes, but sometimes need to be rewritten with respect to other axes. This requires a *orthogonal transformation* [1, 4]. The transformation changes the old mutually perpendicular

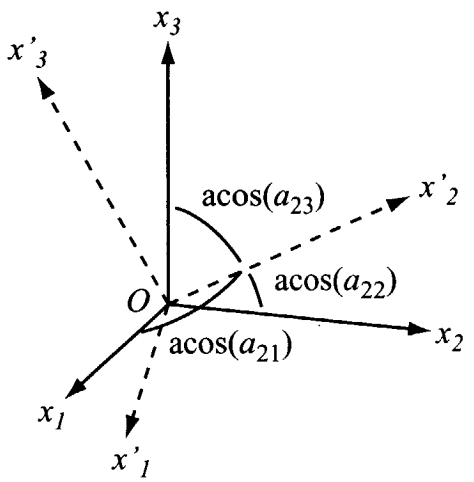

Figure A.4: Orthogonal transformation of the axes system x_j to the system x'_i. The angles of x'_2 with respect to the x_j-axes are indicated.

set of axes x_j to a new orthogonal set x'_i with the same origin; see Figure A.4. It is represented by the matrix a_{ij}, which consists of the direction cosines between the axes of the new set x'_i and those of the old set x_j:

$$\begin{aligned} x'_i &= a_{ij} x_j \\ &= \cos(x'_i, x_j)\, x_j \ , \qquad i,j \in \{1,2,3\} \end{aligned} \tag{A.10}$$

Geometrically, a_{ij} is the projection of x'_i on x_j. Its rows are the vectors x'_i expressed in the old set, x_j. Its columns, however, are the vectors x_j expressed in the new axes set x'_i. As a result, a_{ij} is a *unitary matrix*, which means that the transpose of a_{ij} is equal to its inverse:

$$a_{ji} = a_{ij}^{-1} \tag{A.11}$$

The transformation matrix can be used to write any tensor property with respect to the new axes. For this purpose the following multiplications are to be applied to a vector p_i, a second-rank tensor T_{ij}, and a fourth-rank tensor S_{ijkl}:

$$\begin{aligned} p'_i &= a_{ij} p_j \\ T'_{ij} &= a_{ik} a_{jl} T_{kl} \\ S'_{ijkl} &= a_{im} a_{jn} a_{ko} a_{lp} S_{mnop} \end{aligned} \tag{A.12}$$

A.3.4 Principal values

The appearance of a tensor depends on the chosen axes. A second-rank tensor can nevertheless be represented in a unique way by its *principal values*. These

principal values are the eigenvalues of the tensor matrix, whereas the accompanying eigenvectors point in directions which are called the *principal axes* of the tensor. With respect to the principal axes the tensor is a diagonal matrix with the eigenvalues on the main diagonal.

The tensor T_{ij}, for example, may have the following shape with respect to the axes x_i:

$$T_{ij} = \begin{bmatrix} 5 & \sqrt{3} & 0 \\ \sqrt{3} & 7 & 0 \\ 0 & 0 & 12 \end{bmatrix} \tag{A.13}$$

When T_{ij} is diagonalised, the eigenvalues 4, 8, and 12 are found, as well as the following normalised eigenvectors x'_i [4]:

$$x'_1 = \frac{1}{2}\begin{bmatrix} \sqrt{3} \\ -1 \\ 0 \end{bmatrix}, \quad x'_2 = \frac{1}{2}\begin{bmatrix} 1 \\ \sqrt{3} \\ 0 \end{bmatrix}, \quad x'_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{A.14}$$

These eigenvectors can be used to construct the transformation matrix a_{ij}, transforming x_i to the principal axes of T_{ij}:

$$a_{ij} = \begin{bmatrix} -x'_1- \\ -x'_2- \\ -x'_3- \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{3} & -1 & 0 \\ 1 & \sqrt{3} & 0 \\ 0 & 0 & 2 \end{bmatrix} \tag{A.15}$$

The application of a_{ij} to T_{ij} indeed yields a diagonal matrix T'_{ij} with the eigenvalues on the main diagonal:

$$T'_{ij} = a_{ik}a_{jl}T_{kl} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 12 \end{bmatrix} \tag{A.16}$$

A.4 Elasticity

If a body is deformed by mechanical stress and the stress is then removed, it usually returns to its old shape if the deformation was not too large. This response is called *elastic*. It often obeys *Hooke's law*, stating that the amount strain is linearly proportional to the stress magnitude for sufficiently small stresses.

Monocrystalline silicon and germanium are very elastic and obey Hooke's law closely up to the fracture stress. However, this elasticity is anisotropic because of the crystalline material structure [1, 5]. This means that Hooke's law should be written in its most general form, by using the fourth-rank *elastic compliance* tensor S_{ijkl}:

$$e_{ij} = S_{ijkl}X_{kl} \tag{A.17}$$

Alternatively it can be written inversely, expressing the stress as a function of strain through a tensor C_{ijkl}. This tensor is called the *elastic stiffness*.

The number of terms in the equation is 81, but can fortunately be reduced by using the symmetries of the stress and the cubic crystal. This is discussed more in detail in Section 2.3.3. It appears that only three coefficients of S_{ijkl} are nonzero and independent: S_{1111}, S_{1122}, and S_{1212}, defined with respect to the crystal axes. These coefficients are known from experiments to a good accuracy [5, 7, 6].

The general law of Hooke can be used to define *Young's modulus* of the material in a specific direction [5]. Young's modulus E is defined as the uniaxial stress in a direction divided by the deformation in the same direction. If this direction is chosen as the x'-axis, E can be written as:

$$E = \frac{X'_{11}}{e'_{11}} = \frac{1}{S'_{1111}} \tag{A.18}$$

It should be noted that the compliance coefficient S'_{1111} can be related to the axes of the crystal with the aid of Equation A.12.

A.5 Conclusions

Stress is defined as the force per unit area inside a body. Strain is a measure of the deformation of the body. Both quantities are represented by second-order tensors, of which the appearance depends on the set of axes in which they are defined. If the body is linearly elastic, the relation between stress and strain is given by the compliance coefficients. In crystalline materials such as silicon the compliance is anisotropic and needs to be described by a fourth-rank tensor. The complexity of this tensor can be greatly reduced by using symmetry considerations.

Bibliography

[1] J. F. Nye, *Physical properties of crystals, 2nd ed.*, Clarendon Press, Oxford, 1985.

[2] J. M. Gere and S.P. Timoshenko, *Mechanics of materials,* Chapman and Hall, London, 1991.

[3] J. W. Dalley and W. F. Riley, *Experimental stress analysis,* 2nd ed., McGraw-Hill Kogakusha, Tokyo, 1978.

[4] G. Strang, *Linear algebra and its applications,* Harcourt Brace Jovanovich, San Diego, 1988.

[5] J. J. Wortman and R. A. Evans, Youngs modulus, shear modulus, and Poisson ratio in silicon and germanium, *J. Appl. Phys., 36* (1965) 153-156.

[6] O. Madelung (ed.), *Semiconductors - Basic data,* Springer, Berlin, 1996.

[7] P. J. Burnett, *Toughness of unimplanted and ion-implanted Si,* in: *Properties of silicon, EMIS Datareview series 4,* INSPEC, London, 1988, pp. 30-32.

Appendix B

Band Structure of Cubic Semiconductors

The band structure of a solid describes the energy of an electron in the solid as a function of its wave vector. Together with the Fermi-Dirac statistics it determines the entire electron behaviour [1]. It replaces the description of the electron as a quantummechanical wave by that of a (semi)classical particle. As a result, the wave vector $\mathbf{k}$ can be considered as the electron momentum and the band structure $E_n(\mathbf{k})$ as its equations of motion. These equations are also known as 'energy bands', 'dispersion relations', or 'band equations'.

The band structure determines many properties of the solid, and in particular the electrical conductivity. The conductivity is to be mathematically derived in Appendix C. The band structure therefore needs to be known as a set of analytic equations. These equations are found here for the cubic semiconductors silicon and germanium.

The band theory starts from the assumption that a solid is a large single molecule in which the electrons are not bound to individual atoms but can move freely. Moreover, it is assumed that:

- The solid is a perfectly periodic crystal lattice;
- The sites of the crystal have fixed equilibrium positions;
- The influence of all electrons on a single electron is described through some average periodic field.

On the basis of these assumptions various methods have been developed for actually calculating band structures [1, 2]. These have resulted in quite reliable pictures of the bands of many materials. Those of silicon and germanium are shown in Figure B.1.

Most methods for calculating bands are not very practical to determine the conductivity of semiconductors. They are very complex but yield on the other

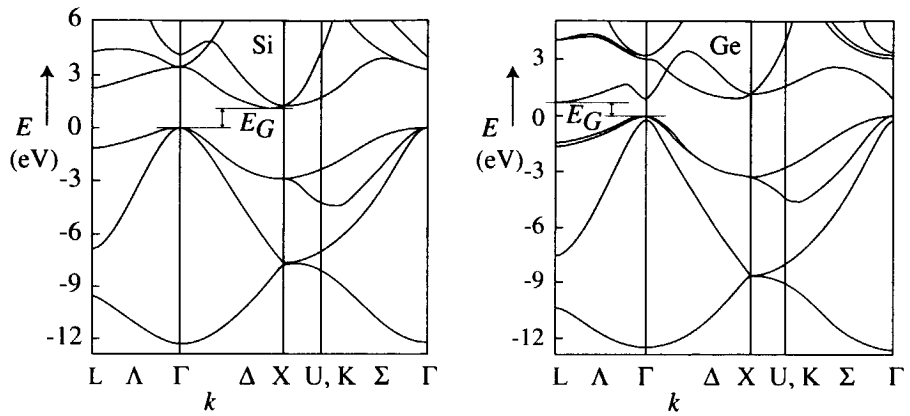

Figure B.1: Band structure of silicon and germanium. Energy E vs. wave number k in different directions of the first Brillouin zone. E_G is the bandgap. For silicon the spin-orbit splitting is imperceptible on this scale. After Madelung [13].

hand more information than necessary. Actually, the conductivity is determined by the few energy bands which are partially filled: the valence and the conduction bands. More specifically, it is determined by the extremal points of those bands, where some electrons are missing from a full band or just populate a band which is further empty. It is therefore sufficient to only describe the band structure around these extrema. A useful description of the band extrema is given by the $\mathbf{k}\hat{\mathbf{p}}$ method. It will be discussed here, on the base of the information found in Reference [1, 3, 4, 5, 6], and [7].

B.1 Overview of the $\mathbf{k}\hat{\mathbf{p}}$-method

The $\mathbf{k}\hat{\mathbf{p}}$-method yields analytical equations of the semiconductor band extrema. For the conduction band they describe the well-known ellipsoids, whereas for the valence band they describe more complicated warped spheres. The equations rely on experimental knowledge of the material: the band gaps, the effective masses at the extrema, and the symmetry of the semiconductor crystal. They can be easily adapted to include the effects of strain, as shown in Section 2.1.

The main idea of the $\mathbf{k}\hat{\mathbf{p}}$-method is to consider electrons in the band extrema as eigenstates of the Hamiltonian operator. However, electrons with a wavevector $\mathbf{k}$ slightly off the extrema are seen as perturbed states. These states are inserted into the Schrödinger equation, after which the eigenvalues can be found by perturbation theory. This yields a characteristic equation with many unknown matrix elements. The number of those elements is strongly reduced, however, by using the symmetry properties of the crystal. The remaining matrix

elements can be determined from experiments. They enter into the characteristic determinant, of which the evaluation yields the explicit equations of the energy bands.

B.2 The kp̂ Hamiltonian

As a first step in the $\mathbf{k\hat{p}}$ method, the Hamiltonian is determined for electrons with wavevectors slightly off the extremal energies.

The state of an electron in band n of a crystal can in general be written as a Bloch function, consisting of a plane wave and a periodic modulating function $u_{n\mathbf{k}}$:

$$\psi_{n\mathbf{k}}(\mathbf{x}) = e^{i\mathbf{k}\cdot\mathbf{x}} u_{n\mathbf{k}}(\mathbf{x}) \tag{B.1}$$

where $\mathbf{x}$ is the position vector and $\mathbf{k}$ the wave vector. This can be inserted into the stationary one-electron Schrödinger equation:

$$\hat{H}\psi_{n\mathbf{k}}(\mathbf{x}) = \left(\frac{\hat{\mathbf{p}}^2}{2m_0} + U(\mathbf{x})\right)\psi_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\,\psi_{n\mathbf{k}}(\mathbf{x}) \tag{B.2}$$

in which $\hat{H}$ is the Hamiltonian, $\hat{\mathbf{p}}$ is the momentum operator $-i\hbar\nabla$, m_0 the free electron mass, $U(\mathbf{x})$ the potential of the periodic crystal, and $E_n(\mathbf{k})$ the total energy.

Equation B.2 can also be written in terms of the modulating function only. The momentum operator is therefore applied to the Bloch function. In general, a product of functions f and g is differentiated as $\nabla^2(fg) = g\nabla^2 f + 2\nabla f\cdot\nabla g + f\nabla^2 g$. Applied to this case, it yields:

$$\left(\frac{\hat{\mathbf{p}}^2}{2m_0} + U(\mathbf{x}) + \frac{\hbar}{m_0}\mathbf{k}\cdot\hat{\mathbf{p}} + \frac{\hbar^2 k^2}{2m_0}\right) u_{n\mathbf{k}}(\mathbf{x}) = E_n(\mathbf{k})\,u_{n\mathbf{k}}(\mathbf{x}) \tag{B.3}$$

If there is information about the band extremum at $\mathbf{k} = \mathbf{k}_0$, it is useful to expand $\mathbf{k}$ into $\mathbf{k} + \mathbf{k}_0 - \mathbf{k}_0$. This yields:

$$\hat{H}_u u_{n\mathbf{k}}(\mathbf{x}) = \left(\hat{H}_0 + \hat{H}_{\mathbf{k}\hat{\mathbf{p}}}\right) u_{n\mathbf{k}_0}(\mathbf{x}) = E'_n(\mathbf{k})\,u_{n\mathbf{k}_0}(\mathbf{x}) \tag{B.4}$$

where $\hat{H}_u$ is the Hamiltonian on the basis of the modulating functions, $\hat{H}_0$ the Hamiltonian at $\mathbf{k}_0$, $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$ is defined as the $\mathbf{k\hat{p}}$ Hamiltonian, and E'_n is a reduced energy. They are given by:

$$\hat{H}_0 = \frac{\hat{\mathbf{p}}^2}{2m_0} + U(\mathbf{x}) + \frac{\hbar}{m_0}\mathbf{k}_0\cdot\hat{\mathbf{p}} + \frac{\hbar^2 k_0^2}{2m_0} \tag{B.5}$$

$$\hat{H}_{\mathbf{k}\hat{\mathbf{p}}} = \frac{\hbar}{m_0}(\mathbf{k} - \mathbf{k}_0)\cdot\hat{\mathbf{p}} \tag{B.6}$$

$$E'_n(\mathbf{k}) = E_n(\mathbf{k}) - \frac{\hbar^2}{2m_0}\left(k^2 - k_0^2\right) \tag{B.7}$$

If $\mathbf{k}$ is relatively close to $\mathbf{k}_0$, $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$ is small and can be mathematically considered as a perturbation on $\hat{H}_0$. This allows us to use perturbation theory and find the energy eigenvalues as a function of $\mathbf{k}$.

B.3 Löwdin perturbation theory

The perturbation theory for bands around an extremum needs to be of second-order and time-independent. In addition, it must include degenerate cases because the valence bands of silicon are degenerate in their maximum. There is a variant of this theory which is especially transparent for this situation and which has been formulated by Löwdin [8].

Applied to the $\mathbf{k}\hat{\mathbf{p}}$ method, it starts by expanding the modulating function $u_{j\mathbf{k}}(\mathbf{x})$ of a Bloch state j at $\mathbf{k}$ into a linear combination of N orthonormal functions $u_{n\mathbf{k}_0}$ at $\mathbf{k}_0$:

$$u_{j\mathbf{k}}(\mathbf{x}) = \sum_{n=1}^{N} c_n u_{n\mathbf{k}_0} \tag{B.8}$$

The coefficients c_n are the Fourier coefficients of the expansion. The functions $u_{n\mathbf{k}_0}$ are exact or approximate eigenstates of the Hamiltonian $\hat{H}_0$ and are supposed to be known.

The linear combination $u_{j\mathbf{k}}$ is inserted into the Schrödinger equation for modulating functions, Equation B.4, multiplied by the complex conjugate $u^*_{m\mathbf{k}_0}$ and integrated over $\mathbf{x}$. This yields the system of linear equations:

$$\sum_{n=1}^{N} \left(H_{mn} - E'_j \delta_{mn}\right) c_n = 0 \tag{B.9}$$

where n ranges over all N states. The H_{mn} are matrix elements and are defined by the following integrals:

$$H_{mn} = \int u^*_{m\mathbf{k}_0} \hat{H}_u u_{n\mathbf{k}_0} d\mathbf{x} = \langle m| \hat{H}_u |n\rangle \tag{B.10}$$

where $\hat{H}_u$ is the full Hamiltonian on the basis of the $u_{n\mathbf{k}_0}$.

The N states are now divided into two classes. The states which should be described fall in class A, and are supposed to be degenerate or to lie close to each other at $\mathbf{k}_0$. All other states form class B and are separated by relatively large energies from the states in class A. This is visualised in Figure B.2. Equation B.9 can therefore also be divided in two classes. If also the diagonal elements are written separately, it reads as:

$$\left(E'_j - H_{mm}\right) c_m = \sum_{n \neq m}^{A} H_{mn} c_n + \sum_{n \neq m}^{B} H_{mn} c_n \tag{B.11}$$

The Fourier coefficients c_m can be written more compactly by defining the off-diagonal matrix elements h_{mn}:

$$h_{mn} = \frac{H_{mn}(1 - \delta_{mn})}{E'_j - H_{mm}} \tag{B.12}$$

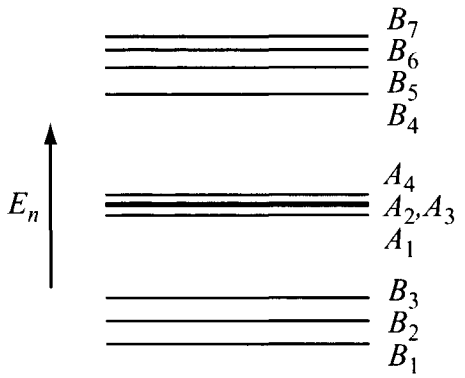

Figure B.2: Division of electron states into two classes. Class A contains one or more states with energy levels E_n close to each other. For these levels a mathematical description is sought. Class B contains all other states, having energy levels far from those of class A. Their actual description is unimportant.

where δ_{mn} is the Kronecker delta. The c_m then reduce to:

$$c_m = \sum_n^A h_{mn} c_n + \sum_n^B h_{mn} c_n \tag{B.13}$$

The sum over B can be brought under the sum over A by an iterative substitution of the c_n in the sum over B by Equation B.13 itself:

$$c_m = \sum_{n \neq m}^A \left(h_{mn} + \sum_b^B h_{mb} h_{bn} + \sum_{\substack{b,c \\ b \neq c}}^B h_{mb} h_{bc} h_{cn} + \cdots \right) c_n \tag{B.14}$$

It is now useful to expand the h_{ij} in this equation into their original quantities, and to take them together in the new matrix U^A_{mn}:

$$U^A_{mn} = H_{mn} + \sum_{b \neq m,n}^B \frac{H_{mb} H_{bn}}{E'_j - H_{bb}} + \sum_{\substack{b,c \neq m,n \\ b \neq c}}^B \frac{H_{mb} H_{bc} H_{cn}}{\left(E'_j - H_{bb}\right)\left(E'_j - H_{cc}\right)} + \cdots \tag{B.15}$$

With this notation, the Fourier coefficients can be compactly written as a sum over the states in A only:

$$c_m = \sum_n^A \frac{U^A_{mn} - H_{mn}\delta_{mn}}{E'_j - H_{mm}} c_n \tag{B.16}$$

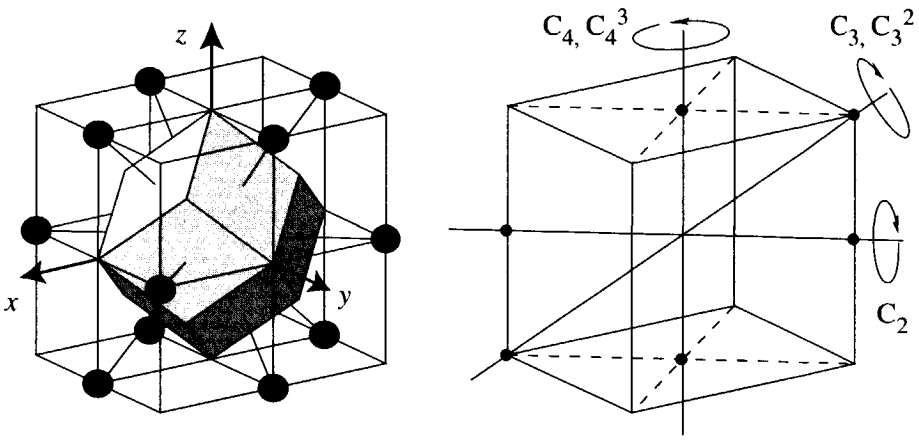

Figure B.3: Wigner-Seitz primitive cell of a face-centered cubic lattice type (left), and the different kinds of symmetry axes of the cubic point group O (right).

If m is chosen within class A, the c_m on the left-hand side of this equation can be brought under the summation sign. After multiplication with $(E'_j - H_{mm})$ this yields:

$$\sum_n^A \left(U^A_{mn} - E'_j \delta_{mn}\right) c_n = 0 \,, \qquad m \in A \tag{B.17}$$

The result of the manipulations is that now, only a limited system of A equations need be solved instead of the infinite system of N equations in Equation B.9. The states in A can be chosen in such way that they correspond to the conduction or the valence bands. These bands are contained by the energies E'_j. According to Equation B.17, they form the eigenvalues of the matrix U^A_{mn}. They can be found by forming the secular determinant and solving:

$$\det\left(U^A_{mn} - E'_j \delta_{mn}\right) = 0 \tag{B.18}$$

The elements of U^A_{mn} are still to be evaluated.

B.4 Cubic symmetry

The evaluation of matrix elements in the $\mathbf{k}\hat{\mathbf{p}}$ method is much simplified by some symmetry considerations from group theory [3, 6]. It is a well-known fact that silicon and germanium crystals have monatomic lattices with tetrahedral bonds. The Bravais lattice of the crystal is therefore a face-centered cubic with a two-point basis (see Figure B.3). This means that its Wigner-Seitz primitive cell

Table B.1: Symmetry operations of the cubic point group O. The effect on a point $(x_1, x_2, x_3) = (123)$ is indicated between the parantheses and includes interchanges and negations of its components.

Class	Symmetry operations
$\hat{E}$	$\hat{S}_1\,(123)$
$3\hat{C}_4, 3\hat{C}_4^3$	$\hat{S}_2\,(1\bar{3}2)\,, \hat{S}_3\,(32\bar{1})\,, \hat{S}_4\,(\bar{2}13)$
	$\hat{S}_5\,(13\bar{2})\,, \hat{S}_6\,(\bar{3}21)\,, \hat{S}_7\,(2\bar{1}3)$
$3\hat{C}_4^2$	$\hat{S}_8\,(1\bar{2}\bar{3})\,, \hat{S}_9\,(\bar{1}2\bar{3})\,, \hat{S}_{10}\,(\bar{1}\bar{2}3)$
$4\hat{C}_3, 4\hat{C}_3^2$	$\hat{S}_{11}\,(312)\,, \hat{S}_{12}\,(\bar{2}3\bar{1})\,, \hat{S}_{13}\,(\bar{3}1\bar{2})\,, \hat{S}_{14}\,(\bar{2}\bar{3}1)$
	$\hat{S}_{15}\,(231)\,, \hat{S}_{16}\,(\bar{3}\bar{1}2)\,, \hat{S}_{17}\,(2\bar{3}\bar{1})\,, \hat{S}_{18}\,(3\bar{1}\bar{2})$
$6\hat{C}_2$	$\hat{S}_{19}\,(21\bar{3})\,, \hat{S}_{20}\,(\bar{2}\bar{1}3)\,, \hat{S}_{21}\,(3\bar{2}\bar{1})$
	$\hat{S}_{22}\,(\bar{3}\bar{2}1)\,, \hat{S}_{23}\,(\bar{1}32)\,, \hat{S}_{24}\,(\bar{1}\bar{3}2)$

has the symmetry of the full cubic group O_h. The group contains the following symmetry elements:

- Rotations around three fourfold axes joining the centres of opposite faces ($3\hat{C}_4$, $3\hat{C}_4^2$ and $3\hat{C}_4^3$);
- Rotations around four threefold axes along the diagonals of the cube ($4\hat{C}_3$ and $4\hat{C}_3^2$);
- Rotations around six twofold axes joining the centres of opposite edges ($6\hat{C}_2$).
- Inversion ($\hat{I}$).

This means that crystals with the full cubic symmetry are transformed into themselves by any of the 24 operations mentioned in Table B.1, and by their inverses.

The symmetry properties of the crystal lattice also applies to the energy bands of the electrons. To support this, two arguments are required.

The first argument is that for a Bloch state $\psi_{n\mathbf{k}}(\mathbf{x})$, a transformation of the crystal lattice corresponds to a similar transformation of the reciprokal lattice. This can be seen by by subjecting the definition of the state to an *arbitrary* transformation $\hat{a}$ of the direct space of the crystal:

$$\hat{a}\psi_{n\mathbf{k}}(\mathbf{x}) = \psi_{n\mathbf{k}}(\hat{a}\mathbf{x}) = \exp(i\mathbf{k}\cdot\hat{a}\mathbf{x})\, u_{n\mathbf{k}}(\hat{a}\mathbf{x}) \tag{B.19}$$

The inner product $\mathbf{k}\cdot\hat{a}\mathbf{x}$ represents the angle between two vectors, and is independent on the axes system in which these vectors are defined. The product is therefore invariant under the transformation, as well as under its inverse $\hat{a}^{-1}$.

It can therefore be expanded as:

$$\begin{aligned}\psi_{n\mathbf{k}}(\widehat{a}\mathbf{x}) &= \exp\left(i\widehat{a}^{-1}\mathbf{k}\cdot\widehat{a}^{-1}\widehat{a}\mathbf{x}\right) u_{n\mathbf{k}}(\widehat{a}\mathbf{x}) \\ &= \exp\left(i\widehat{a}^{-1}\mathbf{k}\cdot\mathbf{x}\right) u'_{n\mathbf{k}}(\mathbf{x}) \\ &= \psi_{n\widehat{a}^{-1}\mathbf{k}}(\mathbf{x})\end{aligned} \tag{B.20}$$

where $u'_{n\mathbf{k}}(\mathbf{x})$ is another modulating function.

The second argument is that the Hamiltonian $\widehat{H}$ is invariant under *symmetry* transformations of the *crystal* $\widehat{S}$. The Hamiltonian contains the Laplace operator ∇^2 and the potential energy $U(\mathbf{x})$. The operator ∇^2 is defined as an inner product of two vectors: $\nabla^2 = \nabla\cdot\nabla$. It depends on $\mathbf{x}$, but not on the coordinate system. The scalar $U(\mathbf{x})$ is the periodic potential of the crystal lattice and therefore possesses all the lattice symmetry. The entire Hamiltonian therefore transforms as the crystal.

The above arguments can now be applied to the Schrödinger equation for an electron state $\psi_{n\mathbf{k}}$. Transforming this equation with the crystal symmetry $\widehat{S}$ yields:

$$\begin{aligned}\widehat{H}\left(\widehat{S}\mathbf{r}\right)\psi_{n\mathbf{k}}\left(\widehat{S}\mathbf{x}\right) &= E_n(\mathbf{k})\,\psi_{n\mathbf{k}}\left(\widehat{S}\mathbf{x}\right) \\ \widehat{H}(\mathbf{x})\,\psi_{n\widehat{S}^{-1}\mathbf{k}}(\mathbf{x}) &= E_n(\mathbf{k})\,\psi_{n\widehat{S}^{-1}\mathbf{k}}(\mathbf{x})\end{aligned} \tag{B.21}$$

This means that if $\psi_{n\mathbf{k}}(\mathbf{x})$ is an eigenstate of the Hamiltonian, the transformed state $\widehat{S}\psi_{n\mathbf{k}}(\mathbf{x})$ is also an eigenstate, corresponding to the same energy. On the other hand, the transformed state must satisfy its 'own' Schrödinger equation:

$$\widehat{H}(\mathbf{x})\,\psi_{n\widehat{S}^{-1}\mathbf{k}}(\mathbf{x}) = E_n\left(\widehat{S}^{-1}\mathbf{k}\right)\psi_{n\widehat{S}^{-1}\mathbf{k}}(\mathbf{x}) \tag{B.22}$$

A comparison with the former equation learns that:

$$E_j(\mathbf{k}) = E_n\left(\widehat{S}^{-1}\mathbf{k}\right) \tag{B.23}$$

This means that the energy bands in the reciprokal space have the symmetr properties of the crystal in the direct space.

B.5 Conduction band equations

In Figure B.1 it can be seen that the conduction band of silicon has a minimum along the [100]-direction at about 85% of the distance to the boundary of the first Brillouin zone. Due to the cubic symmetry of the crystal, this minimum exists in all six {100}-directions. The conduction band of germanium has a minimum along the eight {111}-directions at the zone boundary.

For a minimum located at $\mathbf{k}_0$, the energy in the vicinity may be found with perturbative approximation of Equation B.17. Class A consists in that case of the conduction band in its extremum, which is nondegenerate and therefore

associated with only one state $u_{1\mathbf{k}_0}$. As a result, the subscripts m and n are equal to 1, and E_j is equal to E_C, the conduction band away from the minimum. Equation B.17 therefore reduces to:

$$U_{11}^{C0} = E'_C \tag{B.24}$$

The matrix element U_{11}^{C0} can be worked out with the aid of Equation B.15 and the split-up Hamiltonian of Equation B.4. It should be noted that $u_{1\mathbf{k}_0}$ is the eigenfunction of $\hat{H}_0$ with the energy eigenvalue E_{C0}. This energy is usually called the band edge. To the second order this yields:

$$E'_C(\mathbf{k}) = E_{C0} + \langle 1|\, \hat{H}_{\mathbf{k}\hat{\mathbf{p}}}\, |1\rangle + \sum_{b\neq 1}^{B} \frac{\left|\langle 1|\, \hat{H}_{\mathbf{k}\hat{\mathbf{p}}}\, |b\rangle\right|^2}{(E'_C - E_b)} \tag{B.25}$$

Since the energy is in a minimum at $\mathbf{k}_0$, it should apply that $\nabla_{\mathbf{k}} E_n(\mathbf{k}_0) = 0$. This means that the first-order term $\langle 1|\, \hat{H}_{\mathbf{k}\hat{\mathbf{p}}}\, |1\rangle$ is zero.

The second-order term can be worked out by inserting the expression for $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$ from Equation B.6 and by adding the difference between E'_C and E_C, given by Equation B.7. This turns Equation B.25 into the dispersion relation:

$$E_C(\mathbf{k}) = E_{C0} + \frac{\hbar^2}{2} M_{\alpha\beta} k'_\alpha k'_\beta, \qquad \alpha, \beta \in \{1, 2, 3\} \tag{B.26}$$

where k'_α are the elements of the vector $\mathbf{k} - \mathbf{k}_0$ and $M_{\alpha\beta}$ the elements of the inverse effective mass tensor. In this relation the Einstein convention has been used so that a summation with respect to the subscripts α and β is to be understood (see Section A.3.1). The inverse mass tensor is defined by

$$M_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{m_0} + \frac{1}{m_0^2} \sum_{b\neq 1}^{B} \left. \frac{p_\alpha^{1b} p_\beta^{b1} + p_\beta^{1b} p_\alpha^{b1}}{E_{C0} - E_b} \right|_{\mathbf{k}_0} \tag{B.27}$$

where p_α^{1b} is the α-component of the vector integral $\langle 1|\, \hat{\mathbf{p}}\, |b\rangle$. It should be noted that E'_C has been replaced here by E_{C0}. This is allowed when the 'perturbation' by $\mathbf{k}$ is small on the scale of the band gaps.

Equation B.26 represents the electron energy in the crystal as a function of its wave vector $\mathbf{k}$. For an electron *in free space* the energy would be given by:

$$E = \frac{\hbar^2 k^2}{2m_0} = \frac{\hbar^2}{2} \frac{\delta_{\alpha\beta}}{m_0} k_\alpha k_\beta \tag{B.28}$$

By comparing the equations it appears that an electron in a crystal has an energy shifted by E_{C0} and a wave vector shifted by $\mathbf{k}_0$. It also has a 'mass' that differs from the free-electron mass by the 'interaction' with other electron states.

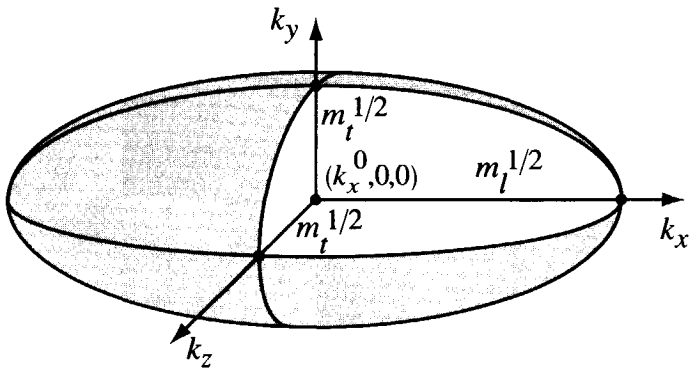

Figure B.4: Ellipsoid representing the equi-energy surfaces of the second-order description of the conduction band minima. m_l and m_t are the longitudinal and transverse effective masses, respectively.

The inverse effective mass tensor can also be written as a function of the band energy $E_C(\mathbf{k})$. For this purpose it is differentiated twice with respect to $\mathbf{k}$. This yields:

$$M_{\alpha\beta} = \frac{1}{\hbar^2} \left. \frac{\partial^2 E_C}{\partial k_\alpha \partial k_\beta} \right|_{\mathbf{k_0}} \tag{B.29}$$

which is useful in calculations on electron transport.

Like other second-order tensors which are symmetrical, the inverse effective mass tensor $M_{\alpha\beta}$ can be represented graphically in the form of an ellipsoid [9]. This ellipsoid represents the points in $\mathbf{k}$-space with the same electron energy. The principal values and the principal axes of the ellipsoid can be found by diagonalising the matrix of $M_{\alpha\beta}$. If the coordinate axes are chosen to coincide with the principal axes, the mass tensor is given by:

$$M_{\alpha\beta} = \begin{bmatrix} m_l^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 0 \\ 0 & 0 & m_t^{-1} \end{bmatrix} \tag{B.30}$$

where the x_1-axis is chosen in the direction of $\mathbf{k}_0$, the position of the minimum in $\mathbf{k}$-space. The quantities m_l and m_t are called the longitudinal and the transversal effective mass, respectively. They have been determined quite accurately by cyclotron experiments and can be found in literature; see Table B.2.

Using the above tensor, Equation B.26 reduces to the following conduction band dispersion relation:

$$E_C(\mathbf{k}) = E_{C0} + \frac{\hbar^2}{2} \left[\frac{(k_1 - k_{0.1})^2}{m_l} + \frac{k_2^2 + k_2^2}{m_t} \right] \tag{B.31}$$

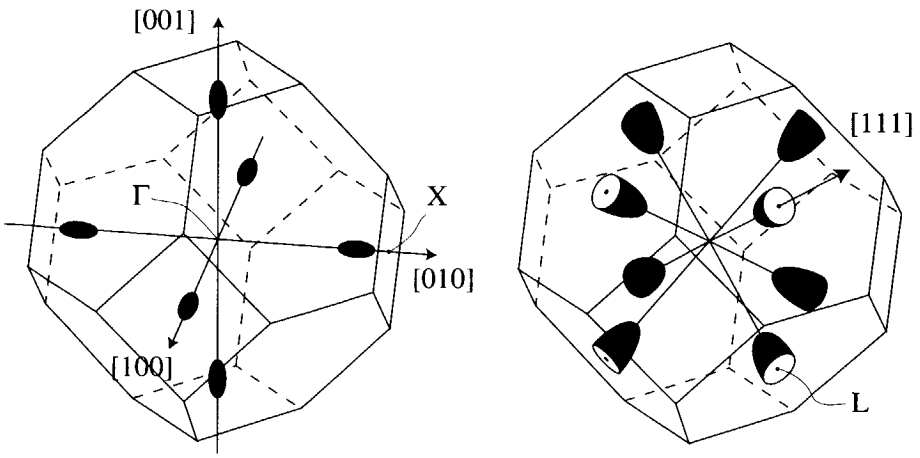

Figure B.5: Schematic of the conduction band equi-energy surfaces drawn in the first Brillouin zone. Left: silicon; right: germanium. Γ, X, and L are particular points in **k**-space.

This equation can be represented by ellipsoids of equal energy as sketched in Figure B.4.

Since the position of the band minima is degenerate in **k**-space, there are m equivalent minima, each centred around its own $\mathbf{k}_0^m$, with a dispersion relation $E_C^m(\mathbf{k})$ and with its own principal axes. Equation B.31 applies to each of those minima. To combine those equations, each should be transformed to the crystal axes system, according to Section A.3.3. In this way, different $E_C^m(\mathbf{k})$ are obtained. Together, they form the well-known 'cigars' of Figure B.5.

The equivalent minima lie on symmetry axes of the crystal. As a result, the principal axes of the ellipsoids of silicon coincide with the main crystal axes of this material. To prove this, the minimum in the [100]-direction at $\mathbf{k}_0^1$ is considered. It is supposed to be described by the effective-mass of Equation B.26, in which the coordinates k_α are defined in the crystal axes system. Now the crystal is subjected to symmetry transformation $\widehat{S}_8\left(1\overline{2}\overline{3}\right)$, defined in Table B.1 and implying a rotation over π around a twofold symmetry axis. The crystal remains invariant under this transformation, whereas the vector $\mathbf{k}_0^1$ remains in place. It means that Equation B.26 is still describing the same energy, although k_2' is replaced by $-k_2'$. This is possible only if the off-diagonal element M_{12}^1 of the inverse effective mass tensor $M_{\alpha\beta}^1$ changes sign under the transformation:

$$\widehat{S}_8\left(1\overline{2}\overline{3}\right) M_{12}^1 = -M_{12}^1 \tag{B.32}$$

when $M_{\alpha\beta}^1$ is defined in the crystal axes system. The tensor was defined in Equation B.27 and depends on the momentum matrix elements p_α^{1b} and p_α^{b1},

Table B.2: Band parameters of silicon and germanium. All values result from experiments, except for the first values of L, M, and N which have been calculated. The band gaps E_G apply to 300 K and 291 K, respectively. See Ref. [10]for a discussion.

Quantity	Unit	Si value	Ref.	Ge value	Ref.
m_l	m_0	0.9163(4)	[11]	1.57(3)	[13]
m_t	m_0	0.1905(1)	[11]	0.0807(8)	[13]
L	$\frac{\hbar^2}{2m_0}$	−5.53	[14]	−30.53	[14]
		−5.641(4)	[12]	−30.34(12)	[15]
M	$\frac{\hbar^2}{2m_0}$	−3.64	[14]	−4.64	[14]
		−3.607(2)	[12]	−4.90(6)	[15]
N	$\frac{\hbar^2}{2m_0}$	−8.32	[14]	−33.64	[14]
		−8.676(6)	[12]	−34.14(12)	[15]
E_G	eV	1.1242	[13]	0.664	[13]
Δ_{SO}	eV	0.0441(3)	[13]	0.289(5)	[16]

$\alpha \in \{1, 2\}$. In general, such an element can be written as:

$$p_\alpha^{1b} = -i\hbar \int \psi_1(\mathbf{x}) \frac{\partial}{\partial x_\alpha} \psi_b(\mathbf{x})\, d\mathbf{x} \tag{B.33}$$

Its integrand is composed of even and odd functions. Since the element does not vanish, the entire integrand is even. It is therefore insensitive to inversions of the coordinate axes resulting from the transformation $\widehat{S}_8\left(1\overline{2}\overline{3}\right)$. This means that:

$$\widehat{S}_8\left(1\overline{2}\overline{3}\right) M_{12}^1 = M_{12}^1 \tag{B.34}$$

which is in correspondence with Equation B.32 if and only if M_{12}^1 is zero. Similar arguments can be used to show that the other diagonal elements vanish. The mass tensor of silicon is therefore diagonal in the crystal axes system, meaning that those axes are also the principal axes of the effective mass tensor.

Symmetry arguments can also be used to prove that M_{22}^1 and M_{33}^1 are equal to each other and different from M_{11}^1 for the minimum at $\mathbf{k}_0^1$. For this purpose, the crystal is rotated over π around the $[01\overline{1}]$-axis (transormation $\widehat{S}_{24}\left(\overline{1}3\overline{2}\right)$) and then inverted. By this transformation, the crystal is turned into itself and the [100]-direction remains on the same place. Equation B.26 is therefore still describing the same energy, although k_2 and k_3 are interchanged. This means that the transverse mass M_{22}^1 is equal to M_{33}^1. None of the symmetry operations of Table B.1, however, can equal it to the longitudinal mass M_{11}^1.

B.6 Valence band equations without spin

The extrema of the valence bands are more difficult to describe than those of the conduction bands because of their degeneracy. This is the result of their location

around $\mathbf{k} = \mathbf{0}$, which is the point in reciprokal space with the full symmetry of the crystal. The degeneracy leads to a system of three energy equations and a 3×3 secular determinant, if the spin-orbit coupling is not included.

This section first presents a suitable form for the valence band states. Then their secular determinant is developed and simplified by means of the crystal symmetry. It will be explicitly solved in Section B.8 after the inclusion of spin-orbit coupling.

The valence band electrons are responsible for the chemical bonding between the atoms of the crystal. These electrons can be found in the outer shell of the electron configuration of the atom, which is only partially filled. The outer shells of independent silicon and germanium atoms have two s- and two p-electrons, which is noted as s^2p^2. In a crystal lattice, the atoms form the well-known tetrahedral bonds. They do this by sharing each of the outer electrons with a neighbouring atom. This, however, requires that one of the s-electrons is promoted to a p-orbital because of the Pauli exclusion principle. As a result, the tetrahedral bonds can be described as linear combinations of one s- and three p-electron states: sp^3.

It appears that the valence band modulating functions in the maximum at $\mathbf{k} = \mathbf{0}$ are linear combinations of the three available p-states [1]. These states $\psi_{n,\mathbf{0}}$ have some very useful symmetry properties. They can be split up into a part depending on the direction in space, and a part depending only on the position *radius* r:

$$\psi_{1,\mathbf{0}}(\mathbf{x}) = x_2 x_3 \, \phi(r) \, ; \quad \psi_{2,\mathbf{0}}(\mathbf{x}) = x_3 x_1 \, \phi(r) \, ; \quad \psi_{3,\mathbf{0}}(\mathbf{x}) = x_1 x_2 \, \phi(r) \qquad \text{(B.35)}$$

With this information it is possible to evaluate the matrix elements and determine the shape of the different valence bands.

Just as the conduction bands, the valence bands E_V^u can be found with the perturbation approach of Section B.3. For this purpose the secular determinant is written again on a base of modulating functions. It thereby makes no difference if the three modulating functions of the valence bands at the maximum are replaced by the set of p-states $\psi_{m,\mathbf{0}}$. To the second order, this yields:

$$\det\left(H_{mn} + \sum_{b \neq m,n}^{B} \frac{H_{mb} H_{bn}}{E'_u - H_{bb}} - E'_u \delta_{mn} \right) = 0 \, ; \qquad m, n \in \{1, 2, 3\} \qquad \text{(B.36)}$$

in which E'_u is the reduced band energy for band u in the sense of Equation B.7.

The matrix elements in this equation can be evaluated by splitting the Hamiltonian $\widehat{H}$ as in Equation B.4 into a part $\widehat{H}_0$ for which the p-states $\psi_{n,\mathbf{0}}$ are the eigenfunctions, and into a perturbation $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$. For the first-order elements H_{mn}

this yields:

$$\begin{aligned} H_{mn} &= \langle m|\, \hat{H}_0 + \hat{H}_{\mathbf{k}\hat{\mathbf{p}}}\, |n\rangle \\ &= \int \psi^*_{m,\mathbf{0}} \hat{H}_0 \psi_{n,\mathbf{0}} d\mathbf{x} + \frac{\hbar}{m_0}\mathbf{k}\cdot \int \psi^*_{m,\mathbf{0}} \hat{\mathbf{p}} \psi_{n,\mathbf{0}} d\mathbf{x} \\ &= E_0 \delta_{mn} \end{aligned} \tag{B.37}$$

The left-hand integral is equal to the eigenenergy E_0 of $\hat{H}_0$. The right-hand integral vanishes when the $\psi_{m,\mathbf{0}}$ of Equation B.35 are inserted because its integrand always odd in one of the coordinates. For the second-order elements in Equation B.36 it applies that:

$$\begin{aligned} H_{mb} &= \int \psi^*_{m,\mathbf{0}} \hat{H}_0 \psi_{b,\mathbf{0}} d\mathbf{x} + \frac{\hbar}{m_0}\mathbf{k}\cdot \int \psi^*_{m,\mathbf{0}} \hat{\mathbf{p}} \psi_{b,\mathbf{0}} d\mathbf{x} \\ &= \frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{p}^{mb}\ ; \qquad m \in \{1,2,3\}\,,\, b \in B \end{aligned} \tag{B.38}$$

where B is the class of states with energies far from the valence bands. The left-hand integral is zero since $m \neq b$ and both states are eigenstates of $\hat{H}_0$. The right-hand integral, however, does not necessarily vanish.

Rewriting the secular determinant with the above information gives:

$$\det\left[\frac{\hbar^2}{m_0^2}\sum_{b\neq m,n}^{B}\frac{\left(\mathbf{k}\cdot\mathbf{p}^{mb}\right)\left(\mathbf{k}\cdot\mathbf{p}^{mb}\right)}{E_0 - E_b} - \left(E'_u - E_0\right)\delta_{mn}\right] = 0 \tag{B.39}$$

where the E'_u in the denominator has been replaced by the unperturbed energy E_0. This is possible if the energy shift due to the $\mathbf{k}\hat{\mathbf{p}}$ 'perturbation' is much smaller than the interband distances $E_0 - E_b$. The inner products within the equation can be rewritten as sums in k_1, k_2, and k_3. For this purpose, also the k_α within the reduced energy E'_u are taken into account (see Equation B.7):

$$\det\left[D^{mn}_{\alpha\beta}k_\alpha k_\beta - \left(E_u - E_0\right)\delta_{mn}\right] = 0\ ,\quad \alpha,\beta \in \{1,2,3\} \tag{B.40}$$

in which the Einstein convention is used to indicate the summation of the terms with respect to the free subscripts α and β. The tensor $D^{mn}_{\alpha\beta}$ in this equation is of the fourth rank, and is defined in analogy to the inverse effective mass tensor $M_{\alpha\beta}$ of Equation B.27:

$$D^{mn}_{\alpha\beta} = \frac{\hbar^2}{2m_0}\delta_{mn}\delta_{\alpha\beta} + \frac{\hbar^2}{m_0^2}\sum_{b\neq m,n}^{B}\frac{p^{mb}_\alpha p^{bn}_\beta}{E_A - E_b} \tag{B.41}$$

where p^{mb}_α is the subscript notation of the vector of matrix elements $\mathbf{p}^{mb}$. The second-rank tensor $D^{mn}_{\alpha\beta}k_\alpha k_\beta$ is the result of the $\mathbf{k}\hat{\mathbf{p}}$ Hamiltonian $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$ only and is called the $\mathbf{k}\hat{\mathbf{p}}$ matrix $\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}$. The energy difference $E_u - E_0$ is the hole energy

with respect to the band edge and is abbreviated to E. Therefore, Equation B.40 can be written in short as the eigenvalue problem:

$$\det\left(\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}} - E\delta_{mn}\right) = 0 \tag{B.42}$$

The $\mathbf{k}\hat{\mathbf{p}}$ matrix can be considerably simplified by applying the symmetry of the crystal to the tensor $D^{mn}_{\alpha\beta}$. Many of its components are either zero or mutually dependend. To show this, the different band energies E_b in the denominator of Equation B.41 can be replaced by a single level E_B. This does not alter the shape of the final expression, but simplifies the calculations [1]. Due to the completeness of the $\psi_{b,\mathbf{0}}$ the tensor then reduces to:

$$D^{mn}_{\alpha\beta} = \frac{\hbar^2}{2m_0}\delta_{mn}\delta_{\alpha\beta} + \frac{\hbar^2}{m_0^2}\frac{\langle m|\,\hat{p}_\alpha\hat{p}_\beta\,|n\rangle}{E_0 - E_B} \tag{B.43}$$

Now the different tensor elements can be evaluated with the aid of the wave functions of Equation B.35 and the fact that the momentum operator can be written as $\hat{p}_\alpha = -i\hbar\partial_\alpha$, where ∂_α is defined for brevity as the nabla operator $\partial/\partial x_\alpha$.

Firstly, the element D^{11}_{13} is considered. It contains the following integral:

$$\langle 1|\,p_1p_3\,|1\rangle = -\hbar^2\int\left(\phi\left(r\right)x_2x_3\right)^*\partial_1\partial_3\left(\phi\left(r\right)x_2x_3\right)d\mathbf{x} \tag{B.44}$$

Since the crystal has cubic symmetry, the energy contribution of the element should be invariant under a rotation of π radians around the x-axis (transformation $\hat{S}_8\left(1\overline{2}\overline{3}\right)$ of Table B.1). Applying this to Equation B.44, however, yields $\hat{S}_8\langle 1|\,\hat{p}_1\hat{p}_3\,|1\rangle = -\langle 1|\,\hat{p}_1\hat{p}_3\,|1\rangle$. This is only possible if the element is zero. A similar argument can be given for the element D^{12}_{11} that also vanishes by the transformation $\hat{S}_8\left(1\overline{2}\overline{3}\right)$, and for D^{13}_{12} which is forced to zero by $\hat{S}_9\left(\overline{1}2\overline{3}\right)$.

Secondly, D^{11}_{11} is considered. This diagonal element is nonzero because the integrand of

$$\langle 1|\,p_1p_1\,|1\rangle = -\hbar^2\int\left(\phi\left(r\right)x_2x_3\right)^*\partial_1\partial_1\left(\phi\left(r\right)x_2x_3\right)d\mathbf{x} \tag{B.45}$$

is invariant under any transformation of the cubic group of Table B.1. Therefore it is defined as the constant L, which is usually determined by experiment:

$$D^{11}_{11} = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum_{b\neq 1}^{B}\frac{p_1^{1b}p_1^{b1}}{E_0 - E_b} \equiv L \tag{B.46}$$

Thirdly, the element D^{11}_{22} is considered, which is determined by the integral:

$$\langle 1|\,p_2p_2\,|1\rangle = -\hbar^2\int\left(\phi\left(r\right)x_2x_3\right)^*\partial_2\partial_2\left(\phi\left(r\right)x_2x_3\right)d\mathbf{x} \tag{B.47}$$

This element is equally nonzero. However, it differs structurally from $\langle 1|\,\hat{p}_1\hat{p}_1\,|1\rangle$ because it contains the variable x_2 in both in the operator and the eigenfunctions. Therefore it is defined as the constant M:

$$D_{22}^{11} = \frac{\hbar^2}{2m_0} + \frac{\hbar^2}{m_0^2}\sum_{b\neq 1}^{B}\frac{p_2^{1b}p_2^{b1}}{E_0 - E_b} \equiv M \tag{B.48}$$

Finally, the off-diagonal element D_{12}^{12} is investigated and found to be nonzero. It is combined with its counterpart D_{21}^{12} into the symmetric constant N:

$$D_{12}^{12} + D_{21}^{12} = \frac{\hbar^2}{m_0^2}\sum_{b\neq 1}^{B}\frac{p_1^{1b}p_2^{b2} + p_2^{1b}p_1^{b2}}{E_0 - E_b} \equiv N \tag{B.49}$$

The remaining tensor elements can be obtained by using the fact that the x_1-, x_2, and x_3-axes are equivalent in the case of cubic symmetry. It therefore applies that:

$$\begin{aligned} L &= D_{11}^{11} = D_{22}^{22} = D_{33}^{33} \\ M &= D_{22}^{11} = D_{33}^{11} = D_{33}^{22} = D_{11}^{22} = D_{11}^{33} = D_{22}^{33} \qquad \text{(B.50)} \\ N &= D_{12}^{12} + D_{21}^{12} = D_{23}^{23} + D_{32}^{23} = D_{13}^{13} + D_{31}^{13} \\ &= D_{12}^{21} + D_{21}^{21} = D_{23}^{32} + D_{32}^{32} = D_{13}^{31} + D_{31}^{31} \qquad \text{(B.51)} \end{aligned}$$

The total $\mathbf{k}\hat{\mathbf{p}}$ matrix $D_{\alpha\beta}^{mn}k_\alpha k_\beta$ can therefore be written as:

$$\begin{aligned} \mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}} &= D_{\alpha\beta}^{mn}k_\alpha k_\beta \qquad \text{(B.52)} \\ &= \begin{bmatrix} Lk_1^2 + M\left(k_2^2 + k_3^2\right) & Nk_1k_2 & Nk_1k3 \\ Nk_1k_2 & Lk_2^2 + M\left(k_1^2 + k_3^2\right) & Nk_2k_3 \\ Nk_1k_3 & Nk_2k_3 & Lk_3^2 + M\left(k_1^2 + k_2^2\right) \end{bmatrix} \end{aligned}$$

The valence band parameters L, M, and N appear in literature with slightly different definitions. They also appear as linear combinations of other parameters, usually A, B, and C, or the Luttinger parameters γ_1, γ_2, and γ_3. This should be kept in mind when numerical values are needed. An overview of the definitions is given in Table B.3.

The above $\mathbf{k}\hat{\mathbf{p}}$ matrix can be inserted into the secular determinant of Equation B.42 to find the three eigenvalues E, and thereby the valence band equations E_V^u. These equations are a function of $\mathbf{k}$ and are therefore dispersion relations. At $\mathbf{k} = \mathbf{0}$ they are indeed threefold degenerate. It appears, however, that this is not completely true in reality because the effects of spin-orbit coupling should be included.

B.7 Effects of spin-orbit coupling

Up to present, it has been neglected that each electron has either a positive or negative spin. Spin adds a magnetic moment to the electron. This moment

Table B.3: Different definitions of the valence band parameters L, M, and N appearing in literature.

Ref.	L	M	N
[1]	$\frac{\hbar^2}{2m_0}(A+2B-1)$	$\frac{\hbar^2}{2m_0}(A-B-1)$	$\frac{\hbar^2}{2m_0}\sqrt{9B^2+3C^2}$
[3]	$A+2B$	$A-B$	$\sqrt{3}D$
[4], [17]	$A+2B-\frac{\hbar^2}{2m_0}$	$A-B-\frac{\hbar^2}{2m_0}$	$\sqrt{9B^2+3C^2}$
[18]	$\hbar^2 A$	$\hbar^2 B$	$\hbar^2 C$
[18]	$-\frac{\hbar^2}{2m_0}(\gamma_1+4\gamma_2)$	$-\frac{\hbar^2}{2m_0}(\gamma_1-2\gamma_2)$	$-\frac{\hbar^2}{2m_0}6\gamma_3$

interacts with the magnetic field created by the orbital angular momentum of the electron, and increases its potential energy. The interaction can be represented by adding a spin-orbit term $\hat{H}_{SO}$ to the spin-free Hamiltonian $\hat{H}_{\mathbf{k}}$: [5, 7]:

$$\hat{H}\psi = \left(\hat{H}_{\mathbf{k}} + \hat{H}_{SO}\right)\psi = E_n\psi \tag{B.53}$$

$$\hat{H}_{SO} = \frac{\hbar}{4m_0^2c_l^2}(\nabla U \times \hat{\mathbf{p}})\cdot\hat{\sigma} = \frac{\Delta_{SO}}{3\hbar}\hat{\mathbf{L}}\cdot\hat{\sigma} \tag{B.54}$$

Here, c_l is the velocity of light, $\hat{\mathbf{L}}$ is the orbital angular momentum operator, and $\hat{\sigma}$ is the vector of Pauli matrices, consisting of [19]:

$$\hat{\sigma}_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},\ \hat{\sigma}_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix},\ \hat{\sigma}_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{B.55}$$

The constant Δ_{SO} is called the spin-orbit splitting energy. It depends on the material and increases with the weight of the constituting atoms. Silicon and germanium are relatively light, so the spin-orbit coupling in those materials is sufficiently small to be taken into account as a perturbation.

The spin-orbit somewhat shifts the band edges and thereby has a more perceptible effect on the valence bands than on the conduction bands. The reason is that it partially lifts the degeneracy of the valence band edges, which has a strong effect the band shapes. The conduction band edges are nondegenerate, however, and therefore retain their shape. In the following it will be described how the spin-orbit coupling influences the valence band states and their secular determinant.

The inclusion of spin doubles the number of valence band wave functions to six. It also doubles the number of p-states of which the bands are constructed. In Equation B.35 these states $\psi_{i,\mathbf{0}}(\mathbf{x})$ were described as $x_j x_k \phi(r)$. They can be

replaced by a new orthogonal set of states including the spinor functions χ^{α}:

$$\Psi_i^+(\mathbf{x}) = \psi_{i,\mathbf{0}}(\mathbf{x})\,\chi^+ = x_j x_k \phi(r) \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad i \neq j \neq k \text{ and} \tag{B.56}$$

$$\Psi_i^-(\mathbf{x}) = \psi_{i,\mathbf{0}}(\mathbf{x})\,\chi^- = x_j x_k \phi(r) \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad i,j,k \in \{1,2,3\} \tag{B.57}$$

where the plus and minus indicate that the spin points up- or downwards, respectively.

The appearance of the spin-orbit Hamiltonian $\hat{H}_{SO}$ is slightly different when it is placed on a basis of modulating functions $u^{\pm}_{i,\mathbf{k}}$. This is the result of the differentiating momentum operator within Equation B.54. Applied to the definition of the Bloch function, this yields:

$$\hat{H}_{SO} u^{\pm}_{i,\mathbf{k}} = \frac{\hbar}{4m_0^2 c_l^2} \left[\nabla U \times (\hat{\mathbf{p}} + \hbar\mathbf{k})\right] \cdot \hat{\sigma} u^{\pm}_{i,\mathbf{k}} \tag{B.58}$$

where the term $\hbar\mathbf{k}$ is new, compared to the operator for the full wave function. The new term is generally neglected, however, because the crystal momentum $|\hbar\mathbf{k}|$ is much smaller than the average electron momentum $|\hat{\mathbf{p}}|$. It can also be retained, which adds a small term to $\hat{\mathbf{p}}$, but leaves the shape of the band equations unchanged. In both cases the spin-orbit Hamiltonian is equal for both Bloch and modulating functions.

The doubling of the number of wave functions also doubles the rank of the secular determinant in Equation B.36:

$$\det\left(H_{mn} + \sum_{b\neq m,n}^{B} \frac{H_{mb}H_{bm}}{E'_u - H_{bb}} - E'_u \delta_{mn} \right) = 0\,; \quad m,n \in \{1,2,..6\} \tag{B.59}$$

An arbitrary matrix element in this determinant now obtains the following shape:

$$H_{mn} = H_{ij}^{\alpha\beta} = \left\langle \Psi_i^{\alpha} \middle| \hat{H}_0 + \hat{H}_{\mathbf{k}\hat{\mathbf{p}}} + \hat{H}_{SO} \middle| \Psi_j^{\beta} \right\rangle\,; \quad \begin{matrix} i,j \in \{1,2,3\} \\ \alpha,\beta \in \{+,-\} \end{matrix} \tag{B.60}$$

The first two terms of this element can be reduced to the spinless case in Equation B.37, becaus the inclusion of spin does not change the eigenvalues of $\hat{H}_0$and $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$. These terms can therefore be written as:

$$\left\langle \Psi_i^{\alpha} \middle| \hat{H}_0 + \hat{H}_{\mathbf{k}\hat{\mathbf{p}}} \middle| \Psi_j^{\beta} \right\rangle = \left\langle \psi_{i,\mathbf{0}} \middle| \hat{H}_0 + \hat{H}_{\mathbf{k}\hat{\mathbf{p}}} \middle| \psi_{j,\mathbf{0}} \right\rangle \cdot (\mathbf{I}_2)_{\alpha\beta} \tag{B.61}$$

where $\psi_{i,\mathbf{0}}$ are the original p-states and $\mathbf{I}_2$ is the 2×2 identity matrix [20].

The first-order elements of the secular determinant are especially determined by the spin-orbit Hamiltonian $\hat{H}_{SO}$. After all, it was seen in Equation B.37 that

the first-order elements of $\hat{H}_0$ equal the unperturbed energy E_0, whereas those of $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$ are zero. The elements of $\hat{H}_{SO}$ can be evaluated by writing them as:

$$\left\langle \Psi_i^\alpha \left| \hat{H}_{SO} \right| \Psi_j^\beta \right\rangle = \lambda \sum_{opq=1}^{3} \left\langle \psi_{i,\mathbf{0}} \chi^\alpha \left| \widehat{\sigma}_o \epsilon_{opq} \partial_p U \partial_q \right| \psi_{j,\mathbf{0}} \chi^\beta \right\rangle , \quad \alpha, \beta \in \{1,2\}$$
$$= \lambda \sum_{opq} \left\langle \psi_{i,\mathbf{0}} \left| \epsilon_{opq} \partial_p U \partial_q \right| \psi_{j,\mathbf{0}} \right\rangle \left(\widehat{\sigma}_o\right)_{\alpha\beta} \tag{B.62}$$

where λ is equal to $-i\hbar^2/4m_0^2c_l^2$, ∂_j is defined as the gradient $\partial/\partial x_j$. The ϵ_{opq} is the complete antisymmetric tensor. It represents the cross product of Equation B.54 and gives rise to a sum of six terms. The spin superscrips $\alpha\beta$ now take the values '1' or '2' instead of '+' or '-'. This shows that the effect of the two spinors χ^α and χ^β is to select an element $[\alpha, \beta]$ of the Pauli matrix $\widehat{\sigma}_o$.

The next step in the evaluation is to expand the spinless part of the matrix elements into integrals and examine their symmetry. Within the integrands, the gradient operator ∂_q works on the spherically symmetric part of the wavefunction $\phi(r)$. This yields:

$$\partial_q \phi(r) = \frac{\partial \phi(r(\mathbf{x}))}{\partial x_q} = \frac{x_q}{r} \frac{\partial \phi}{\partial r} = \frac{x_q}{r} \partial_r \phi \tag{B.63}$$

When this is inserted, it appears that the diagonal elements $[m, m]$ are odd in two coordinates and therefore vanish. The off-diagonal elements, however, contain even terms and are nonzero. This happens when $i = o$ and $j = p$. If (i, j, k) is an even permutation of $(1, 2, 3)$, the elements are equal to:

$$\sum_{ijk} \left\langle \psi_{i,\mathbf{0}} \left| \epsilon_{ijk} \partial_j U \partial_k \right| \psi_{j,\mathbf{0}} \right\rangle \left(\widehat{\sigma}_i\right)_{\alpha\beta}$$
$$= \left(\widehat{\sigma}_k\right)_{\alpha\beta} \int d\mathbf{x} \phi^*(r) x_j x_k \left[\partial_i U \frac{x_i x_j x_k}{r} \partial_r - \partial_j U x_k \left(1 + \frac{x_i^2}{r} \partial_r \right) \right] \phi(r) \tag{B.64}$$
$$= \left\langle \psi_{i,\mathbf{0}} \left| \partial_i U \partial_j - \partial_j U \partial_i \right| \psi_{j,\mathbf{0}} \right\rangle \left(\widehat{\sigma}_k\right)_{\alpha\beta} \tag{B.65}$$

For an odd permutation of (i, j, k) the sign is reversed. In the latter equation it is possible to recognise a commutation of the potential gradient $\partial_i U = \nabla U$ and the momentum operator $\widehat{p}_j = \widehat{\mathbf{p}}$. They correspond to the total time derivatives of momentum and position, respectively. The first-order spin-orbit matrix elements of Equation B.62 can therefore be written as:

$$\left\langle \Psi_i^\alpha \left| \hat{H}_{SO} \right| \Psi_j^\beta \right\rangle = \mp i \frac{\Delta_{SO}}{3} \left(\widehat{\sigma}_k\right)_{\alpha\beta} , \qquad i \neq j \neq k \tag{B.66}$$

in which Δ_{SO} is the spin-orbit splitting energy, defined as:

$$\Delta_{SO} = i \frac{3\hbar}{4m_0^2 c_l^2} \left\langle \psi_{i,\mathbf{0}} \left| [\nabla U, \widehat{\mathbf{p}}]_- \right| \psi_{j,\mathbf{0}} \right\rangle , \quad i \neq j \tag{B.67}$$

The second-order elements of the secular determinant need not to be changed with respect to the spinless case. They only need to contain the influence

of $\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$, because $\hat{H}_{SO}$ has already been taken into account in the first order. They yield the new $\mathbf{k}\hat{\mathbf{p}}$ matrix $\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}^{6\times6}$, resulting from scaling $\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}$ in the way of Equation B.61.

The secular determinant can now be written in matrix notation, in analogy to Equation B.42:

$$\det\left[\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}^{6\times6} + \mathbf{H}_{SO} - E\delta_{mn}\right] = 0 \tag{B.68}$$

The spin-orbit matrix $\mathbf{H}_{SO}$ is generated by Equation B.66 and reads:

$$\mathbf{H}_{SO} = \frac{\Delta_{SO}}{3}\begin{bmatrix} \mathbf{0} & -i\widehat{\sigma}_3 & i\widehat{\sigma}_2 \\ i\widehat{\sigma}_3 & \mathbf{0} & -i\widehat{\sigma}_1 \\ -i\widehat{\sigma}_2 & i\widehat{\sigma}_1 & \mathbf{0} \end{bmatrix} \tag{B.69}$$

The energy Δ_{SO} is equal to the splitting of the band edges. This can be derived from inserting $\mathbf{k} = \mathbf{0}$ into Equation B.68 and calculating the energy eigenvalues. Two band edges are shifted upwards with $\Delta_{SO}/3$ and are still degenerate, while one is lowered with an amount $2\Delta_{SO}/3$. The splitting energy has been determined experimentally. The resulting values for silicon and germanium are shown in Table B.2.

B.8 Valence band equations including spin

The valence band dispersion relations can now be calculated analytically by solving the eigenvalue determinant of Equation B.68, i.e. finding the energy eigenvalues of the matrix $\mathbf{H} = \mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}^{6\times6} + \mathbf{H}_{SO}$ [20]:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} - i\widehat{\sigma}'_3 & h_{13} + i\widehat{\sigma}'_2 \\ h_{12} + i\widehat{\sigma}'_3 & h_{22} & h_{23} - i\widehat{\sigma}'_1 \\ h_{13} - i\widehat{\sigma}'_2 & h_{23} + i\widehat{\sigma}'_1 & h_{33} \end{bmatrix} \tag{B.70}$$

Each element in this matrix is a 2×2 matrix itself. The $\widehat{\sigma}'_i$ stand for $\widehat{\sigma}_i\Delta_{SO}/3$, and the diagonal elements h_{ii} for:

$$h_{ii} = \left[Lk_i^2 + M\left(k_j^2 + k_k^2\right)\right]\mathbf{I}_2 \tag{B.71}$$

in which (i, j, k) is a cyclic permutation of $(1, 2, 3)$. The off-diagonal elements are:

$$h_{ij} = Nk_ik_j\mathbf{I}_2\ , \qquad i \neq j \tag{B.72}$$

The secular equation is found by Gaussian elimination, during which care must be taken to multiply the operators from the correct side. It may be used that $\widehat{\sigma}_i\widehat{\sigma}_i = \mathbf{I}_2$, and that $\widehat{\sigma}_i\widehat{\sigma}_j = \pm i\widehat{\sigma}_k$ for even and odd permutations of (i, j, k), respectively. This results in two decoupled cubic equations from which the Pauli matrices have disappeared:

$$E^3 + pE^2 + qE + r = 0 \tag{B.73}$$

Its coefficients p, q, and r are real, and are given by:

$$p = -(h_{11} + h_{22} + h_{33}) \tag{B.74}$$

$$q = h_{11}h_{22} + h_{22}h_{33} + h_{33}h_{11} - (h_{12} + h_{23} + h_{31}) - 3\left(\frac{\Delta_{SO}}{3}\right)^2 \tag{B.75}$$

$$r = -h_{11}h_{22}h_{33} - 2h_{12}h_{23}h_{31} + h_{11}h_{23} + h_{22}h_{31} + h_{33}h_{12} + + (h_{11} + h_{22} + h_{33})\left(\frac{\Delta_{SO}}{3}\right)^2 + 2\left(\frac{\Delta_{SO}}{3}\right)^3 \tag{B.76}$$

This cubic secular equation should have three real solutions, one for each band. These yield the valence band dispersion relations [21, 22]:

$$E_V^{\mathrm{HH}}(\mathbf{k}) = E_0 + 2\sqrt{Q}\cos\left(\frac{\Theta}{3}\right) - \frac{p}{3} \tag{B.77}$$

$$E_V^{\mathrm{LH}}(\mathbf{k}) = E_0 + 2\sqrt{Q}\cos\left(\frac{\Theta - 2\pi}{3}\right) - \frac{p}{3} \tag{B.78}$$

$$E_V^{\mathrm{SO}}(\mathbf{k}) = E_0 + 2\sqrt{Q}\cos\left(\frac{\Theta + 2\pi}{3}\right) - \frac{p}{3} \tag{B.79}$$

where HH stands for 'heavy-hole band', LH for 'light-hole band', and SO for 'split-off band'. The angle Θ is defined as:

$$\Theta = \arccos\left(\frac{-R}{\sqrt{Q^3}}\right)$$

in which the quantities Q and R are defined as:

$$Q = \frac{1}{9}\left(p^2 - 3q\right) \tag{B.80}$$

$$R = \frac{1}{54}\left(2p^3 - 9pq + 27r\right) \tag{B.81}$$

The degenerate edge of the heavy and light hole bands is often defined as the zero-energy level. Since it is positioned at $\mathbf{k} = \mathbf{0}$, it can be described as:

$$E_V^{\mathrm{HH}}(\mathbf{0}) = E_V^{\mathrm{LH}}(\mathbf{0}) = E_{V0} = E_0 + \frac{\Delta_{SO}}{3} \equiv 0 \tag{B.82}$$

The bands are plotted in Fig. B.6 for values of $\mathbf{k}$ around $\mathbf{0}$. It can be seen that the bands are quite anisotropic and not quadratic, especially the HH-band. The labels HH and LH correspond to the steepness of the respective bands, which is inversely proportional to the effective masses. SO refers to the spin-orbit splitting by which the top of this band lies Δ_{SO} (44 meV) below the other two bands.

The bands can also be represented by surfaces of equal energy in $\mathbf{k}$-space. For different energies they are plotted in Figure B.7, B.8 and B.9. Two energy

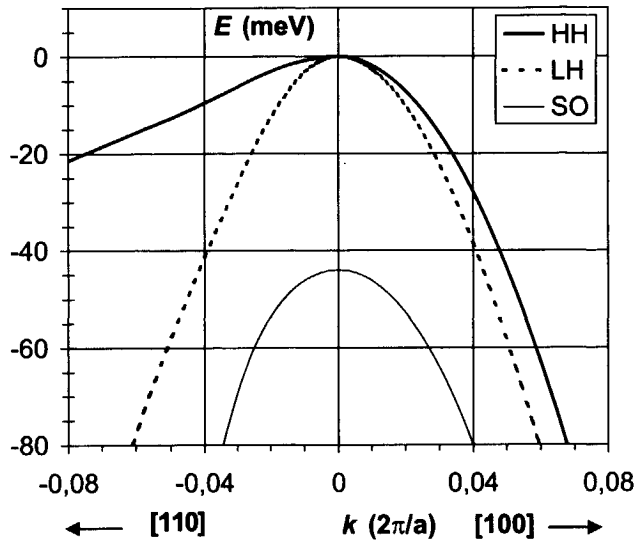

Figure B.6: Valence band dispersion relations of Si in the [100]- and [110]-directions, as obtained with the $\mathbf{k}\hat{\mathbf{p}}$-method. E is the energy, k the wave number, and a the atomic constant.

ranges are of special interest. Near the HH- and LH-band edges the influence of the SO-band is relatively weak and the shapes of the surfaces only arises from the interaction between the HH- and the LH-bands. Closer to Δ_{SO} the SO-band also interacts, which has a strong influence on the band shapes. At a distance of some Δ_{SO} the bands start to decouple, which means that their equi-energy surfaces approach the spheres of the free-electron case. Similar figures have been published by Hinckley [7, 23] and Dijkstra [24].

B.9 Conclusions

The energy bands of silicon and germanium can be approached near the bandgap with the $\mathbf{k}\hat{\mathbf{p}}$ method. This method yields dispersion relations which are expressed in measurable quantities such as the effective masses. For this purpose it employs the symmetry properties of the semiconductor crystal. The dispersion relation of the conduction band shows minima at different places in $\mathbf{k}$-space which are nondegenerate and can be described by a parabolic equation. Graphically, they can be represented by ellipsoids of equal energy. The dispersion relations of the valence bands are triply degenerate at the origin and therefore strongly influence each other. They are partially decoupled by the spin-orbit effect. Mathematically, they can be described in a closed form by the solution of a cubic secular equation. Graphically, they are represented by warped spherical surfaces, of which the degree of warping depends on the energy considered.

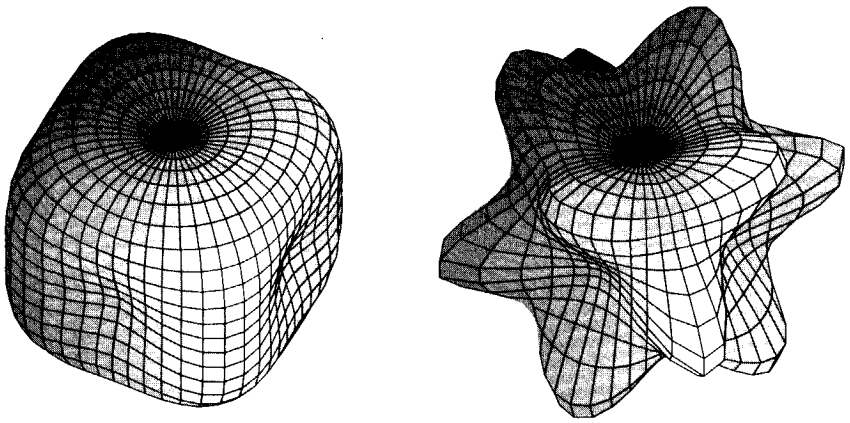

Figure B.7: Equi-energy surfaces of the heavy-hole band at 1 meV (left) and 44 meV (right) under the band edge. Not on scale.

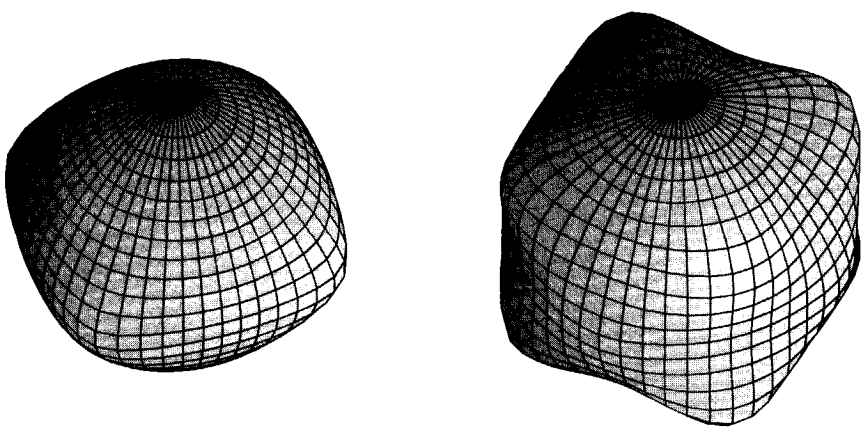

Figure B.8: Equi-energy surfaces of the light-hole band at 1 meV (left) and 44 meV (right) under the band edge. Not on scale.

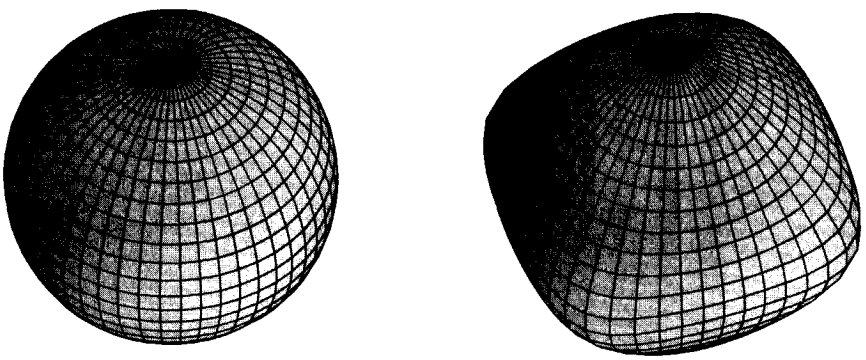

Figure B.9: Equi-energy surfaces of the split–off hole band at 45 meV (left) and 88 meV (right) under the HH band edge. Not on scale.

Bibliography

[1] I. M. Tsidilkovski, *Band structure of semiconductors*, Int. Ser. Scie. Sol. St., 19, Pergamon Press, Oxford, 1982.

[2] N. M. Ashcroft and N. D. Mermin, *Solid state physics*, Saunders College, Philadelphia, 1976.

[3] G. L. Bir and G. E. Pikus, *Symmetry and strain-induced effects in semiconductors*, Wiley, New York, 1974.

[4] G. Dresselhaus, A. F. Kip, and C. Kittel, Cyclotron resonance of electrons and holes in silicon and germanium crystals, *Phys. Rev., 98* (1955) 368-384.

[5] J. M. Luttinger and W. Kohn, Motion of electrons and holes in perturbed periodic fields, *Phys. Rev., 97* (1955) 869-883.

[6] I. Ipatova and V. Mitin, *Solid-state electronics*, Addison-Wesley, Reading, Massachusetts, 1996.

[7] J. Singh, *Physics of semiconductors and their heterostructures*, McGraw-Hill, New York, 1993.

[8] P.-O. Löwdin, A note on the quantum-mechanical perturbation theory, *J. Chem. Phys., 19* (1951) 1396-1401.

[9] J. F. Nye, *Physical properties of crystals*, Clarendon Press, Oxford, 1985.

[10] M. A. Green, Intrinsic concentration, effective densities of states, and effective mass in silicon, *J. Appl. Phys., 67* (1990) 2944-2954.

[11] J. C. Hensel, H. Hasegawa, and M. Nakayama, Cyclotron resoance in uniaxially stressed silicon, II. Nature of the covalent bond, *Phys. Rev., 138* (1965) 225-238.

[12] J. C. Hensel in: R. G. Humphreys, Valence band averages in silicon: anisotropy and nonparabolicity, *J. Phys. C: Solid State Phys., 14* (1981) 2935-2942.

[13] O. Madelung (ed.), *Semiconductors - Basic data*, Springer, Berlin, 1996.

[14] M. V. Fischetti, and S. E. Laux, Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys, *J. Appl. Phys., 80* (1996) 2234-52.

[15] J. C. Hensel and K. Suzuki, Quantum resonance in the valence bands of germanium. II. Cyclotron resonances in uniaxially stressed crystals, *Phys. Rev. B, 9* (1974) 4219-4257.

[16] A. R. Goñi, K. Syassen, and M. Cardona, Direct-band absorption in germanium under pressure, *Phys. Rev. B, 39* (1989) 12,921-12,924.

[17] J. C. Hensel and G. Feher, Cyclotron resonance experiments in uniaxially stressed silicon: valence band inverse mass parameters and deformation potentials, *Phys. Rev., 129* (1963) 1041-1062.

[18] J. M. Luttinger, Quantum theory of cyclotron resonance in semiconductors: general theory, *Phys. Rev., 102* (1956) 1030-1041.

[19] W. Greiner, *Quantum mechanics, an introduction,* Springer, Berlin, 1989.

[20] J. E. Dijkstra and W. Th. Wenckenbach, Hole transport in strained Si, *J. Appl. Phys., 81* (1997) 1259-1263.

[21] M. Abramowitz and I. M. Stegun,*Handbook of mathematical functions; with formulas, graphs, and mathematical tables,* Dover, New York, 1965, p. 17.

[22] R. C. Weast and S. M. Selby (eds.), *Handbook of tables for mathematics,* 4th ed., Chemical Rubber Co., Cleveland, Ohio, 1970.

[23] J. M. Hinckley, *The effect of strain in pseudomorphic p-$Si_{1-x}Ge_x$: Physics and modeling of the valence bandstructure ad hole transport,* Ph. D. thesis, The University of Michigan, Ann Arbor, 1990.

[24] J. E. Dijkstra, *Monte Carlo simulation of hole transport in Si, Ge and $Si_{1-x}Ge_x$,* Ph. D. thesis, Delft University of Technology, Delft, The Netherlands, 1997.

Appendix C

Charge Transport in Stressed Silicon

The electron and hole currents in a semiconductor can be derived from its quantummechanical behaviour, as was argued in Section 2.2.1. For this purpose, a description of the energy bands is needed together with a statistical approach and the Boltzmann equation. Such a derivation is described in most textbooks on solid state and semiconductor physics [1, 2, 3]. In most of them, however, it is made with some assumptions that are unwelcome in the study of semiconductors under mechanical stress. These assumptions include:

- Isotropic material,
- Homogeneous material, and
- Parabolic edges of the energy bands.

In this appendix the transport of charge carriers is calculated without making these assumptions. Many other assumptions, however, still hold under stress. They are mentioned in Section 2.2.1 and are used in the following. The derivations lead to general expressions for the charge carrier concentrations and the conductivities in terms of band structure parameters. These are used to find the mobility tensors, and a set of generalised drift-diffusion equations. Finally, the derivations lead to generalised expressions for some auxiliary relations such as the Poisson equation.

The following derivations are based on the band descriptions of Appendix B, as well as on the generalised band edge diagram of Section 2.1.7. In addition, they consequently apply the Boltzmann distribution function and the density of states concept to each band separately. Much in this appendix is based on the excellent work of Marshak and Van Vliet [4].

C.1 Charge concentration in nonequilibrium

The number of electrons in an energy band depends on how much of them have the energy required, and how much states are available. Since those numbers are considerable, the energy of the electrons can be described by a distribution function and the available states by a density of states function. The shape of the distribution function is estimated by making the relaxation time approximation. The density of states function is deduced from the band dispersion relations. Often, its average over all energies is calculated and called the effective density of states, which can also be expressed in a density of states effective mass.

C.1.1 The relaxation time approximation

In the semiclassical model the behaviour of an electron is specified by its position $\mathbf{x}$ and wave vector $\mathbf{k}$. The distribution of its momentum is given by a function $f_n(\mathbf{x}, \mathbf{k}, t)$ which can also be considered as the probability of finding a particle of a momentum $\hbar\mathbf{k}$ at a specific position $\mathbf{x}$ and time t.

The shape of the distribution can be estimated by making the *relaxation time approximation* ([1], Ch.13). In this approximation, electrons in a conduction band m collide with the atomic lattice with a chance $dt/\tau_m(\mathbf{x}, \mathbf{k})$ per infinitesimal time interval dt, where $\tau_m(\mathbf{x}, \mathbf{k})$ is the mean time between two collisions, called relaxation time. It is assumed that:

1. The distribution after the collision does not depend on the distribution before.
2. The collisions do not influence the shape of the local distribution in thermodynamic equilibrium.

The first of those assumptions is illustrated in Figure C.1.

The distribution in this equilibrium is given by Fermi-Dirac statistics because the wave functions of electrons are antisymmetric [5]. The distribution function f_n can therefore be written as:

$$f_n(\mathbf{k}, \mathbf{x}) = f_n(E, \mathbf{x}) = \left[1 + \exp\left(\frac{E(\mathbf{k}, \mathbf{x}) - E_{Fn}(\mathbf{x})}{k_B T(\mathbf{x})}\right)\right]^{-1} \tag{C.1}$$

in which k_B is the Boltzmann constant, $T(\mathbf{x})$ the local temperature, and $E_{Fn}(\mathbf{x})$ the electron quasi Fermi level or electrochemical potential. It should be noted that f_n depends on $\mathbf{k}$ only through the electron energy E. For quasi Fermi levels lying more than $2.3\, k_B T$ under E the distribution can be approached by Maxwell-Boltzmann statistics with less than 10% error:

$$f_n(\mathbf{k}, \mathbf{x}) \simeq \exp\left(\frac{E_{Fn}(\mathbf{x}) - E(\mathbf{k}, \mathbf{x})}{k_B T(\mathbf{x})}\right) \tag{C.2}$$

The materials for which this is true are called nondegenerate. For silicon it applies when the doping concentration is approximately less than 10^{18} atoms

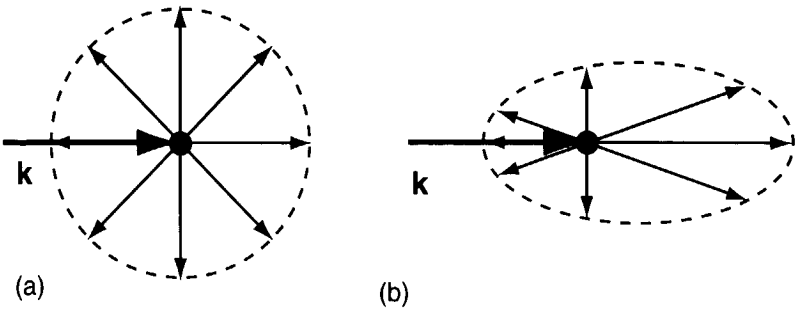

Figure C.1: Illustration of the relaxation time approximation and a counterexample. An electron with wave vector $\mathbf{k}$ scatters and may be sent in any direction. (a) All directions are equally likely. (b) There is a most likely direction which depends on the direction of incidence.

per cm^3. Maxwell-Boltzmann statistics considerably simplify the analytical calculations in the following.

For holes the distribution function f_p is the chance of not finding an electron, which means that $f_p = 1 - f_n$. By using the hole quasi Fermi level $E_{Fp}(\mathbf{x})$ it is written as:

$$f_p(\mathbf{k},\mathbf{x}) = \left[1 + \exp\left(\frac{E_{Fp}(\mathbf{x}) - E(\mathbf{k},\mathbf{x})}{kT(\mathbf{x})}\right)\right]^{-1}$$
$$\simeq \exp\left(\frac{E(\mathbf{k},\mathbf{x}) - E_{Fp}(\mathbf{x})}{k_B T(\mathbf{x})}\right) \tag{C.3}$$

where the latter is, of course, only true for nondegenerate material.

In thermodynamic equilibrium the quasi Fermi levels E_{Fn} and E_{Fp} need to be equal to a value E_F. In that case the distribution functions are indicated as f_n^0 and f_p^0.

C.1.2 Concentration integrals

The behaviour of conduction electrons in a semiconductor is not only determined by its momentum distribution, but also by the shape of the band in which it resides. Because of the Pauli exclusion principle, only a limited number of electrons is allowed in a band on a specific energy interval dE. For an energy E in band m this number is represented by the density of states $g_C^m(E,\mathbf{x})$ ([1], p.143):

$$g_C^m(E,\mathbf{x}) = \int_C \frac{d\mathbf{k}}{4\pi^3}\delta\left(E - E_C^m(\mathbf{k},\mathbf{x})\right) \tag{C.4}$$

where the integral is over any primitive cell, where δ is the Dirac delta function, and where $E_C^m(\mathbf{k},\mathbf{x})$ represents the shape of conduction band m. For this shape explicit expressions are found in Section 2.1 and Appendix B.

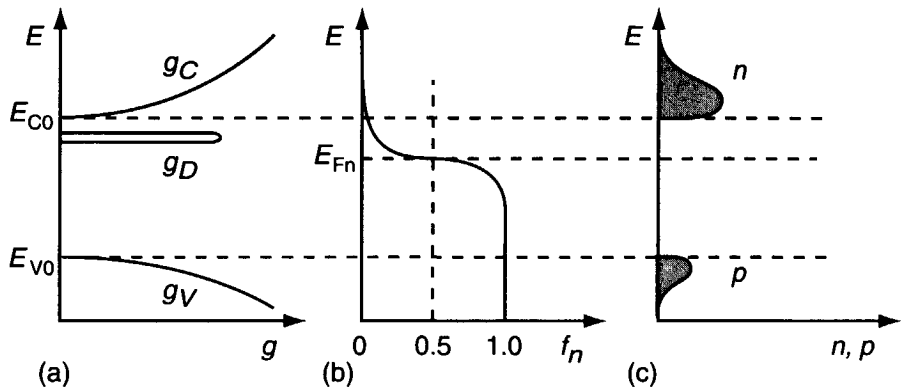

Figure C.2: Schematic calculation of the electron concentration in a conduction band. (a) Density of states as a function of energy E, given by the valence band (g_V), the doping (g_D), and the conduction band (g_C). (b) Fermi-Dirac distribution for electrons. (c) Electron population of the conduction band, resulting from the multiplication of a and b. The gray area is the total concentration n. The hole population is merely given for comparison.

For holes in a band u, the density of states is defined similarly:

$$g_V^u(E, \mathbf{x}) = \int_C \frac{d\mathbf{k}}{4\pi^3} \delta\left(E - E_V^u(\mathbf{k}, \mathbf{x})\right) \tag{C.5}$$

where $E_V^u(\mathbf{k}, \mathbf{x})$ gives the dispersion relation of band u.

The actual concentration of electrons due to band m is given by the number of quantum states around E, multiplied by the probability those states are filled, and integrated over all possible energies (see Figure C.2). This yields:

$$n^m = \int f_n g_C^m dE = \int_C \left[1 + \exp\left(\frac{E_C^m(\mathbf{k}, \mathbf{x}) - E_{Fn}(\mathbf{x})}{k_B T(\mathbf{x})}\right)\right]^{-1} \frac{d\mathbf{k}}{4\pi^3} \tag{C.6}$$

Equally, the hole concentration due to band u is given by:

$$p^u = \int f_p g_V^u dE = \int_C \left[1 + \exp\left(\frac{E_{Fp}(\mathbf{x}) - E_V^u(\mathbf{k}, \mathbf{x})}{kT(\mathbf{x})}\right)\right]^{-1} \frac{d\mathbf{k}}{4\pi^3} \tag{C.7}$$

As said before, the quasi Fermi levels E_{Fn} and E_{Fp} are equal in thermodynamic equilibrium. In other situations they are no more than matching parameters used to obtain the correct total concentrations of charge carriers.

C.1.3 Effective density of states

In semiconductor device analysis it is often useful to express a carrier concentration due to a band into an *effective* density of states and the energy of the

band *edge*. The effective density of states expresses the amount of charge carriers a band can contain, from its edge level up to infinite energy. Its concept, however, is usually defined for homogeneous, isotropic, non-degenerate semiconductors with parabolic bands. To use it beyond those assumptions, it has to be redefined and provided with a nonideality factor.

For this purpose the expression for the total electron energy in conduction band m is recalled from Equation 2.52: $E_C^m(\mathbf{k},\mathbf{x}) = E_{C0}^m(\mathbf{x}) + W_n^m(\mathbf{k},\mathbf{x})$, where E_{C0}^m is the energy of the band edge and W_n^m the kinetic energy. Also, it is assumed that the density of states of a band below its edge energy is zero. Then, in analogy with the classical expression, the local density of states $N_C^m(\mathbf{x})$ is defined as:

$$N_C^m(\mathbf{x}) = \int_0^\infty g_C^m(W_n^m,\mathbf{x}) \exp\left(\frac{-W_n^m}{k_B T(\mathbf{x})}\right) dW_n^m \tag{C.8}$$

Using this definition, the electron concentration n^m can be rewritten as:

$$n^m(\eta^m,\mathbf{x}) = N_C^m(\mathbf{x}) K_n^m(\eta^m,\mathbf{x}), \qquad \eta^m(\mathbf{x}) = E_{Fn}(\mathbf{x}) - E_{C0}^m(\mathbf{x}) \tag{C.9}$$

where η^m represents the difference between the band edge and the quasi Fermi level, and where the nonideality factor $K_n^m(\eta^m,\mathbf{x})$ is given by:

$$K_n^m(\eta^m,\mathbf{x}) = \int_0^\infty \frac{g_C^m(W_n^m,\mathbf{x})\, dW_n^m}{1+\exp\left(\frac{W_n^m-\eta^m}{k_B T}\right)} \cdot \frac{1}{\int_0^\infty g_C^m(W_n^m,\mathbf{x}) \exp\left(\frac{-W_n^m}{k_B T}\right) dW_n^m} \tag{C.10}$$

This factor approaches unity for nondegenerate conditions. In that case, the concentration is equal to:

$$n^m(\eta^m,\mathbf{x}) \simeq N_C^m(\mathbf{x}) \exp\left(\frac{\eta^m(\mathbf{x})}{k_B T(\mathbf{x})}\right) \tag{C.11}$$

Likewise, the local, effective density of states of a valence band u can be defined as:

$$N_V^u(\mathbf{x}) = \int_0^\infty g_V^u(W_p^u,\mathbf{x}) \exp\left(\frac{-W_p^u}{k_B T(\mathbf{x})}\right) dW_p^u \tag{C.12}$$

Then, the hole concentration p^u can be written for nondegenerate conditions as:

$$p^u(\mathbf{x}) = N_V^u(\mathbf{x}) \exp\left(\frac{E_{V0}^u(\mathbf{x}) - E_{Fp}(\mathbf{x})}{kT(\mathbf{x})}\right) \tag{C.13}$$

C.1.4 Density of states mass

It is customary to relate N_C^m and N_V^u to the local effective density-of-state masses m_{dC}^m and m_{dC}^u. They are *defined* as:

$$N_C^m(\mathbf{x}) \equiv 2\left(\frac{m_{dC}^m(\mathbf{x})\, k_B T(\mathbf{x})}{2\pi\hbar^2}\right)^{3/2}, \quad N_V^u(\mathbf{x}) \equiv 2\left(\frac{m_{dC}^u(\mathbf{x})\, k_B T(\mathbf{x})}{2\pi\hbar^2}\right)^{3/2} \tag{C.14}$$

The origin of those expressions lies in the analytical evaluation of Equation C.8 for the case of a parabolic energy band. For a conduction band m, for instance, we can combine Equation B.26 and 2.52 to find:

$$W_n^m = \frac{\hbar^2}{2} \sum_{ij} M_{ij} \left(k_i - k_i^0\right) \left(k_j - k_j^0\right) \tag{C.15}$$

where M_{ij} is the inverse-effective mass tensor. Since this tensor can be written in a diagonal form, the equation can be written in spherical coordinates by applying the Herring transformation [6, 7] :

$$\kappa_i = M_{ii}^{\frac{1}{2}} \left(k_i - k_i^0\right) \tag{C.16}$$

This yields:

$$W_n^m = \frac{\hbar^2}{2} \sum_j \kappa_j^2 = \frac{\hbar^2}{2} \kappa^2 \tag{C.17}$$

where κ is a radius defined as $\kappa^2 = \kappa_x^2 + \kappa_y^2 + \kappa_z^2$. Applying the transformation to the infinitesimal volume element $d\mathbf{k}$ gives:

$$d\mathbf{k} = \det\left(M_{ii}^{\frac{1}{2}}\right)^{-1} d\kappa = \left(m_l m_t m_t\right)^{\frac{1}{2}} d\kappa \tag{C.18}$$

where it has been used that two diagonal elements of M_{ij}^{-1} are equal to the inverse transverse mass m_t and one to the longitudinal mass m_l. The element can also be written in spherical coordinates $(\kappa, \varphi, \theta)$:

$$d\mathbf{k} = \left(m_{dC}^m\right)^{\frac{3}{2}} \kappa^2 \sin\theta d\kappa d\varphi d\theta \tag{C.19}$$

in which $m_{dC}^m = \left(m_l m_t m_t\right)^{1/3}$ is defined as the density of states effective mass. This expression can be inserted into the definition of N_C^m in Equation C.8. From Equation C.17 the κ^2 can now be replaced by W_n^m, and W_n^m can be normalised to $\zeta = W_n^m / \left(k_B T\right)$. This yields:

$$N_C^m = \frac{\left(m_{dC}^m k_B T\right)^{\frac{3}{2}}}{\sqrt{2}\pi^2\hbar^3} \int_0^\infty \zeta^{\frac{1}{2}} \exp\left(-\zeta\right) d\zeta \tag{C.20}$$

The integral is recognised as the Gamma function of order 3/2, $\Gamma\left(3/2\right)$, and is equal to $\sqrt{\pi}/2$ [8]. Inserting this value in Equation C.20 gives the relation between N_C^m and m_{dC}^m of Equation C.14.

It should be noted that for the valence bands the density of states masses cannot be expressed in such a simple function of the effective mass tensor elements. Instead, the integral of Equation C.12 needs to be evaluated numerically, after which the mass is calculated from Equation C.14.

C.2 Charge transport equations

It was mentioned in Section 2.2.1 that the current densities as a function of the applied fields can be derived from the semiclassical transport model. This is done by solving the Boltzmann equation for a first-order perturbation caused by external fields. The resulting distribution is a measure of the mean electron velocity, to which the current density is proportional. This current density can be expressed in a gradient of the quasi Fermi level and in the conductivity of the material. The conductivity can be split up into a carrier concentration and a mobility, which can sometimes be evaluated analytically in terms of band parameters. In addition, the gradient of the quasi Fermi level can be split up into band edge parameters, resulting in a generalised form of the classical drift-diffusion equations.

C.2.1 Solution of the Boltzmann equation

The semiclassical transport model describes the development in time of the electron distribution function f_n by means of the Boltzmann transport equation:

$$\frac{d}{dt} f_n(\mathbf{x}, \mathbf{k}, t) = \widehat{S}_{sc} f_n(\mathbf{x}, \mathbf{k}, t) \tag{C.21}$$

where $\widehat{S}_{sc}$ is an operator representing the scattering of the electrons. Both sides of this equation can be worked out to find a more explicit expression.

On the left hand side the total temporal derivative can be developed as:

$$\frac{d}{dt} f_n(\mathbf{x}, \mathbf{k}, t) = \frac{\partial f_n}{\partial t} + \frac{\partial f_n}{\partial x_i}\frac{dx_i}{dt} + \frac{\partial f_n}{\partial k_i}\frac{dk_i}{dt} \tag{C.22}$$

Some of those terms can be substituted using Hamilton's (classical) equations of motion in which the momentum $\mathbf{p}$ has been substituted by the crystal momentum $\hbar\mathbf{k}$:

$$\frac{dx_i}{dt} = \frac{1}{\hbar}\frac{\partial \widehat{H}}{\partial k_i} = v_i \tag{C.23}$$

$$\frac{dk_i}{dt} = -\frac{1}{\hbar}\frac{\partial \widehat{H}}{\partial x_i} \tag{C.24}$$

where v_i is the electron velocity. For an electron in the conduction band m the Hamiltonian $\widehat{H}$ is equal to its total energy and can be substituted by $E_C^m(\mathbf{x}, \mathbf{k})$ of Equation 2.52.

On the right hand side of the Boltzmann equation it is assumed that the distribution function is perturbed by the external fields only to linear order. It is also assumed that the scattering is described by the relaxation time approximation mentioned in Section C.1. Finally it is assumed that the relaxation time $\tau_m(\mathbf{x}, \mathbf{k})$ of band m depends on $\mathbf{k}$ only through the electron kinetic energy $W_n^m(\mathbf{k}, \mathbf{x})$ which is part of its total energy E_C^m (see Equation 2.52). This means

that a collision obliterates any information about the direction of the electron velocity. In that case the scattering operator can be written as:

$$\widehat{S}_{sc} = -\frac{f_n - f_n^0}{\tau_m\left(\mathbf{x}, W_n^m\right)} = -\frac{f_n^1}{\tau_m} \tag{C.25}$$

where f_n^0 is the local Fermi-Dirac distribution in equilibrium and f_n^1 is its perturbation due to the fields.

Inserting the above expressions in Equation C.21, the transport equation for an electron in band m becomes:

$$\frac{\partial f_n}{\partial t} + \frac{1}{\hbar}\frac{\partial f_n}{\partial x_i}\frac{\partial E_C^m}{\partial k_i} - \frac{1}{\hbar}\frac{\partial f_n}{\partial k_i}\frac{\partial E_C^m}{\partial x_i} = -\frac{f_n^1}{\tau_m} \tag{C.26}$$

It is now assumed that the currents are stationary. This means that the perturbation can be written as:

$$f_n^1 = -\frac{\tau_m}{\hbar}\left(\frac{\partial f_n}{\partial x_i}\frac{\partial E_C^m}{\partial k_i} - \frac{\partial f_n}{\partial k_i}\frac{\partial E_C^m}{\partial x_i}\right) \tag{C.27}$$

In a similar way the distribution of holes can be approached. The perturbation f_p^1 in the hole distribution function f_p due to external fields appears to be:

$$f_p^1 = -\frac{\tau_u}{\hbar}\left(\frac{\partial f_p}{\partial x_i}\frac{\partial E_V^u}{\partial k_i} - \frac{\partial f_p}{\partial k_i}\frac{\partial E_V^u}{\partial x_i}\right) \tag{C.28}$$

where $E_V^u = E_V^u(\mathbf{x}, \mathbf{k})$ is the dispersion relation of valence band u, $\tau_u = \tau_u\left(\mathbf{x}, W_p^u\right)$ its relaxation time, and W_p^u the hole kinetic energy.

C.2.2 Current density and conductivity

Once the distribution function is known in terms of statistical and band functions it is possible to derive the electron current density as a function of the quasi Fermi level.

The electron current density vector due to a band m is determined by the concentration of electrons n^m, their mean velocity $\langle v_i^m \rangle$, and their charge q. It is therefore defined as:

$$J_i^m = -q n_m \langle v_i^m \rangle \tag{C.29}$$

In the semiclassical model the fraction of the electrons with a velocity $v_i^m(E)$ is given by the distribution function $f_n(E(\mathbf{k}))$. Therefore $\langle v_i^m \rangle$ can be written as the average:

$$\begin{aligned}\langle v_i^m \rangle &= \frac{1}{n^m}\int_0^\infty g_C^m(E)\, v_i^m(E)\, f_n(E)\, dE \\ &= \frac{1}{n^m}\int_C \frac{d\mathbf{k}}{4\pi^3} v_i^m\left(E_C^m\right) f_n\left(E_C^m\right)\end{aligned} \tag{C.30}$$

where the density of states function of Equation C.4 was inserted. In the preceding section, f_n was split up into an equilibrium and a perturbation term: $f_n = f_n^0 + f_n^1$. In addition, the velocity was given by the Hamiltonian expression of Equation C.23. When this is used in Equation C.30, J_i^m can be rewritten as:

$$J_i^m = -\frac{q}{\hbar}\int_C \frac{d\mathbf{k}}{4\pi^3}\frac{\partial E_C^m}{\partial k_i}\left(f_n^0 + f_n^1\right) = -\frac{q}{\hbar}\int_C \frac{d\mathbf{k}}{4\pi^3}\frac{\partial E_C^m}{\partial k_i} f_n^1\left(E_C^m\right) \tag{C.31}$$

because the current density in equilibrium is zero.

The distribution function can be eliminated from this equation by firstly replacing f_n^1 by Equation C.27:

$$J_i^m = \frac{q}{\hbar^2}\int_C \frac{d\mathbf{k}}{4\pi^3}\tau_n \frac{\partial E_C^m}{\partial k_i}\left(\frac{\partial f_n}{\partial x_j}\frac{\partial E_C^m}{\partial k_j} - \frac{\partial f_n}{\partial k_j}\frac{\partial E_C^m}{\partial x_j}\right) \tag{C.32}$$

Secondly, it was assumed in the relaxation time approximation that the shape of f_n is equal to that of the Fermi-Dirac distribution function. The gradients of f_n can therefore be calculated by applying the chain rule to Equation C.1, which yields:

$$\frac{\partial f_n}{\partial x_j} = \frac{\partial f_n}{\partial E_C^m}\left[\frac{\partial E_C^m}{\partial x_j} - \frac{\partial E_{Fn}}{\partial x_j} - \left(\frac{E_C^m - E_{Fn}}{T}\right)\frac{\partial T}{\partial x_j}\right] \tag{C.33}$$

$$\frac{\partial f_n}{\partial k_j} = \frac{\partial f_n}{\partial E_C^m}\frac{\partial E_C^m}{\partial k_j} \tag{C.34}$$

Inserting these in Equation C.32 results in:

$$J_i^m = -\frac{q}{\hbar^2}\int \frac{d\mathbf{k}}{4\pi^3}\tau_n \frac{\partial E_C^m}{\partial k_i}\left[\frac{\partial E_C^m}{\partial k_j}\frac{\partial E_{Fn}}{\partial x_j} + \left(\frac{E_C^m - E_{Fn}}{T}\right)\frac{\partial T}{\partial x_j}\right]\frac{\partial f_n}{\partial E_C^m} \tag{C.35}$$

where $\partial f_n / \partial E_C^m$ can simply be evaluated from Equation C.1. It is seen that J_i^m is driven by the gradients in both the quasi Fermi level and the temperature. The above expression therefore leads to the so-called hydrodynamic equations which enable the modelling of the purely electric as well as the thermoelectric effects in the material.

In the following, however, it is assumed that the thermoelectric effects are negligible by supposing isothermal conditions. It yields the equation:

$$J_i^m(\mathbf{x}) = \sigma_{ij}^m(\mathbf{x})\frac{\partial E_{Fn}}{q\partial x_j}(\mathbf{x}) \tag{C.36}$$

$$\sigma_{ij}^m(\mathbf{x}) = -\frac{q^2}{\hbar^2}\int \frac{d\mathbf{k}}{4\pi^3}\left[\tau_n \frac{\partial E_C^m}{\partial k_i}\frac{\partial E_C^m}{\partial k_j}\frac{\partial f_n}{\partial E_C^m}\right](\mathbf{x}) \tag{C.37}$$

where the tensor σ_{ij}^m is defined as the *electrical conductivity* due to band m. The gradient in the quasi Fermi level was left outside the integral since it does not depend on the wave number. In the case of nondegenerate material f_n can be replaced by the Maxwell-Boltzmann distribution giving:

$$\sigma_{ij}^m(\mathbf{x}) \simeq \frac{q^2}{\hbar^2 k_B T}\int \frac{d\mathbf{k}}{4\pi^3}\left[\tau_n \frac{\partial E_C^m}{\partial k_i}\frac{\partial E_C^m}{\partial k_j}\exp\left(\frac{E_{Fn} - E_C^m}{k_B T}\right)\right](\mathbf{x}) \tag{C.38}$$

From the above equations it appears that the driving force behind the current is the gradient in the quasi Fermi level. The current is influenced by the material characteristics $E_C^m(\mathbf{k},\mathbf{x})$ and $\tau_n(E_C^m,\mathbf{x})$. It is also influenced by parameters that can be adjusted from outside: T and E_{Fn}, the latter being equivalent to the electron concentration.

Similar expressions can be derived for holes in band u:

$$J_i^u(\mathbf{x}) = \sigma_{ij}^u(\mathbf{x}) \frac{\partial E_{Fp}}{q\partial x_j}(\mathbf{x}) \tag{C.39}$$

$$\sigma_{ij}^u(\mathbf{x}) \simeq \frac{q^2}{\hbar^2 k_B T} \int \frac{d\mathbf{k}}{4\pi^3} \left[\tau_p \frac{\partial E_V^u}{\partial k_i} \frac{\partial E_V^u}{\partial k_j} \exp\left(\frac{E_V^u - E_{Fp}}{k_B T}\right)\right](\mathbf{x}) \tag{C.40}$$

C.2.3 Mobility

The conductivity is not always an ideal parameter to characterise the electrical properties of a semiconductor. The reason for this is its strong dependency on the quasi Fermi level, which is easily modified from outside by injecting charge carriers. Therefore the concept of *mobility* was introduced. For electrons in band m and holes in band u it is defined as:

$$\mu_{ij}^m(E_{Fn},\mathbf{x}) = \frac{\sigma_{ij}^m(E_{Fn},\mathbf{x})}{qn^m(E_{Fn},\mathbf{x})}, \qquad \mu_{ij}^u(E_{Fp},\mathbf{x}) = \frac{\sigma_{ij}^u(E_{Fp},\mathbf{x})}{qp^u(E_{Fp},\mathbf{x})} \tag{C.41}$$

An evaluation of μ_{ij}^m with the aid of Equation C.6 and C.37 learns that it is a material constant. It is independent of E_{Fn}, but only in the case of nondegenerate material.

For this nondegenerate case and for parabolic bands the mobility can be evaluated analytically. Equation C.38 is therefore first rewritten as:

$$\sigma_{ij}^m = \frac{q^2}{\hbar^2 k_B T} \exp\left(\frac{E_{Fn} - E_{C0}^m}{k_B T}\right) \int \frac{d\mathbf{k}}{4\pi^3} \tau_n \frac{\partial W_n^m}{\partial k_i} \frac{\partial W_n^m}{\partial k_j} \exp\left(\frac{-W_n^m}{k_B T}\right) \tag{C.42}$$

where it has been used that $E_C^m = E_{C0}^m + W_n^m$. The expression for n^m in Equation C.9 permits to write:

$$\mu_{ij}^m(\mathbf{x}) = \frac{q}{\hbar^2 k_B T} \frac{1}{N_C^m(\mathbf{x})} \int \frac{d\mathbf{k}}{4\pi^3} \left[\tau_n \frac{\partial W_n^m}{\partial k_i} \frac{\partial W_n^m}{\partial k_j} \exp\left(\frac{-W_n^m}{k_B T}\right)\right](\mathbf{x}) \tag{C.43}$$

The relaxation time is supposed to depend only on the electron kinetic energy W_n^m (see Section C.1) and is usually approached by some power law in s:

$$\tau_m(W_n^m,\mathbf{x}) = \tau_m^0(\mathbf{x}) \left(\frac{W_n^m(\mathbf{k},\mathbf{x})}{k_B T}\right)^s \tag{C.44}$$

where s is approximately equal to $-1/2$ for nondegenerate silicon [3].

The derivatives $\partial W_n^m/\partial k_i$ are obtained from Equation C.15. If the principal axes of the inverse effective mass tensor M_{ij} are perpendicular to the coordinate axes, M_{ij} is diagonal and $\partial W_n^m/\partial k_i$ is equal to $\hbar^2 M_{ii}\left(k_i - k_i^0\right)$.

The integral is now calculated further along the lines pointed out in Section C.1.4. This means that a Herring transformation is applied to normalise the wave vector k_i to κ_i. Then the integration is performed over the solid angle. The remaining integrand contains a factor $\kappa_i \kappa_j$ which is nonzero only if $i = j$ due to the symmetry of the integrand. In addition, it is equal to $\kappa^2/3$ due to the normalisation of the vector components. Now κ^2 is replaced by the energy W_n^m, which is then normalised with $k_B T$ to ζ. The result is:

$$\mu_{ij}^m = \frac{2\sqrt{2} q \tau_n^0 (k_B T)^{\frac{3}{2}}}{3h^3 \pi^2 N_C^m} (m_d^m)^{\frac{3}{2}} M_{ii} \int_0^\infty \zeta^{s+\frac{3}{2}} \exp(-\zeta)\, d\zeta \tag{C.45}$$

The integral can be recognised as the Gamma function $\Gamma(s + 5/2)$ and is equal to 1 for $s = -1/2$ [8]. The effective density of states of a parabolic band was known already from Equation C.14. Together with the values for M_{ii} this yields the diagonal mobility tensor:

$$\mu_{ij}^m = \frac{4q\tau_n^0}{3\sqrt{\pi} m_{l,t}} \delta_{ij} \tag{C.46}$$

where the choice for the longitudinal mass m_l or the transversal mass m_t depends on whether the current direction i is parallel or perpendicular to the axis of revolution of band m.

C.2.4 Generalised drift-diffusion equations

In Section C.2.2 and C.2.3 the electron and hole current densities were derived as a function of the quasi Fermi levels:

$$J_i^m = n^m \mu_{ij}^m \frac{\partial E_{Fn}}{\partial x_j}\,, \qquad J_i^u = p^u \mu_{ij}^u \frac{\partial E_{Fp}}{\partial x_j} \tag{C.47}$$

However, sometimes it is convenient to separate them into a drift and a diffusion part. This will enable a comparison with the classical drift-diffusion equations valid for isotropic, homogeneous material. For this purpose the gradient in the quasi Fermi level $\partial E_{Fn}/\partial x_i$ should be developed.

This is possible by taking the gradient of the nonequilibrium electron concentration $n^m(\eta^m, \mathbf{x})$, given by Equation C.9, where η^m was an abbreviation of the energy difference $E_{Fn} - E_{C0}^m$. The gradient of n^m can formally be written as:

$$\frac{dn^m}{dx_j} = \frac{\partial n^m}{\partial \eta^m} \frac{\partial \eta^m}{\partial x_j} + \frac{d^\eta n^m}{dx_j} \tag{C.48}$$

where $d^\eta n/dx_j$ is *defined* as the gradient of n^m after all variables *except* after η^m. When η^m is replaced by $E_{Fn} - E_{C0}^m$, a rearrangement of terms gives:

$$\frac{\partial E_{Fn}}{\partial x_j} = \frac{\partial E_{C0}^m}{\partial x_j} + \left(\frac{\partial n^m}{\partial \eta^m}\right)^{-1} \frac{dn^m}{dx_j} - \left(\frac{\partial n^m}{\partial \eta^m}\right)^{-1} \frac{d^\eta n^m}{dx_j} \tag{C.49}$$

The gradient in the band edge $\partial E_{C0}^m/\partial x_j$ can be expressed as a function of the electrostatic drift field F_j^{el} by differentiating the expression found in Section 2.1.7:

$$\frac{\partial E_{C0}^m}{\partial x_j} = -\frac{\partial \chi^m}{\partial x_j} - q\frac{\partial \phi}{\partial x_j} \tag{C.50}$$

where χ^m is the electron affinity of conduction band m. The gradient of the electrostatic potential ϕ is by definition equal to $-F_j^{\text{el}}$.

Combining the above information with Equation C.47, the electron current density can be written as:

$$J_i^m = n^m \mu_{ij}^m \left(-\frac{\partial \chi^m}{\partial x_j} - q\frac{\partial \phi}{\partial x_j}\right) + qD_{ij}^m \frac{dn^m}{dx_j} - n^m \mu_{ij}^m \frac{\partial \Gamma_n^m}{\partial x_j} \tag{C.51}$$

where D_{ij}^m is the *diffusion tensor*, defined by the *generalised Einstein relation:*

$$qD_{ij}^m \equiv n^m \mu_{ij}^m \left(\frac{\partial n^m}{\partial \eta^m}\right)^{-1} \tag{C.52}$$

and where Γ_n describes the inhomogeneity of the material. Its gradient is defined as:

$$\frac{\partial \Gamma_n^m}{\partial x_j} \equiv \left(\frac{\partial n^m}{\partial \eta^m}\right)^{-1} \frac{d^\eta n^m}{dx_j} \tag{C.53}$$

In combination with Equation C.52 it reveals the diffusive nature of the last term of Equation C.51:

$$n^m \mu_{ij}^m \frac{\partial \Gamma_n^m}{\partial x_j} = qD_{ij}^m \frac{d^\eta n^m}{dx_j} \tag{C.54}$$

A similar derivation can be made for the hole current density J_i^u. The results are:

$$J_i^u = p^u \mu_{ij}^u \left(-\frac{\partial \chi^m}{\partial x_j} - q\frac{\partial \phi}{\partial x_j} - \frac{\partial E_G^{mu}}{\partial x_j}\right) - qD_{ij}^u \frac{dp^u}{dx_j} + p^u \mu_{ij}^u \frac{\partial \Gamma_p^u}{\partial x_j} \tag{C.55}$$

where E_G^{mu} was the bandgap between the valence band u and the conduction band m. The hole diffusion tensor D_{ij}^u and the additional hole diffusion function $\partial \Gamma_p/\partial x_j$ are defined as:

$$qD_{ij}^u \equiv p^u \mu_{ij}^u \left(\frac{\partial p^u}{\partial \eta^u}\right)^{-1}, \quad \frac{\partial \Gamma_p^u}{\partial x_j} \equiv \left(\frac{\partial p^u}{\partial \eta^u}\right)^{-1} \frac{d^\eta p^u}{dx_j} \tag{C.56}$$

The η^u in those expressions is equal to $E_{V0}^u - E_{Fp}$.

The generalised Einstein relation and the additional diffusion functions $\partial \Gamma_n^m/\partial x_j$ and $\partial \Gamma_p^u/\partial x_j$ can be much simplified when the material is nondegenerate. It

was shown in Section C.1 that in that case the electron concentration can be written as:

$$n^m = N_C^m \exp\left(\frac{\eta^m}{k_B T}\right) \tag{C.57}$$

As a result, its derivatives are equal to:

$$\frac{\partial n^m}{\partial \eta^m} = \frac{n^m}{k_B T}\,, \qquad \frac{d^\eta n^m}{dx_j} = \frac{d \ln N_C^m}{dx_j} \tag{C.58}$$

When also the definitions of mobility in Equation C.41 are used, it is possible to write the generalised drift-diffusion equations for electrons and holes as:

$$J_i^m = \sigma_{ij}^m \left[-\frac{\partial \phi}{\partial x_j} - \frac{1}{q}\frac{\partial \chi^m}{\partial x_j} + \frac{k_B T}{q}\left(\frac{d}{dx_j}\ln n^m - \frac{d}{dx_j}\ln N_C^m\right)\right] \tag{C.59}$$

$$J_i^u = \sigma_{ij}^u \left[-\frac{\partial \phi}{\partial x_j} - \frac{1}{q}\frac{\partial \chi^m}{\partial x_j} - \frac{1}{q}\frac{\partial E_G^{mu}}{\partial x_j} - \frac{k_B T}{q}\left(\frac{d}{dx_j}\ln p^u - \frac{d}{dx_j}\ln N_V^u\right)\right] \tag{C.60}$$

The gradient $-\partial\phi/\partial x_j$ is equal to the electrostatic field F_j^{el}. If the material is isotropic and homogeneous, it can be verified that the equations reduce to the classical case of Equation 2.54 and 2.55.

C.3 Auxiliary relations

The current density equations are generally accompanied by a number of auxiliary relations in order to solve some unknown parameters. These relations include the *pn* product, the continuity equations, and the Poisson equations. Their appearance changes in the presence of mechanical stress.

C.3.1 The *pn* product

The *pn* product is often used as a boundary condition to calculate the minority carrier concentration just behind a junction [9]. In addition, it is a material constant in thermodynamic equilibrium for the case of nondegenerate material. It can be determined experimentally and used to characterise the band structure. Although the concept is also applicable to highly doped material [4], it is worked out here only for the nondegenerate case. Even then it can be defined in three different ways.

Firstly, the *pn* product can be defined for one conduction band m and one

valence band u. With the aid of Equation C.9 and C.13 it can be written as:

$$\begin{aligned}(n_i^{mu})^2(\mathbf{x}) &= n^m(\mathbf{x})\,p^u(\mathbf{x}) \\ &= (n_{i0}^{mu})^2(\mathbf{x})\exp\left(\frac{E_{Fn}-E_{Fp}}{k_BT}\right)(\mathbf{x}) \qquad \text{(C.61)}\end{aligned}$$

$$\text{where}: (n_{i0}^{mu})^2(\mathbf{x}) = N_C^m N_V^u \exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_BT}\right)(\mathbf{x}) \qquad \text{(C.62)}$$

Here, $(n_{i0}^{mu})^2$ is the pn product of the bands m and u in thermodynamic equilibrium and is independent of the quasi Fermi levels.

Secondly, the pn product can be defined for one conduction band m with respect to *all* valence bands u:

$$\begin{aligned}(n_i^m)^2(\mathbf{x}) &= n^m(\mathbf{x})\sum_u p^u(\mathbf{x}) \\ &= (n_{i0}^m)^2(\mathbf{x})\exp\left(\frac{E_{Fn}-E_{Fp}}{k_BT}\right)(\mathbf{x}) \qquad \text{(C.63)}\end{aligned}$$

$$\text{where}: (n_{i0}^m)^2(\mathbf{x}) = N_C^m(\mathbf{x})\sum_u N_V^u(\mathbf{x})\exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_BT}\right)(\mathbf{x}) \qquad \text{(C.64)}$$

where the equilibrium product $(n_{i0}^m)^2$ is again independent of the quasi Fermi levels. In a similar way the pn product can be defined with respect to one conduction band m and all valence bands u.

Finally, the pn product can be defined for *all* conduction bands and *all* valence bands:

$$\begin{aligned}n_i^2(\mathbf{x}) &= \sum_m\sum_u n^m(\mathbf{x})\,p^u(\mathbf{x}) \\ &= n_{i0}^2(\mathbf{x})\exp\left(\frac{E_{Fn}-E_{Fp}}{k_BT}\right)(\mathbf{x}) \qquad \text{(C.65)}\end{aligned}$$

$$\text{where}: n_{i0}^2(\mathbf{x}) = \sum_m\sum_u N_C^m(\mathbf{x})\,N_V^u(\mathbf{x})\exp\left(\frac{E_{V0}^u - E_{C0}^m}{k_BT}\right)(\mathbf{x}) \qquad \text{(C.66)}$$

This definition is equal to the intrinsic carrier concentration that is traditionally used to describe stress-free material. It turns out, however, that it is not very practical to calculate the characteristics of stressed devices.

C.3.2 Continuity equations

The continuity equations describe how much charge is flowing to and from a specific point. They state that electrons and holes are either generated, stay in the same band, or recombines with each other. They therefore describe the

changes in the concentrations in time. This can be formulated as:

$$\frac{\partial n^m}{\partial t}(\mathbf{x}) = G_n^m(\mathbf{x}) - R_n^m(\mathbf{x}) + \frac{1}{q}\sum_{i=1}^{3}\frac{\partial J_i^m}{\partial x_i}(\mathbf{x}) \tag{C.67}$$

$$\frac{\partial p^u}{\partial t}(\mathbf{x}) = G_p^u(\mathbf{x}) - R_p^u(\mathbf{x}) - \frac{1}{q}\sum_{i=1}^{3}\frac{\partial J_i^u}{\partial x_i}(\mathbf{x}) \tag{C.68}$$

where G_n^m and G_p^u are the local generation rates and R_n^m and R_p^u the recombination rates of the conduction band m and the valence band u.

Both the generation and the recombination of electron-hole pairs can be described by lifetimes. The generation and recombination lifetimes of bulk materials are quite different, the former being up to 100 times higher [10].

Generation and recombination can occur through different mechanisms. In silicon with free carrier concentrations up to $5\text{x}10^{18}$ cm^{-3} the most important mechanism is that of Shockley, Read, and Hall [11]. It occurs through deep-level impurities, causing states around the middle of the forbidden band. Under the assumptions that:

- The system is in steady state,
- The number of traps is small compared to the majority carrier concentration, and
- The injected minority concentration does not exceed the doping level (low-level injection),

the recombination rates of the minorities R_n and R_p can be written as:

$$R_n^m = \frac{n^m - n_0^m}{\tau_{\text{SRH}}^m}; \qquad R_p^u = \frac{p^u - p_0^u}{\tau_{\text{SRH}}^u} \tag{C.69}$$

where n_0^m and p_0^u are the electron and hole concentrations in thermal equilibrium.

The τ_{SRH}^m and τ_{SRH}^u are the electron and hole Shockley-Read-Hall lifetimes of bands m and u, respectively. They are related to the crystal defects and to the doping concentration. Therefore they are quite technology-dependent. Their value begins to decrease at concentrations greater than 10^{17} cm^{-3}. Often, it is assumed that they are equal for all bands of a type. They can be written as a function of the impurity concentration N_T, the thermal velocity of the carrier v_{th}, and the capture cross section of the impurity γ. This yields:

$$\tau_{\text{SRH}}^m = \frac{1}{\gamma_n v_{th} N_T}; \quad \tau_{\text{SRH}}^u = \frac{1}{\gamma_p v_{th} N_T} \tag{C.70}$$

At high free carrier concentrations the generation and recombination is dominated by the Auger mechanism. This creates electron-hole pairs by the impact of a hot electron, or annihilates a pair by the inverse mechanism. The associated

lifetimes depend on the Auger coefficients C_n and C_p and on the square of the majority carrier concentration. For low-level injection they can be written as:

$$\tau^m_{\text{Aug}} = \frac{1}{C_p p_0^2}; \quad \tau^u_{\text{Aug}} = \frac{1}{C_n n_0^2} \tag{C.71}$$

Both the SRH and the Auger lifetimes have been determined experimentally on many occasions [10]. They can also be derived from quantummechanical considerations, where photons and phonons perturb the electron eigenfunctions [12, 13].

C.3.3 The Poisson equation

The Poisson equation relates the total space charge concentration ρ^{ch} to the electric field $\mathbf{F}^{\text{el}}$, and is especially useful for calculating device behaviour with a finite-element method. Under low frequency conditions it reads:

$$\nabla \cdot \left(\varepsilon_0 \varepsilon_r \mathbf{F}^{\text{el}}\right) = \rho^{\text{ch}} \tag{C.72}$$

where ε_0 is the electrical permittivity of vacuum and ε_r is the dielectric constant of the material. The field is related to the electrostatic potential ϕ as $\mathbf{F}^{\text{el}} = -\nabla\phi$.The space charge is given by the sum of the free carrier concentrations and the ionized doping atom concentration N_A^- and N_D^+. Therefore Equation C.72 can also be written as:

$$\varepsilon_r \nabla^2 \phi + \nabla \varepsilon_r \cdot \nabla \phi = \frac{q}{\varepsilon_0} \left(\sum_m n^m - \sum_u p^u + N_A^- - N_D^+ \right) \tag{C.73}$$

Normally, the material is isotropic and homogeneous, so that ε_r is a scalar and its gradient is zero. In a stressed material, however, it is generally a tensor and when the stress in inhomogeneous, its gradient is nonzero.

C.4 Conclusions

Even for anisotropic, inhomogeneous material with degenerate bands the current density is proportional to the conductivity and the gradient in the quasi Fermi level, as described by Equation C.36 and C.39. For nondegenerate material it is practical to split the conductivity into a carrier concentration and a mobility representing the influence of the material. Sometimes it is also practical to split the quasi Fermi level gradient into a drift and a diffusion component. It then appears that the currents are not only driven by the electrostatic field and the concentration gradient, but also by gradients in the effective density of states, the band gap, and the electron affinity. The inhomogeneity behind those forces also leads to an extra term in the Poisson equation.

Bibliography

[1] N. M. Ashcroft and N. D. Mermin, *Solid state physics*, Saunders College, Philadelphia, 1976.

[2] A. van der Ziel, *Solid State Physical Electronics*, 3rd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

[3] M. Lundstrom, *Fundamentals of carrier transport, Modular series on solid state devices, X*, Addison-Wesley, Reading, Mass., 1992.

[4] A. H. Marshak and C. M. van Vliet, Electrical current and carrier density in degenerate materials with nonuniform band structure, *Proc. IEEE, 72* (1984) 148-164.

[5] W. Greiner, *Quantum mechanics, an introduction*, Springer, Berlin, 1989.

[6] C. Herring and E. Vogt, Transport and deformation-potential theory for many-valley-semiconductors with anisotropic scattering, *Phys. Rev., 101* (1956) 944-961.

[7] C. Herring, Transport properties of a many-valley semiconductor, *Bell Syst. Techn. J., 34* (1955) 237-290.

[8] M. Abramowitz and I. M. Stegun, *Handbook of mathematical functions; with formulas, graphs, and mathematical tables*, 9th pr., Dover, New York, 1970.

[9] A. H. Marshak and R. Shrivastava, Law of the junction for degenerate material with position-dependent band gap and electron affinity, *Solid-St. Electron., 22* (1979) 567-571.

[10] D. K. Schroder, *Semiconductor material and device characterization*, Wiley, New York, 1990.

[11] P. A. Gough, *Fundamental equations*, in: P. A. H. Hart (ed.), Bipolar and bipolar-MOS integration, Elsevier, Amsterdam, 1994.

[12] J. Singh, *Physics of semiconductors and their heterostructures*, McGraw-Hill, New York, 1993.

[13] J. Singh, *Semiconductor optoelectronics, McGraw-Hill*, New York, 1995.

Appendix D

Current through a Bipolar Transistor

Probably the most important characteristic of a bipolar transistor is the exponential relation between the collector current I_c and the base-emitter voltage V_{be} in forward bias. This relation is usually written as:

$$I_c = I_S \exp\left(\frac{qV_{be}}{k_B T}\right) \tag{D.1}$$

in which I_S is the saturation current, q the elementary charge, k_B the Boltzmann constant, and T the absolute temperature. In practice, this relation appears to be accurate over many decades of current. It can be influenced by mechanical stress only through I_S. Other transistor characteristics are determined by the forward current gain, the Early voltage, the base transit time, and the depletion layer capacitances. Their changes with stress, however, are far less important for the transistor behaviour in a circuit. Therefore they are not modelled here.

In this appendix the saturation current is analytically modelled in terms of transport parameters. For this purpose the Gummel-Poon model is followed [1]-[4]. Although it considers a device of only one dimension, it has proven to be remarkably accurate. It has the advantage over the simpler Ebers-Moll model [5] that it allows many parameters to vary with position: the doping, the mobility, and the intrinsic carrier concentration. For this reason it can be used to model the effects of inhomogeneous stress.

The appendix starts by discussing the one-directional abstraction of the transistor geometry. Then it presents assumptions about its construction and its operating conditions. It derives the current densities in the base, and relates them to the voltage differences between the transistor terminals. Finally, it derives the collector current in the form of Equation D.1, and the saturation current in terms of geometry and transport parameters.

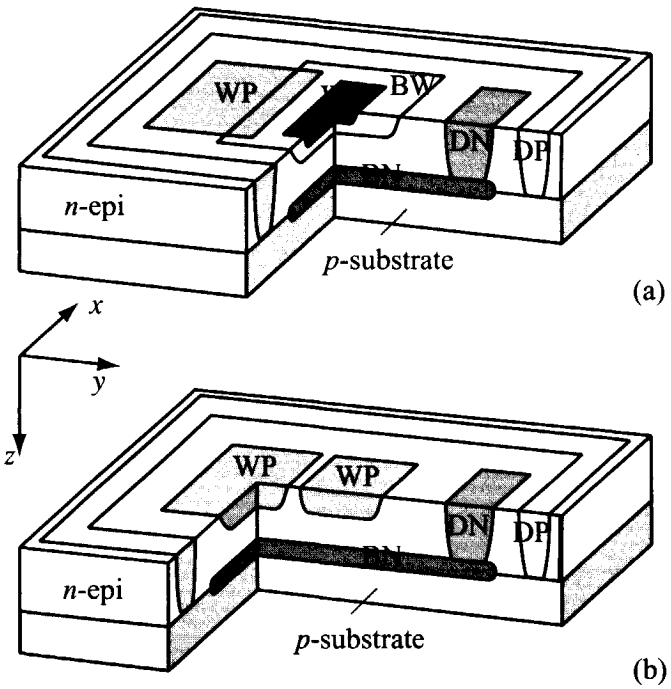

Figure D.1: 3D schematics of conventional bipolar transistors. (a) Vertical *npn*. DN: collector plug, BN: buried layer, BW: intrinsic basis, WP: extrinsic basis, WN: emitter, DP: isolation. (b) Lateral *pnp*. DN: base plug, WP: collector and emitter.

D.1 The one-dimensional approximation

A vertical *npn* transistor is often constructed as the one shown in Figure D.1.a. Its collector is formed by a homogeneous epitaxial layer on a *p*-type substrate. This epi-layer has a donor concentration of about 10^{16} cm^{-3} and is contacted by a highly-doped buried layer and a deep collector plug. In the epi-collector an intrinsic base has been implanted with an acceptor concentration of about 10^{17} cm^{-3} in the active region. It is contacted with a higher doped extrinsic base. In the intrinsic base the emitter has been implanted with a donor concentration of approximately 10^{19} cm^{-3}. Around the collector, base, and emitter a ring with a high acceptor concentration has been diffused in order to isolate the transistor from other devices.

In the same fabrication process a *pnp*-transistor is usually constructed laterally, as in Figure D.1.b. Now the epi-layer serves as the base and two donor implantations as the collector and emitter.

The three-dimensional structures of Figure D.1 can be better modelled when

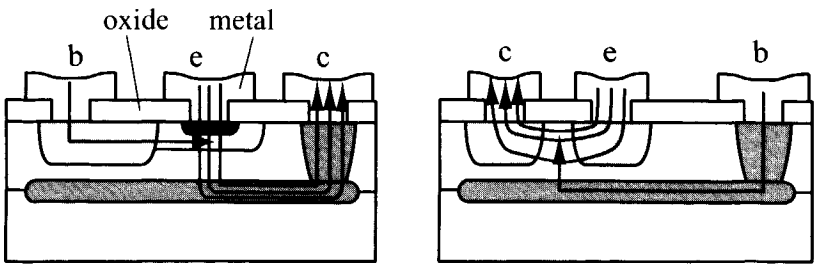

Figure D.2: 2D schematic of a vertical *npn* transistor (a) and a lateral *pnp* transistor (b). The main flows of the charge carriers are indicated with arrows. The metal connection e is the emitter, b the base, and c the collector.

they are simplified to the two-dimensional cross-sections of Figure D.2. This considerably reduces the time of calculations needed. It also introduces some minor errors due to the neglect of edge effects and the spreading of the collector current [3].

The figure also shows the direction of the main charge carrier flows in the transistors if they are biased in forward. The main current direction in the base is perpendicular to the substrate surface for the vertical transistor. For the lateral *pnp*, however, it is directed along the surface. This difference plays a role when the response to mechanical stress is considered.

The two-dimensional structure can only be modelled analytically if it is reduced to the one-dimensional cross-section in Figure D.3. Again, the edge-effects are neglected. Figure D.3.a represents the effective doping profile of the vertical *npn* for a cross section through the emitter. Figure D.3.b sketches the electrical connections in forward bias. It also indicates the principal current flows inside the transistor. The collector current I_c mainly consists of electrons crossing the emitter-base junction due to the positive V_{be}. They diffuse through the base as minority charge carriers. Finally, they are collected at the collector-base junction. The base current I_b mainly consists of holes injected in the base which are able to cross the base-emitter junction. Since the emitter doping is much higher than that of the base, I_b is much smaller than I_c. Figure D.3.c, finally, sketches the band edge diagram throughout the transistor and indicates the quasi Fermi levels in forward bias.

D.2 Assumptions

The Gummel approach for deriving the collecor current makes some assumptions about the transistor construction and the operating conditions:

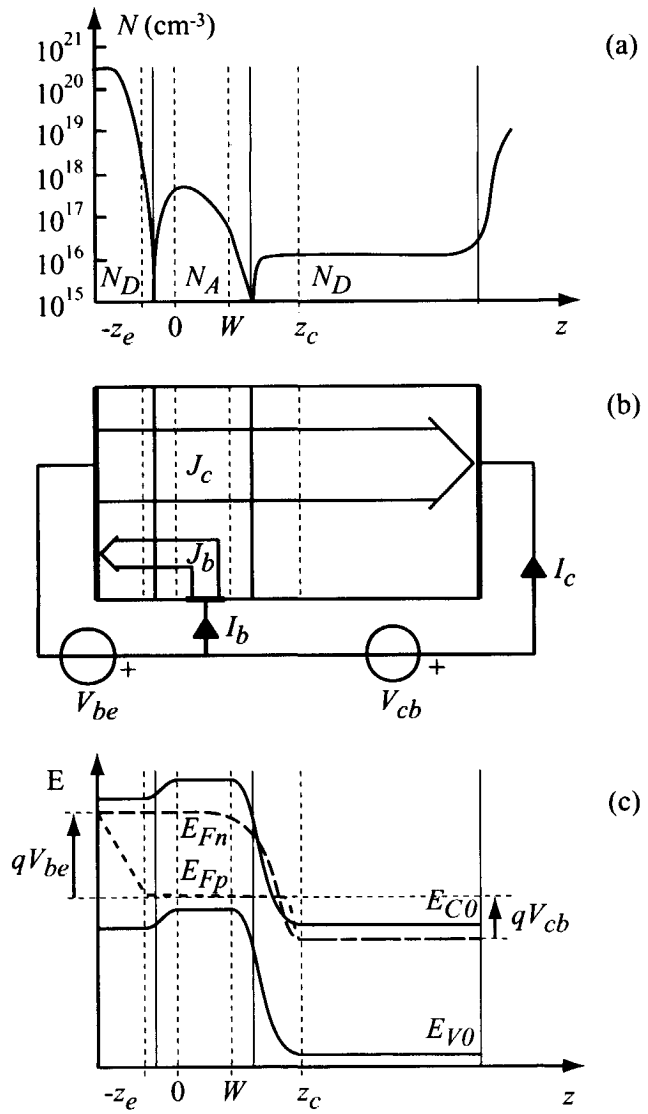

Figure D.3: 1D simplification of a vertical *npn* transistor in forward bias. (a) Profile of effective doping for a line through the emitter, (b) Flow of the main currents, (c) Band edge diagram and quasi-Fermi levels. The dashed vertical lines indicate the boundaries of the depletion layers.

- The base material is nondegenerate, making Maxwell-Boltzmann statistics applicable;
- The bias voltages are low enough to neglect avalanche multiplication;
- The minority concentration in the base is less than the doping concentration (low-level injection);
- The base is wide enough to neglect the effects of velocity saturation (see the comment in the following);
- The recombination in the base and space charge regions is negligible;
- The forward current amplification factor β_F is much higher than one;
- The majority current in the base flowing into the collector is small compared to the minority current.

D.3 Current densities in the base

The last two assumptions in the previous section imply that the collector current is dominated by the minority current injected from the emitter into the base. This current will be derived here for the case of an *npn* transistor.

In the base, at a position $\mathbf{x}$ in the quasi-neutral region, an electron current is flowing in band m with a density $J_i^m(\mathbf{x})$. In its most general form this current density can be written as (see Equation C.36 and C.41):

$$J_i^m(\mathbf{x}) = n^m(\mathbf{x})\,\mu_{ij}^m(\mathbf{x})\,\frac{\partial E_{Fn}}{\partial x_j}(\mathbf{x}) \tag{D.2}$$

where n^m is the contribution of the band to the non-equilibrium electron concentration in the base, and μ_{ij}^m is the mobility tensor of the band. E_{Fn} is the quasi Fermi level for electrons.

For nondegenerate conditions the electron concentration n^m can be replaced by the exponential expression of Equation C.9 and C.11:

$$n^m(\mathbf{x}) = N_C^m(\mathbf{x}) \exp\left(\frac{E_{Fn}(\mathbf{x}) - E_{C0}^m(\mathbf{x})}{k_B T}\right) \tag{D.3}$$

in which N_C^m is the effective density of states of the conduction band m, E_{C0}^m its edge energy, k_B the Boltzmann constant and T the absolute temperature. The quasi Fermi level can be related to the potential in the base by multiplying both sides of Equation D.2 by the total hole concentration p. Since low-level injection is assumed, p is approximately equal to the local doping concentration N_A:

$$p(\mathbf{x}) = N_A(\mathbf{x}) = \sum_u N_V^u(\mathbf{x}) \exp\left(\frac{E_V^u(\mathbf{x}) - E_{Fp}(\mathbf{x})}{k_B T}\right) \tag{D.4}$$

where the sum runs over all u valence bands and where the symbols have an equivalent meaning as in Equation D.3. Inserting this into Equation D.2 gives:

$$J_i^m N_A = \left(n_{i0}^m\right)^2 \mu_{ij}^m \exp\left(\frac{E_{Fn} - E_{Fp}}{k_B T}\right) \frac{dE_{Fn}}{dx_j} \tag{D.5}$$

where $\left(n_{i0}^m\right)^2 = \left(n_{i0}^m\right)^2(\mathbf{x})$ is the pn-product of conduction band m with respect to all valence bands (see Section C.3.1). The subscript "0" indicates the situation of thermodynamic equilibrium.

The Fermi level gradient is applied in the z-direction along a unit vector l_i, which itself is defined in the axes system of the mobility tensor. The resulting current vector has only one interesting component: the one crossing the junction and therefore along l_i. The magnitude of this component J^m is equal to the inner product $J_i^m l_i$. The vectorial equation D.5 can therefore be rewritten in scalar form. With some rearrangements this yields:

$$\frac{J^m N_A}{\left(n_{i0}^m\right)^2 l_i l_j \mu_{ij}^m} = \exp\left(\frac{E_{Fn} - E_{Fp}}{k_B T}\right) \frac{dE_{Fn}}{dz} \tag{D.6}$$

D.4 Relation with the terminal voltages

The above expression for the local current density can be transformed to an equation for the complete device by integrating z over the effective base width W (see Figure D.3). Because of the one-dimensional approximation and the negligible recombination, J^m does not change over the base width and is independent of z. It is therefore equal to the current density J_c^m flowing over the base-collector junction. It can be placed outside the integral, yielding:

$$J_c^m \int_0^W \frac{N_A}{l_i l_j \mu_{ij}^m \left(n_{i0}^m\right)^2} dz = \int_0^W \exp\left(\frac{E_{Fn} - E_{Fp}}{k_B T}\right) \frac{dE_{Fn}}{dz} dz \tag{D.7}$$

Now, the quasi Fermi level of holes E_{Fp} is also assumed to be constant over the base width, at least in comparison with E_{Fn}. This is firstly the result of the assumption of low-level injection. Under such injection the concentration of electrons remains inferior to the concentration of the doping. Secondly, it is the result of the assumption of high, current gain, meaning that the electron current through the base is much higher than the hole current. Finally, it is the result of the hole mobility being nearly equal to the electron mobility. The consideration of Equation D.2 now learns that for keeping all currents flowing, the Fermi gradient of the electrons need to be much larger than that of the holes.

The above equation will hardly change if the space charge regions are included within the integration interval. This is best seen at the left-hand side. Firstly, the depleted regions contribute little to the total number of holes. The reason is that they are mostly thin compared to the effective base width W. In addition, they have a low hole concentration compared to the quasi-neutral

base because of the high internal field. Secondly, the values of μ_{ij}^m and $(n_{i0}^m)^2$ are quite independent of the material and region type. Therefore Equation D.7 can be written in a good approximation as:

$$J_c^m = \left(\int_0^W \frac{N_A}{l_i l_j \mu_{ij}^m \left(n_{i0}^m\right)^2} dz \right)^{-1} \int_{E_{Fn}(z_e)}^{E_{Fn}(z_c)} \exp\left(\frac{E_{Fn} - E_{Fp}^b}{k_B T} \right) dE_{Fn} \quad \text{(D.8)}$$

where E_{Fp}^b is the constant value of E_{Fp}.

The quasi Fermi levels at the space charge boundaries apply to the local majority charge carriers. They can therefore be related to the voltages applied to the transistor. By definition it is true that:

$$V_{be} = \frac{1}{q}\left(E_{Fn}(z_e) - E_{Fp}^b\right), \quad V_{cb} = \frac{1}{q}\left(E_{Fp}^b - E_{Fn}(z_c)\right) \quad \text{(D.9)}$$

where V_{be} and V_{cb} are the voltages over the base-emitter and collector-base junctions, respectively. This means that:

$$\int_{E_{Fn}(z_e)}^{E_{Fn}(z_c)} \exp\left(\frac{E_{Fn} - E_{Fp}^b}{k_B T} \right) dE_{Fn} = -k_B T \left[\exp\left(\frac{qV_{be}}{k_B T} \right) - \exp\left(\frac{-qV_{cb}}{k_B T} \right) \right] \quad \text{(D.10)}$$

The voltages are equal to those between the device terminals if the ohmic voltage drop inside each transistor region is negligible with respect to the junction voltages.

The first term on the right-hand side of Equation D.10 represents the energy of the electrons flowing forward from the emitter into the base, while the second term represents the reverse flow from the collector. This reverse flow is generally undesirable and should be suppressed. It starts being negligible when V_{ce} exceeds 200 mV. This is a quite reasonable condition in the practice of analog circuit design.

As a result, Equation D.8 is simplified to:

$$J_c^m = -J_S^m \exp\left(\frac{qV_{be}}{k_B T} \right) \quad \text{(D.11)}$$

where J_S^m is the saturation current density due to band m:

$$J_S^m = k_B T \left(\int_0^W \frac{N_A(\mathbf{x})}{l_i l_j \mu_{ij}^m(\mathbf{x}) \left(n_{i0}^m(\mathbf{x})\right)^2} dz \right)^{-1} \quad \text{(D.12)}$$

It should be noted that every parameter under the integral sign may be a function of position, meaning that the expression of J_S^m is valid for inhomogeneous and anisotropic base material.

In the above derivation the effects of velocity saturation have been neglected. This saturation may occur in the space charge regions where the electric field

is so high that the electron velocity reaches a saturation value. Inclusion of the effect would result in an additional integral [1, 3, 7]. The effect of this integral may be represented by an increase in the effective base width. This increase amounts 20 nm at maximum. For most transistors this is still negligible.

D.5 Collector and saturation current

In order to find the total collector current density the contributions of all bands should be summed. In addition, the current density should be integrated over the junction area A. Finally, the sign should be reversed since, conventionally, the collector current I_c is considered positive when it enters the terminal. This yields:

$$I_c = I_S(\mathbf{x}) \exp\left(\frac{qV_{be}}{k_B T}\right) \tag{D.13}$$

where I_S is the saturation current, given by:

$$I_S = \sum_m \int_A J_S^m(\mathbf{x})\, dxdy \tag{D.14}$$

Often it is not too far from the truth to assume that μ_{ij}^m and n_{i0}^m in the base are independent of the exact position. If in addition the doping profile is assumed to be uniform, Equation D.14 reduces to:

$$I_S = k_B T \frac{A}{WN_A} \sum_m l_i l_j \mu_{ij}^m (n_{i0}^m)^2 \tag{D.15}$$

The equation can be reduced even further if the material is isotropic, which is the case in stress-free silicon. The mobility is then equal to the scalar average value μ^{n} and the pn-product equal to n_{i0}^2. By applying the Einstein relation of Equation C.52, μ^{n} can also be expressed in an average diffusion constant D^{n}. The saturation current can then be written as:

$$I_S = q \frac{A}{W} \frac{D^{\mathrm{n}} n_{i0}^2}{N_A} \tag{D.16}$$

which is the expression appearing in most textbooks on the subject [5].

D.6 Conclusions

Under some reasonable assumptions, the collector current of a bipolar transistor can be related to the exponent of the base-emitter voltage. This relation contains a linear dependence on the saturation current, which is a transistor parameter. It can be expressed in terms of the geometry, the mobility tensor, and the pn-product in thermodynamic equilibrium. In its evaluation, care should be taken to correctly sum the contributions of the different energy bands and the various tensor components.

Bibliography

[1] H. K. Gummel, A charge control relation for bipolar transistors, *Bell Syst. Techn. J.*, *49* (1970) 115-120.

[2] H. K. Gummel and H. C. Poon, An integral charge control model of bipolar transistors, *Bell Syst. Techn. J.*, *49* (1970) 827-852.

[3] G. A. M. Hurkx, *Bipolar transistor physics*, in: *Bipolar and bipolar-MOS integration*, P. A. H. Hart (ed.), Elsevier, Amsterdam, 1994, pp. 71-175.

[4] H. Kroemer, Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region, *Solid-St. Electr.*, *28* (1985) 1101-1103.

[5] S. M. Sze, *Physics of semiconductor devices*, Wiley, New York, 1981, pp. 133-189.

[6] J. L. Moll and I. M. Ross, The dependence of transistor parameters on the distribution of base layer resistivity, *Proc. IRE*, *44* (1956) 72-78.

[7] G. A. M. Hurkx, An improved analytical description of the effect of velocity saturation on the collector current of a bipolar transistor, *Solid-St. Electr.*, *35* (1992) 1397-1399.

List of symbols

α	Shear deformation constant of the conduction bands
α_h	Sensitivity of I_S to hydrostatic pressure
β	Transistor DC current gain
β_0	Maximum transistor DC current gain
γ_n, γ_p	Impurity capture cross section of electrons, holes
Γ_n^m	Inhomogeneity function of the conduction band m
Γ_p^u	Inhomogeneity function of the valence band u
δ_{ij}	Kronecker delta
δ_v	Vertical deflection of the beam tip
Δ_{SO}	Spin orbit splitting energy
ε_0	Electrical permittivity of vacuum
ε_r	Dielectric constant
ζ_{ijkl}, ζ_{OP}	First-order piezojunction tensor
$\zeta_{ijklmn}, \zeta_{OPQ}$	Second-order piezojunction tensor
ζ_i^I	First-order piezojunction coefficients
$\zeta_i^{I'}$	Linear combination of first-order piezojunction coeff.
$\zeta_i^{I''}$	Linear combination of first-order piezojunction coeff.
ζ_i^{II}	Second-order piezojunction coefficients
$\zeta_i^{II'}$	Linear combination of second-order piezojunction coeff.
$\zeta_i^{II''}$	Linear combination of second-order piezojunction coeff.
η^m	Energy difference between E_{Fn} and E_{C0}^m
θ_1, θ_2	Strain-induced rotation of the orhtogonal sides of a square
Θ	Help variable in valence band equations
κ, κ_i	Wave vector transformed by the Herring transformation
$\mu^{\mathrm{n}}, \mu^{\mathrm{p}}$	Average electron, hole mobility in isotropic material
$\mu_{ij}^{\mathrm{n}}, \mu_{ij}^{\mathrm{p}}$	Average electron, hole mobility
μ_{ij}^m	Mobility of an electron in conduction band m
μ_{ij}^u	Mobility of a hole in valence band u
Ξ_{ij}^{mn}	First-order deformation potential tensor
Ξ_l	Longitudinal deformation potential of the conduction bands
Ξ_t	Transverse deformation potential of the conduction bands
π_{ijkl}, π_{OP}	First-order piezoresistive tensor
π_{ijklmn}, π_{OPQ}	Second-order piezoresistive tensor
π_i	First-order piezoresistive coefficients

Π_{ijkl}, Π_{OP}	First-order piezoconductive tensor
Π_{ijklmn}, Π_{OPQ}	Second-order piezoconductive tensor
ρ^{ch}	Total space charge concentration
$\rho^{\mathrm{n}}_{ij}, \rho^{\mathrm{p}}_{ij}$	Total electron, hole resistivity
ρ^{0}_{ij}	Resistivity in the absence of mechanical stress
$\sigma^{\mathrm{n}}_{ij}, \sigma^{\mathrm{p}}_{ij}$	Total electron, hole conductivity
σ^{0}_{ij}	Conductivity in the absence of mechanical stress
σ^{m}_{ij}	Conductivity due to conduction band m
σ^{u}_{ij}	Conductivity due to valence band u
$\hat{\sigma}, \hat{\sigma}_i$	Vector of Pauli matrices
τ^{m}_{Aug}	Auger lifetime of an electron in band m
τ^{u}_{Aug}	Auger lifetime of a hole in band u
τ^{m}_{SRH}	Shockley-Read-Hall lifetime of an electron in band m
τ^{u}_{SRH}	Shockley-Read-Hall lifetime of a hole in band u
τ_m	Relaxation time of an electron in band m
τ^{0}_{m}	Energy independent component of the relaxation time τ_m
τ_u	Relaxation time of a hole in band u
τ^{0}_{u}	Energy independent component of the relaxation time τ_u
χ^{m}	Electron affinity of conduction band m
$\phi(r)$	Spherically symmetric factor of the electron wave function
ϕ	Electrostatic potential
φ	Angle of rotation with respect to the [110] crystal direction
$\psi_{n\mathbf{k}}$	Wave function of an electron in band n
$\Psi^{+}_{i}, \Psi^{-}_{i}$	Electron wave functions including spin
$a_{ij}, \hat{a}$	Orthogonal transformation matrix or operator
a, b, d	Alternative set of valence band deformation potentials
A	Area of the resistor contact or the emitter-base junction
A_{ij}	Matrix with orientation of the experim. stress and current
b, d, a	Alternative set of valence band deformation potentials
B_{ij}	Matrix with orientation of the experim. stress and current
c_l	Velocity of light
c_n	Fourier coefficients
C_{ij}	Matrix describing the experimental stress and current orientation
C_n	Auger coefficient of electrons
C_p	Auger coefficient of holes
d, a, b	Alternative set of valence band deformation potentials
$D^{\mathrm{n}}, D^{\mathrm{p}}$	Electron, hole diffusion constant in isotropic material
D^{mn}_{ij}	Inverse effective mass tensor of the valence bands
D^{m}_{ij}	Diffusion tensor constant of conduction band m
D^{u}_{ij}	Diffusion tensor constant of valence band u
e_{ij}	Strain tensor
E	Undefined energy, or Young's modulus
E^{m}_{C0}	Energy at the edge of conduction band m
E_{Fn}	Quasi-Fermi level of electrons
E_{Fp}	Quasi-Fermi level of holes

$E_{l,0}$	Energy of an electron at rest, at a potential of zero
E_{V0}^{u}	Energy at the edge of valence band u
E_0	Energy eigenvalue of the unperturbed band edge
E_C^m	Energy of an electron in conduction band m
E_F	Fermi level in thermodynamic equilibrium
E_G^{mu}	Energy gap between conduction band m and valence band u
E_l	Energy of the local vacuum level
E_n	Energy eigenvalue of an electron in band n
E_u'	Reduced energy eigenvalue of band u
E_V^u	Valence band, or energy of a hole in valence band u
f_n	Distribution function of electrons
f_n^0, f_p^0	Electron, hole distribution in thermodynamic equilibrium
f_n^1, f_p^1	Perturbation of the electron, hole distribution function
f_p	Distribution function of holes
F_C	Total available density of states of the conduction bands
F_C^m	Available density of states (DOS) of conduction band m
F_V	Total available density of states of the valence bands
F_V^u	Available density of states (DOS) of valence band u
$\mathbf{F}, F_j$	Gradient in the quasi-Fermi level
$\mathbf{F}^{el}$	Electrostatic field
g_C^m	Density of states function in conduction band m
g_V^u	Density of states function in valence band u
g_m	Transconductance
G_{ij}^m	Part of the conductivity due to band m independent of E_{Fn}
G_{ij}^u	Part of the conductivity due to band u independent of E_{Fp}
G_n^m	Generation rate of electrons in band m
G_p^u	Generation rate of holes in band u
h_{ij}	Help parameters in valence bands or in perturbation theory
$\hbar$	Dirac's constant
$\hat{H}$	Full Hamiltonian operator
$\mathbf{H}$	Full Hamiltonian matrix
$\hat{H}_{\mathbf{k}\hat{\mathbf{p}}}$	$\mathbf{k}\hat{\mathbf{p}}$ perturbation Hamiltonian
$\mathbf{H}_{\mathbf{k}\hat{\mathbf{p}}}$	$\mathbf{k}\hat{\mathbf{p}}$ matrix
$\hat{H}_{\mathbf{k}\hat{\mathbf{p}}\varepsilon}$	Mixed $\mathbf{k}\hat{\mathbf{p}}$ and strain perturbation Hamiltonian
$\hat{H}_0$	Hamiltonian of unperturbed electrons
$\hat{H}_C^m$	Semiclassical Hamiltonian of an electron in band m
$\hat{H}_\varepsilon$	Strain perturbation Hamiltonian
$\hat{H}_\varepsilon$	Strain perturbation Hamiltonian
$\hat{H}_V^u$	Semiclassical Hamiltonian of a hole in band u
H_{mn}	Matrix element of the Hamiltonian operator $\hat{H}$
HH	Heavy hole band index
$\mathbf{I}_2$	Identity matrix of rank 2
I_c	Collector current
I_i	Collector current of orientation i normalised on stress-free value
I_S	Saturation current

$\mathbf{J}^{\mathrm{n}}$, J_i^{n}	Total electron current density
$\mathbf{J}^{\mathrm{p}}$, J_i^{p}	Total hole current density
J_i^m	Electron current density due to band m
J_i^u	Hole current density due to band u
J_R	Current density in the resistor R
k_B	Boltzmann's constant
$\mathbf{k_0}$, k_i^0	Wave vector of an electron at the band extremum
$\mathbf{k}$, k_i	Wave vector of an electron
$\mathbf{k}'$, k_i'	Wave vector in the strained coordinate system
K_i^m	Unit vector in $\mathbf{k}$-space towards the edge of band m
l	Length of cantilever beam
l, m, n	Valence band deformation potentials
l_i	Unit vector in the current direction
L	Length of a resistor
$\hat{\mathbf{L}}$	Angular momentum operator
L, M, N	Valence band effective mass parameters
LH	Light hole band index
m	Index of a conduction band or an arbitrary band
m, n, l	Valence band deformation potentials
m_{dC}^m	Density of states effective mass of conduction band m
m_0	Free electron mass
m_l	Longitudinal effective mass of the conduction bands
m_t	Transverse effective mass of the conduction bands
M_{ij}^m	Inverse effective mass tensor of conduction band m
$\bar{M}_{ij}$	Diagonalised effective mass tensor
M, N, L	Valence band effective mass parameters
n	Total electron concentration
n, l, m	Valence band deformation potentials
n_i	Intrinsic carrier concentration
n_{i0}	Intrinsic carrier concentration at thermodynamic equilibrium
n_{i0}^{mu}	Intrinsic conc. of cond. band m and valence band u
n_{i0}^m	Intrinsic conc. of cond. band m and all valence bands
n_0	Electron concentration at thermodynamic equilibrium
N, L, M	Valence band effective mass parameters
N_A	Acceptor doping concentration
N_A^-	Concentration of ionised acceptor atoms
N_C	Effective density of states (DOS) of all conduction bands
N_C^m	Effective density of states (DOS) of conduction band m
N_D	Donor doping concentration
N_D^+	Concentration of ionised donor atoms
N_T	Concentration of generation-recombination centres
N_V	Effective density of states (DOS) of all valence bands
N_V^u	Effective density of states (DOS) of valence band u
$\hat{\mathbf{p}}$, $\hat{p}_i$	Momentum operator
p	Total hole concentration

p, q, r	Help variables in valence band equations
p_α^{mn}	Matrix element of the momentum vector operator
p_0	Hole concentration at thermodynamic equilibrium
q	Unit charge
q, r, p	Help variables in valence band equations
q_i	Normalised vector in the direction of the uniaxial stress
Q, R	Help variables in valence band equations
r	Radius in spherical coordinate system
r, p, q	Help variables in valence band equations
R	Resistance
R, Q	Help variables in valence band equations
R_i	Resistance of orientation i normalised on stress-free value
R_n^m	Recombination rate of electrons in band m
R_p^u	Recombination rate of holes in band u
$\hat{S}$	Symmetry operator of the crystal
$\hat{S}_{sc}$	Scattering operator
S_{ijkl}	Elastic compliance
SO	Split-off band index
T	Absolute temperature
T_0	Reference temperature at time $t = 0$
T_1	Transistor temperature extracted from Gummel plots
T_2	Transistor temperature measured by Pt-100
u	Index of a valence band
$u_{n\mathbf{k}}$	Modulating part of the electron Bloch function in band n
u_i	Displacement of a point in the strain definition
U	Electron potential energy
U_{ij}	Strain derivative of the electron potential energy
U_{mn}^A	Hamiltonian perturbation matrix of states in Class A
v_i	Velocity
$v_t h$	Thermal velocity of electrons
V	Voltage
V_{AF}	Forward Early voltage
V_{AR}	Reverse Early voltage
V_{be}	Base-emitter voltage
V_{ce}	Collector-emitter voltage
V_{g0}	Reference bandgap voltage
V_R	Voltage over the resistor R
w	Width of cantilever beam
W	Effective base width
W_n^m	Kinetic energy of an electron in band m
W_p^u	Kinetic energy of a hole in band u
x, y, z	Coordinates attached to the sides of a component
$\mathbf{x}$, x_i	Position vector
$\mathbf{x}'$, x_i'	Position vector in the strained coordinate system
x_t	Distance from a transistor to the tip of the cantilever beam

X	Magnitude of uniaxial stress
X_{kl}	Mechanical stress
z_c	Depletion zone boundary in the collector
z_e	Depletion zone boundary in the emitter

Acknowledgements

This thesis has been realised with the help of many people, who I would like to thank here. In the first place I should mention Paddy French. He created the research project, invited me to join his group, and gave guidance and a daily support. In addition, he has been a driving force behind many social activities that make work more pleasant. At the same time I should mention professor Middelhoek, who was my supervisor and supported me with his vision and experience.

I would like to thank the Technology Foundation STW, the applied science foundation of NWO and the technology programme of the Dutch Ministry of Economic Affairs for the funding of this project, DEL.3908. I thank the members of the STW user's committee for their regular feedback: dr. L.J. Korstanje (STW), ir. M. van der Lee (Enraf), H. Cassier (Alcatel Microelectronics), ir. R. de Boer (SMARTEC), and prof.dr.ir. A.H.M. van Roermund (TU Eindhoven).

There are many people who helped me at particular stages of my research. They include Martin Middelhoek, who showed me how to interpret measurement results and provided me with his software. They include Fabiano Fruett and Gerard Meijer, with whom I regularly discussed the advances and the application side of the research. Anton Bakker helped me with the design of some circuits. I should mention prof.dr. W.Th. Wenckebach for his comments on solid-state physics, and ir. P.V. Pistecky for his help on mechanics. I thank Andreas Friedrich of the Swiss Federal Institute of Technology Lausanne (EPFL) with whom I had clarifying discussions on the piezo-tunneling diode. I thank Henk van Zeijl and Leon van den Oever for their assistance with the device simulations. In addition, I thank Jim Knutti and Henry Allen for the practical training at their company, Silicon Microsystems.

I am grateful to the people who provided the technical support of the project. They include the technicians of the Electronic Instrumentation Laboratory: Piet Trimp, Ger de Graaf, Maureen Meekel, Jeroen Bastemeijer, and Jef Mollinger. Thanks to Lina Sarro for the technological advices and the coordination of the processing of my devices. Also many thanks to Ruud Klerks and Wim van der Vlist for the unusual sawing and wire bonding, and to the other members of the DIMES Technology Centre for the processing. I thank Ger Schotte and the other members of the Mechanical Workshop for the realisation and fine-tuning of the bending apparatus. Similarly, I thank Frans Broos and his group

for the design and fabrication of the printed circuit boards. Finally, I must thank Evelyn Sharabi, Inge van Egmond, and Willem van der Sluys for their professional and friendly administrative support.

Many thanks to the people who corrected parts of my thesis in an early stage: Wilko Kindt, Sandra Bellekom, Jeroen van der Meer, Maarten Bakker, Mirjam Nieman, Richard den Dulk, and Pieter Creemer. Their comments have undoubtly improved the quality of the work.

I wish to thank all collegues of the Electronic Instrumentation Laboratory for the pleasant working environment. I really enjoyed their company during the many coffee breaks, lunches, dinners, and trips to conferences. I thank my roommates Davies, Gabriel, and Michael for their company. In particular, I remember the roommates at the start of my project: Sandra, Wilko, Paul, and Colin.

I thank all my friends, especially Jan and Maarten, who where always willing to put things in a wider perspective during a good meal and a glass of beer. I thank Isabelle for all the love and support she gave me.

Finally, I am very grateful to my parents. They have always stimulated me in my studies, and have supported me in every field to make this thesis possible.

About the Author

Fredrik Creemer was born in Amsterdam, the Netherlands, in 1972. From 1984 to 1990 he followed secondary education at the Hermann Wesselink College in Amstelveen and graduated with the 'Gymnasium-β' diploma. In 1990 he joined Delft University of Technology to study Electrical Engineering. He obtained his M.Sc. or 'ingenieurs' degree in 1995 (cum laude). He joined the Université Paris-Sud where he received the 'Diplôme d'Études Approfondis' in electronics in 1996. From 1996 to 2001 he worked towards his Ph.D. degree at Delft University within the Electronic Instrumentation Laboratory of DIMES.

During his studies Fredrik Creemer was a trainee of Océ Copiers in Venlo, the Netherlands. In 1995 he worked for Philips Semiconductors in Nijmegen in the fabrication of CMOS circuits. In Paris he was a trainee at the Institut d'Électronique Fondamentale where he designed an optoelectronic switch. He also worked as a trainee at Exar Corporation, USA, in the production of micromachined silicon pressure sensors. Currently, he is a designer of analogue circuits at SystematIC design, Delft.

List of Publications

Journal articles

- J. F. Creemer, S. Middelhoek, and P. M. Sarro, The photoelectromagnetic effect in planar silicon structures, *Sensors and Actuators A, 55* (1996) 115-120.

- J. F. Creemer and P. J. French, The piezojunction effect in bipolar transistors at moderate stress levels: a theoretical and experimental study, *Sensors and Actuators A, 82* (2000) 181-185.

- J. F. Creemer, F. Fruett, G. C. M. Meijer, and P. J. French, The piezojunction effect in silicon sensors and circuits and its relation to piezoresistance, *IEEE Sensors J., 1* (2001) 98-108.

- J. F. Creemer and P. J. French, A new model of the effect of mechanical stress on the saturation current of bipolar transistors, *Sensors and Actuators A*, To be published.

Conference presentations

- J. F. Creemer and P. J. French, Reduction of uncertainty in the measurement of the piezoresistive coefficients of silicon with a three-element rosette, *Proc. SPIE, Smart structures and materials, Smart Electronics and MEMS*, San Diego, CA, USA, March 2-4, 1998, Vol. 3328, pp. 392-402.

- J. F. Creemer and P. J. French, Theoretical and experimental study of the piezojunction effect in bipolar transistors under moderate stress levels, *Dig. 10th International Conference on Solid-State Sensors and Actuators (Transducers99)*, Sendai, Japan, June 7-10, 1999, Vol. 1, pp. 204-208.

- J. F. Creemer and P. J. French, The piezojunction effect in mechanical and bandgap sensors, *Proc. 2nd annual workshop on Semiconductor Advances for Future Electronics (SAFE99)*, Mierlo, the Netherlands, Nov. 24-25, 1999, pp. 105-110.

- J. F. Creemer and P. J. French, The orientation dependence of the piezojunction effect in bipolar transistors, *Proc. 30th European Solid-State Device Research Conference (ESSDERC 2000)*, Cork, Ireland, Sept. 11-13, 2000, pp. 416-419.

- F. Creemer and P. French, The piezojunction effect in silicon and its applications to sensors and circuits, *Proc. 1st annual workshop on Semiconductor Sensor and Actuator Technology (SeSens 2000)*, Veldhoven, the Netherlands, Dec. 1, 2000, pp. 627-631.

- J. F. Creemer and P. J. French, An analytical model of the piezojunction effect for arbitrary stress and current orientations, *Dig. 11th International Conference on Solid-State Sensors and Actuators (Transducers01)*, Munich, Germany, June 10-14, 2001, Vol. 1, pp. 256-259.

- J. F. Creemer and P. J. French, Anisotropy of the piezojunction effect in silicon transistors, *Techn. Dig. 15th IEEE International Micro Electro Mechanical Systems Conference (MEMS 2002)*, Las Vegas, NV, USA, Jan. 20-24, 2002, To be presented.

Summary

This thesis presents a new model of the piezojunction effect in silicon. The model can be used to design better mechanical sensors or more robust electronic circuits. Although it is derived from fundamental theory, it describes the behaviour of bipolar transistors under mechanical stress with a set of practically manageable equations. The model has been validated with experiments.

Silicon bipolar transistors are very important components of analogue electronic circuits. Their main electrical characteristic is determined by the saturation current, which is influenced by mechanical stress. This influence is called the piezojunction effect. It is closely related to the well-known piezoresistive effect.

Stress often appears in integrated circuits, where it is an unwanted by-product of the processing and packaging. Through the piezojunction effect stress can cause offset, which is especially a problem in bandgap references. On the other hand, stress is used in micromachined sensors to measure mechanical quantities such as pressure or acceleration. Often, to detect stress, resistors called strain gauges are used. However, stress could also be detected by transistors, which have a much a better source resistance and a smaller active area. This would increase the accuracy of the sensors and decrease their size and power consumption.

In the present practice of circuit and sensor design the piezojunction effect is rarely considered. The reason is that an adequate model of the effect is missing. Previous models are only accurate for very high stresses (over 1 GPa), or for specific stress orientations. The new model in this thesis applies to stresses of any orientation and with moderate magnitudes (up to 200 MPa).

The piezojunction effect in bipolar transistors depends on one key quantity: the conductivity of minority charge carriers in the base. The stress sensitivity of this conductivity can be deduced from the strain, that is, a deformation of the semiconductor crystal lattice. This deformation changes the periods of the electron wave functions and often perturbs their symmetry. As a result, the energy bands of the semiconductor are shifted, split, and deformed. This greatly influences the electron conductivity and thereby on the transistor saturation current. The influence on the saturation current was numerically calculated from deformation potentials found in literature.

The numerical calculations agree with the changes observed in measurements. However, they are complex, contain some time-consuming integrations, and are sensitive to parameter variations. Therefore, they are impractical for design purposes. For those reasons we introduced the macroscopic model.

The macroscopic model describes the stress dependence of the minority conductivity with a polynomial tensor series. The coefficients of this series are called the piezojunction coefficients. This description is very similar to the model of the majority conductivity, in which the stress sensitivity is described by the piezo*resistive* coefficients. Both types of coefficients form tensors of material constants. This fact, combined with the crystal symmetry, greatly limits the number of coefficients that need to be determined and gives the expansion a particularly simple shape. The coefficients were calculated numerically in terms of the energy bands and the deformation potentials.

The model has been validated by experiments in which transistors were characterised under different stresses. For this purpose, vertical *npn* and *pnp* transistors were fabricated in wafers with different crystal orientations. The wafers were sawn into long, thin beams. These were bent as cantilever beams by a newly designed apparatus. During this bending the transistors were characterised electrically by measuring the Gummel plots.

After analysis of the measurements the different saturation currents could be plotted as a function of the applied stress. The data points lie by good approximation on parabolic curves, which have different shapes for tensile and compressive stress. The curves depend greatly on the orientations of the stress and the current. In addition, they differ significantly for *npn* and *pnp* transistors.

Despite the anisotropy, all curves can be described by one polynomial tensor series with cubic symmetry. This confirms the validity of the macroscopic model. In addition, the description provides numerical values for the piezojunction coefficients with a good accuracy.

The accuracy of the measurement set-up was mainly limited by the spreading of the current in the base of the transistors and by parasitic torsion of the beams. The validity of the results is supported by coefficients reported by other authors. It is also supported by a characterisation of the well-known piezo*resistive* effect for which the same set-up was used.

This work probably provides the first accurate model of the piezojunction effect at moderate stress levels. The model is based on fundamental physical principles, but it is sufficiently practical for application in the design of sensors or circuits. The model also shows many parallels with the piezoresistive effect.

Further improvements can be expected when the model is provided with Fermi-Dirac statistics and a better description of electron scattering. In addition, it could be improved by more experiments to find the missing piezojunction coefficients and to determine the other coefficients more accurately. However, the model will prove its practical use most clearly in the design of an actual mechanical sensor.

Samenvatting

Dit proefschrift presenteert een nieuw model van het piëzojunctie-effect in silicium. Het model kan worden gebruikt voor het ontwerpen van betere mechanische sensoren en robuustere elektronische schakelingen. Hoewel het model is afgeleid van een fundamentele theorie, beschrijft het het gedrag van bipolaire transistors onder mechanische spanning met enkele praktisch hanteerbare vergelijkingen. Het model is gevalideerd met experimenten.

Bipolaire transistors van silicium zijn zeer belangrijke componenten van analoge elektronische schakelingen. Hun voornaamste elektrische karakteristiek wordt bepaald door de verzadigingsstroom, die beïnvloed wordt door mechanische spanning. Deze invloed wordt het piëzojunctie-effect genoemd. Het is sterk gerelateerd aan het bekende piëzoweerstands-effect.

Mechanische spanning treedt vaak op in geïntegreerde schakelingen, waar ze een ongewild bijproduct is van de fabricage en verpakking. Via het piëzojunctie-effect kan ze offset veroorzaken, wat vooral een probleem is in bandgap-referenties. De spanning kan daarentegen ook benut worden in sensoren gemaakt met behulp van microbewerking om mechanische grootheden te meten zoals druk of versnelling. Gewoonlijk wordt deze spanning gedetecteerd door weerstanden die rekstrookjes worden genoemd. Echter, ze kan ook worden gedetecteerd door transistors, die een veel betere bronweerstand hebben en een kleiner actief oppervlak. Dit zou de nauwkeurigheid van de sensoren vergroten en hun omvang en vermogensdissipatie verkleinen.

In de huidige ontwerpen van schakelingen en sensoren wordt het piëzojunctie-effect zelden gebruikt of onderzocht. De reden hiervoor is een gebrek aan een geschikt model. Vroegere modellen zijn alleen nauwkeurig voor zeer hoge spanningen (meer dan 1 GPa), of voor specifieke oriëntaties van de spanning. Het nieuwe model in dit proefschrift beschrijft spanningen met een willekeurige oriëntatie en een gematigde grootte (tot 200 MPa).

Het piëzojunctie-effect in bipolaire transistors hangt af van één grootheid in het bijzonder: de geleiding van minderheidsladingsdragers in de basis. De spanningsgevoeligheid van deze geleiding kan afgeleid worden uit de rek, ofwel de vervorming van het kristalrooster van de halfgeleider. Deze vervorming verandert de periode van de elektron-golffuncties en verbreekt gewoonlijk hun symmetrie. Hierdoor worden de energiebanden van de halfgeleider verschoven, gesplitst en

vervormd. Dit heeft grote invloed op de geleiding van elektronen en daardoor op de verzadigingsstroom van de transistor. De invloed op de verzadigingsstroom is numeriek berekend vanuit de deformatiepotentialen uit de literatuur.

De numerieke berekeningen komen overeen met de veranderingen die zijn waargenomen in de metingen. Ze zijn echter complex, bevatten integraties die veel tijd kosten en zijn gevoelig voor variaties in de ingangsvariabelen. Ze zijn daarom onpraktisch voor ontwerp-doeleinden. Vanwege deze redenen wordt het macroscopische model geïntroduceerd.

Het macroscopische model beschrijft de spanningsafhankelijkheid van de minderhedengeleiding met een polynomische tensorreeks. De coëfficiënten van deze reeks worden de piëzojunctie-coëfficiënten genoemd. Deze beschrijving vertoont veel gelijkenis met het model van de meerderhedengeleiding, waar de spanningsafhankelijkheid wordt beschreven met de piëzo*weerstands*-coëfficiënten. Beide types coëfficiënten vormen tensoren van materiaalconstanten. Dit feit, gecombineerd met de kristalsymmetrie, reduceert het aantal coëfficiënten dat bepaald hoeft te worden sterk en geeft de reeks een bijzonder simpele vorm. De coëfficiënten zijn numeriek berekend in termen van de energiebanden en de deformatiepotentialen.

Het model is gevalideerd met experimenten waarin transistors zijn gekarakteriseerd onder verschillende mechanische spanningen. Hiervoor zijn verticale *npn*- en *pnp*-transistors gefabriceerd in plakken met verschillende kristaloriëntaties. De plakken zijn in lange, dunne stroken gezaagd. Deze stroken zijn gebogen als uitkragende liggers met behulp van een speciaal ontworpen apparaat. De buiging resulteerde in goed gedefinieerde spanningen in de transistors, die elektrisch werden gekarakteriseerd door de Gummelplots te meten.

Na analyse van de Gummelplots konden de verschillende verzadigingsstromen geplot worden als functie van de mechanische spanning. De datapunten liggen in goede benadering op parabolische curves, die verschillend van vorm zijn voor trek- en duwspanning. De curves zijn erg afhankelijk van de oriëntatie van zowel de spanning als de stroom. Bovendien tonen ze grote verschillen tussen *npn*- en *pnp*-transistors.

Ondanks de anisotropie worden alle plots beschreven door één polynomische tensorreeks met kubische symmetrie. Dit bevestigt de juistheid van de vergelijkingen en de aannames van het macroscopische model. Daarnaast levert de beschrijving numerieke waarden van de piëzojunctie-coëfficiënten op met een behoorlijke nauwkeurigheid.

De nauwkeurigheid van de meetopstelling werd voornamelijk beperkt door de spreiding van de stroom in de basis van de transistors en door parasitaire torsie in de stroken. De geldigheid van de resultaten wordt bevestigd door coëfficiënten die door andere auteurs zijn gerapporteerd. Ze wordt ook ondersteund door de karakterisering met dezelfde opstelling van het bekende piëzo*weerstands*-effect.

Dit werk bevat waarschijnlijk het eerste nauwkeurige model van het piëzojunctie-effect voor gematigde spanningsniveaus. Dit model is gebaseerd op fundamentele fysische principes, maar is praktisch genoeg om te worden gebruikt

voor het ontwerpen van sensoren en schakelingen. Het model laat daarnaast vele parallellen zien met het piezoweerstandseffect.

Indien het model verbeterd moet worden, zou het kunnen worden voorzien van Fermi-Dirac-statistiek en een betere beschrijving van de verstrooiing van elektronen. Daarnaast zou het verbeterd kunnen worden door meer experimenten. Hierdoor kunnen de nu ontbrekende piëzojunctie-coëfficiënten worden gevonden en de andere nauwkeuriger worden bepaald. Echter, het model zal zijn praktische nut het duidelijkst bewijzen wanneer men er een mechanische sensor mee ontwerpt.

The current-voltage characteristics of bipolar transistors are very sensitive to mechanical stress. This sensitivity is caused by the piezojunction effect. The effect can be used for strain gauges in micromachined mechanical sensors. On the other hand, it should be avoided in integrated circuits where stress arises from packaging or processing and leads to offset.

This work describes the piezojunction effect in silicon from theoretical models as well as from experimental results. It introduces a new, analytical model of the effect. The new model is founded on solid-state physics, but can easily be related to measurements and is applicable to the design of sensors or circuits.

The work will be of interest to researchers and engineers in the fields of semiconductor device physics, silicon sensors, and analogue circuits. Especially, it is intended for those interested in the following questions:

- How does stress deform the semiconductor crystal?
- How does stress change the electron energy bands?
- How does stress modify the carrier transport?
- How do the changes in transport and energy bands relate?

Each of those subjects is supported by an elaborate appendix, where the stress-free case is introduced.

Fredrik Creemer studied Electrical Engineering at Delft University of Technology, the Netherlands. He obtained his Ph.D. degree from the DUT based on research on the piezojunction effect. Currently, he is a designer of analogue electronic circuits at SystematIC design in Delft.

ISBN 90-5166-883-X
9 789051 668834

Eburon