TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Quantization effects in semiconductor inversion and accumulation layers

Citation for published version (APA):
Pals, J. A. (1972). *Quantization effects in semiconductor inversion and accumulation layers*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Electrical Engineering]. Technische Hogeschool Eindhoven. https://doi.org/10.6100/IR126747

DOI:
10.6100/IR126747

Document status and date:
Published: 01/01/1972

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

QUANTIZATION EFFECTS IN SEMICONDUCTOR INVERSION AND ACCUMULATION LAYERS

J. A. PALS

QUANTIZATION EFFECTS IN SEMICONDUCTOR INVERSION AND ACCUMULATION LAYERS

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE TECHNISCHE WETENSCHAPPEN AAN DE TECHNISCHE HOGESCHOOL TE EINDHOVEN, OP GEZAG VAN DE RECTOR MAGNIFICUS, PROF. DR. IR. G. VOSSERS, VOOR EEN COMMISSIE AANGEWEZEN DOOR HET COLLEGE VAN DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP VRIJDAG 15 DECEMBER 1972 TE 16.00 UUR

DOOR

JAN ALBERTUS PALS

GEBOREN TE GORINCHEM

DIT PROEFSCHRIFT IS GOEDGEKEURD DOOR DE PROMOTOREN
PROF. DR. F. M. KLAASSEN EN
PROF. DR. F. VAN DER MAESEN

Aan mijn ouders

Aan Nelly

Het onderzoek beschreven in dit proefschrift is verricht in het Natuurkundig Laboratorium van de N.V. Philips' Gloeilampenfabrieken te Eindhoven in de groep onder leiding van dr. ir. P. A. H. Hart en dr. E. Kooi.

De medewerkers van dit laboratorium, die bijdragen geleverd hebben in de voortgang van het onderzoek, betuig ik mijn dank. Met name wil ik noemen dr. M. V. Whelan, met wie ik diverse aspekten van het onderzoek heb besproken, W. J. J. A. van Heck, die mij veel experimenteel werk uit handen heeft genomen en J. G. van Lierop, die steeds op korte termijn de verschillende halfgeleider-elementen vervaardigd heeft.

De direktie van het Natuurkundig Laboratorium betuig ik mijn erkentelijkheid voor de mij geboden mogelijkheid het onderzoek in deze vorm af te ronden.

CONTENTS

1. INTRODUCTION TO THESIS AND SURVEY OF PREVIOUS WORK ON QUANTIZATION IN SURFACE LAYERS 1

1.1. Introduction 1
1.2. Theoretical work on quantized surface layers 2
1.3. Experimental verification of surface quantization 5

2. THEORETICAL INVESTIGATION OF QUANTIZATION IN SEMICONDUCTOR INVERSION AND ACCUMULATION LAYERS 8

2.1. Theory for charge carriers in an inversion layer 8
2.1.1. Potential well at the surface due to the ionized impurities in the depletion layer 8
2.1.2. Continuum model for the motion of charge carriers in an inversion layer 12
2.1.3. Charge carriers in an inversion layer with a quantized motion perpendicular to the surface 13
2.2. Numerical solution of the equations for an inversion layer . . 17
2.2.1. Calculation of free-carrier density and potential with the continuum model 17
2.2.2. Calculation with the quantum model for the motion of charge carriers 19
2.3. Inversion and accumulation layers in the electric quantum limit 25
2.3.1. A general solution in dimensionless variables for inversion layers in the electric quantum limit 25
2.3.2. An analytical solution with variational calculus 27
2.3.3. A solution for accumulation layers for a non-degenerate semiconductor 30
2.3.4. Range of validity of the electric quantum limit at $T = 0$ K 33

3. EXPERIMENTAL VERIFICATION OF QUANTIZATION BY MEASURING GATE-CAPACITANCE VARIATIONS OF AN MOS TRANSISTOR 41

3.1. Calculation of the gate capacitance of an MOS transistor . . . 41
3.2. Description of the measurements 46
3.3. Comparison between measured and calculated variations in the average inversion-layer thickness 50
3.4. Discussion 56

4. EXPERIMENTAL VERIFICATION OF QUANTIZATION BY MEASURING THE ANOMALOUS BEHAVIOUR IN THE GATE–BULK TRANSFER CAPACITANCE OF AN MOS TRANSISTOR 59

4.1. Calculation of the two-port capacitances of an MOS system with an inversion layer 59
4.2. Description of the measurements 65
4.3. Comparison between measurements and calculations 67
4.4. The influence of surface states and oxide charges 71
4.5. Discussion 74

5. CONCLUSIONS AND REMARKS 77

List of symbols 79

References 82

Summary 85
Samenvatting 87

1. INTRODUCTION TO THESIS AND SURVEY OF PREVIOUS WORK ON QUANTIZATION IN SURFACE LAYERS

1.1. Introduction

Application of an electric field normal to the surface of a semiconductor attracts the majority charge carriers in the semiconductor to the surface or repels them from the surface into the bulk of the material, depending on the charge of these carriers and the sign of the field. As a result of this redistribution of charge carriers a space-charge layer is created at the surface, causing a band bending there. In the resulting potential well charge carriers may be bound to the surface. When majority carriers are bound to the surface we speak of an accumulation layer. When the majority carriers are repelled from the surface into the bulk a depletion layer is formed and the band bending may become so large with increasing applied external field that the surface may become inverted. Minority carriers are then bound in the potential well at the surface and we speak of an inversion layer.

It was realized by Schrieffer as long ago as 1957 that the charge carriers bound to the surface in the potential well should in principle have a quantized motion in the direction perpendicular to the surface [1]). With a quantized motion perpendicular to the surface we have a kind of two-dimensional electron or hole gas in the surface layer. The motion parallel to the surface is free while the motion perpendicular to the surface is bound. The total number of carriers in this surface layer can be varied by varying the externally applied electric field.

From a theoretical point of view the inversion or accumulation layers at the surface of a semiconductor are very useful for studying the behaviour of a two-dimensional electron or hole gas with a quantized motion in one direction.

The understanding of the behaviour of charge carriers in surface layers can have significant practical implications. Such layers find application in many modern semiconductor devices e.g. metal–oxide–semiconductor transistors (MOSTs) [2]) and charge-coupled devices [3]). The operation of these devices is based principally on the conduction along the surface of the charge carriers in an inversion layer. For understanding for instance the surface mobility of these carriers the quantization of their motion perpendicular to the surface must be considered.

In this thesis some theoretical results are given which are used to solve the equations describing an inversion layer or accumulation layer with a quantized motion perpendicular to the surface. New experiments which we performed to demonstrate quantization in inversion layers are described. An account is also given of some experiments on accumulation layers.

In the following sections of this chapter previous work, both theoretical and experimental, on quantized surface layers is briefly reviewed.

Chapter 2 is concerned with the theoretical description of the quantization of carriers in the self-consistent potential well of an inversion or accumulation layer. A numerical procedure we used for solving the coupled Schrödinger and Poisson equations is described. We show that at the electric quantum limit, with only the lowest energy level occupied, a general solution of the equations can be given for all semiconductor materials and surface orientations.

In chapter 3 gate-capacitance measurements on silicon MOSTs are given which verify the quantization in an inversion layer at liquid-nitrogen temperature and room temperature. Measurements on accumulation layers at liquid-helium temperature prove quantization in these channels.

In chapter 4 we give a more direct experimental method for demonstrating quantization in an inversion layer. The measurement method is based on the anomalous behaviour of the externally measurable transfer capacitance between gate and bulk contact of the MOST, when the Fermi level in the device passes through the lowest energy level for motion perpendicular to the surface.

Chapter 5, containing some final remarks and conclusions, ends the thesis.

1.2. Theoretical work on quantized surface layers

Since Schrieffer's realization [1]) that the motion perpendicular to the surface in a surface inversion layer must be quantized, a number of authors have mentioned the possible influence of this quantization on the properties of such a surface layer [4-7]).

The first theoretical work carried out was on surface inversion layers. The minority carriers are bound in a potential well at the surface and are separated from the bulk of the semiconductor by a depletion layer. The simplest way to begin the calculations is by approximating the potential at the surface by a linearly graded well [4,6,8-10]). This is only a reasonable approximation for small inversion-layer charge densities, when the potential well is hardly changed by the space-charge density of the carriers in the inversion layer. For a linearly graded potential well the Schrödinger equation for the motion perpendicular to the surface can be solved analytically yielding the known Airy functions [11]) as wave functions. However, for high concentrations of the inversion-layer charge carriers the influence of the space charge of these carriers on the potential well has to be taken into account. A fully self-consistent formulation of the problem has been given by Stern and Howard in the form of coupled Schrödinger and Poisson equations [12]). The Schrödinger equation is then solved for a potential well which is partly determined by the space-charge density of the carriers with a wave function determined by the Schrödinger equation itself. This method is similar to the Hartree approximation for calculation of the wave functions of electrons in the self-consistent potential of an atom [13]). The first solutions to this coupled set of equations were given for the electric quantum

limit, which means that for motion perpendicular to the surface only the lowest energy level is occupied [12,14]. In general, the equations can only be solved numerically, but analytical approximations can be obtained at the electric quantum limit. Later on, this restriction to the electric quantum limit was abandoned and the occupation of higher energy levels was also taken into account [15–17].

We give a numerical method with which we solved the equations. For the electric quantum limit we show that a general solution can be obtained by introducing dimensionless quantities and we also give approximate analytical solutions obtained with variational calculus.

The problem of quantization in an accumulation layer has also been investigated [18–25]. This problem is in general much more difficult than the problem for an inversion layer. In the case of an inversion layer we have only minority carriers bound to the surface and separated from the bulk by a depletion layer. On the other hand an accumulation layer consists of majority carriers in direct contact with the bulk of the semiconductor. In addition to the carriers in the bound energy levels of the surface accumulation layer we also have the free carriers with a continuous energy spectrum. The wave functions of these carriers which are travelling waves in the interior of the semiconductor are disturbed by the self-consistent potential well at the surface and therefore the free carriers also make a contribution to the space charge in the accumulation layer. The complexity of the problem is greatly reduced in the electric quantum limit when, owing to the sufficiently low temperature, there are no free carriers in the interior of the semiconductor. For this situation we give a solution of the accumulation-layer equations.

In all calculations an effective-mass approximation is used for the carriers in the surface layer. This approximation may be questionable because of the small dimension of the inversion or accumulation layer perpendicular to the surface and due to the rapidly changing potential in these layers. The complex band structure of the material also has to be taken into account. This may give rise to difficulties, especially for the valence band of silicon and germanium which has a degenerate maximum. Owing to the interaction between the two branches the constant-energy surfaces become warped surfaces for high energies which cannot be described with an effective-mass approximation [8,26–29].

One of the characteristic properties of a surface layer is that the density of states, defined as the number of available states per unit energy interval and per unit square, has discontinuities at the energy levels for the motion perpendicular to the surface as illustrated in fig. 1.1.

In this respect the surface inversion or accumulation layers of semiconductors behave like thin films of material [28,29]. In these films the motion perpendicular to the surface is quantized when the thickness of the film is of the same order as the wavelength of the carriers in the film. Considerable work has been done

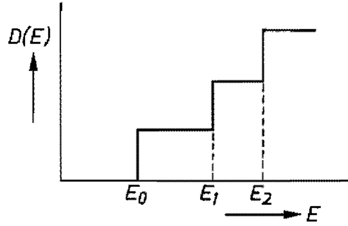


Fig. 1.1. Density of states for a quantized two-dimensional electron or hole gas. E_0, E_1 and E_2 are the quantized energy levels for the motion perpendicular to the surface.

on these size-quantized films [32–38]) and many of the results are similar to those for inversion or accumulation layers.

The first calculations of the mobility of carriers in an inversion layer at the surface of a semiconductor did not take into account the quantization of the motion perpendicular to the surface [39–43]). Later it was realized that quantization plays an important role and has to be taken into account. Much theoretical work has since been devoted to solving the problem of calculating the mobility of carriers moving in a surface potential well, which quantizes the motion perpendicular to the surface. Most investigators consider the two-dimensional analogue of the well-known three-dimensional impurity scattering, for which the scattering centres are charged centres at the semiconductor surface [12,15,21,29,44]). An important effect in this case is the screening of the scattering centres by the two-dimensional charge-carrier gas in the surface layer [12,15,45]). Scattering of the charge carriers by lattice vibrations has also been studied; the higher the temperature the more pronounced this scattering becomes [44,46–48]). A third type of scattering which may be important is that due to surface roughness. This type of scattering has not been investigated to anything like the same extent [49,50]). With these theories the mobility in surface layers can be qualitatively understood, although the experimental behaviour cannot exactly be explained.

A very interesting phenomenon, on which almost all direct experimental demonstrations of quantization are based, occurs if a magnetic field is applied perpendicular to the surface of a semiconductor with an inversion or accumulation layer. The motion of the charge carriers, which is quantized perpendicular to the surface in the potential well, now also becomes quantized parallel to the surface owing to the Landau quantization. The energy spectrum now becomes completely discrete and the density of states shown in fig. 1.1 changes into an array of delta functions separated by the Landau splitting energy $\hbar\omega_c$, as illustrated in fig. 1.2. The cyclotron frequency ω_c is equal to $q\,B/m_d$, where m_d is the effective mass parallel to the surface. This complete quantization of the motion in an inversion or accumulation layer is a characteristic difference compared to three-dimensional Landau splitting in which the motion parallel

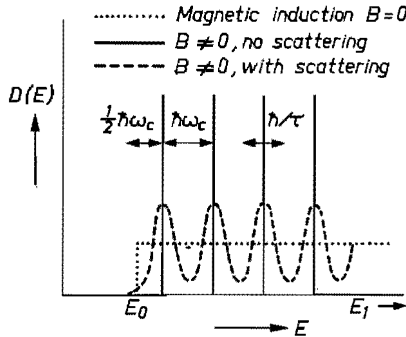


Fig. 1.2. Density of states for a quantized two-dimensional electron or hole gas with a magnetic field perpendicular to the surface causing a Landau level splitting $\hbar\omega_c$.

to the applied magnetic field is still free. In reality the delta-function peaks in the density of states are broadened by the scattering process [10]). If the characteristic collision time for the scattering is equal to τ the levels have a width of the order $\hbar/\tau$. The Landau level splitting is only significant if the splitting between the energy levels is larger than the broadening of the levels and larger than the thermal energy:

$$\hbar\omega_c > \hbar/\tau, kT.$$

To fulfil this condition in practice temperatures below 4·2 K and high magnetic fields B of the order of 1–10 Wb/m^2 are required.

1.3. Experimental verification of surface quantization

After the theoretical demonstration of the importance of quantization in semiconductor inversion and accumulation layers it was some considerable time before experimental verification of the quantization became possible. One of the main difficulties was the preparation of a suitable surface covered with an insulating layer and a metal layer which makes it possible to apply the electric fields normal to the surface necessary for the formation of an inversion or accumulation layer with a variable surface charge density. The invention of the silicon planar process made it possible to make metal–oxide–semiconductor structures on silicon with a high-quality surface, while similar techniques also made these structures possible with other materials.

The first experiments, demonstrating quantization were carried out at temperatures below 4·2 K and with a high magnetic field perpendicular to the surface. The field causes a Landau splitting and the density of states is peaked. With an increasing number of charge carriers in the inversion or accumulation layer the Fermi level shifts through this varying density of states. This causes variations in the conductance parallel to the surface, the so-called Shubnikov–De Haas oscillations. The quantization of the motion perpendicular to

the surface was thus first demonstrated for silicon *n*-type inversion layers [51–55]. Later these experiments were also successful for other materials such as indium antimonide [56,57] and tellurium [58].

The varying density of states owing to the Landau splitting also has an influence on the gate capacitance of the MOS structure. When the Fermi level passes a minimum in the density of states, a minimum in the capacitance occurs, as was shown for silicon *n*-type inversion layers [59,60] and also for the compound $Hg_{0\cdot 8}Cd_{0\cdot 2}Te$ [61].

The typical structure in the density of states due to the simultaneous quantization of the motion perpendicular to the surface by the electric field and of the motion parallel to the surface by the magnetic field is also revealed by tunnel experiments on metal–insulator–semiconductor structures with a quantized accumulation layer. The metal was Pb and the semiconductor indium antimonide [62,63] or indium arsenide [64].

The energy difference between the quantized levels for the motion perpendicular to the surface was recently determined directly with infrared photoconductivity measurements below 4·2 K. If the energy of a photon of the infrared light falling on an inversion layer is equal to the energy difference between two quantized levels, the charge carriers may be excited to a higher energy level. This causes a change in conductivity due to the different mobilities of charge carriers on the various energy levels. By measuring the photoconductance response as a function of the wavelength of the infrared light the energy distance between the quantized levels was determined for *n*-type inversion layers on silicon [65] and indium antimonide [56,66].

Besides these fairly direct experimental checks on quantization, all of which are effected below 4·2 K, there are a number of experimental results that can be understood by taking the quantization into account.

These include a considerable number of mobility measurements for inversion layers. Most of these measurements are done for silicon inversion layers both *n*-type [67–70] and *p*-type [8,27,71,72] and with different surface orientations at different temperatures and with a varying surface charge density. The quantization is taken into account in interpreting these measurements. The agreement between the measurements and theoretical results for the mobility is still only qualitative. The scattering of quantized carriers at the surface of a semiconductor is still not well enough understood for quantitative agreement.

Piezoresistance measurements of silicon *n*-type inversion layers also give an indication of the practical importance of the quantization of the motion perpendicular to the surface [73,74].

Another type of measurements are magnetoresistance measurements of inversion layers with a magnetic field parallel to the surface and perpendicular to the current, which show an anomalous decrease of the resistance with increasing magnetic field [75–77]. This phenomenon, which occurs with a trans-

verse Hall voltage across the inversion layer, is also attributed to the quantization of the motion perpendicular to the surface [78,79]. Magnetoresistance measurements for other directions of the magnetic field also give indications that the charge carriers in the accumulation layer of *n*-type indium antimonide have a quantized motion perpendicular to the surface [80–82].

Tunnel experiments with lead–insulator–indium-arsenide junctions with a magnetic field parallel to the surface give an experimental value of the spread of the wave functions of the quantized carriers which is in good agreement with the theoretical calculations [83].

We might summarize this section by stating that a considerable volume of experimental results is now available that demonstrate more or less directly the importance of quantization of the motion perpendicular to the surface for carriers bound in an inversion or accumulation layer at the surface of a semiconductor. Our measurements give new evidence for quantization in inversion and accumulation layers. One of the remarkable features of our measurements is that quantization is clearly demonstrated at relatively high temperatures. These temperatures are appreciably higher than those reported in previous work.

2. THEORETICAL INVESTIGATION OF QUANTIZATION IN SEMICONDUCTOR INVERSION AND ACCUMULATION LAYERS

2.1. Theory for charge carriers in an inversion layer

To formulate the equations governing the charge carriers in an inversion layer we shall only describe the case for electrons in an inversion or accumulation layer at the surface of a p-type semiconductor and n-type semiconductor respectively. If we want to have the equations for holes in an inversion layer of an n-type semiconductor we only have to change the characteristic quantities of the electrons by the corresponding quantities for the holes. The charge of the carriers has to be changed in this case, the conduction band with its effective masses has to be replaced by the valence band and the effective masses for the holes and the net bulk acceptor concentration has to be replaced by the net donor concentration.

2.1.1. *Potential well at the surface due to the ionized impurities in the depletion layer*

The electrons in an inversion layer at the surface of a semiconductor are bound to the surface in a potential well caused by an externally applied electric field normal to the surface and/or a positive oxide charge in the oxide layer covering the semiconductor. The resulting band bending is schematically sketched in fig. 2.1. The energy at the conduction band edge E_c and the energy at the valence band edge E_v are given as a function of z, the distance measured from the surface of the semiconductor and normal to the surface. We suppose that all quantities of the inversion layer are independent of the coordinates parallel to the surface. E_c is separated from E_v by the band gap E_g. The potential $V(z)$ in the semiconductor is chosen to be the potential at the edge of the conduction band:

$$E_c(z) = -q\ V(z). \tag{2.1}$$

For a given semiconductor device at a certain temperature there are two independent quantities which can be varied by changing the external conditions of the device. The electric-field strength at the surface can be arbitrarily varied and a bias voltage can be applied between the inversion layer at the surface and the bulk of the semiconductor. The way in which this can be achieved practically will become clear in chapter 3 when we describe the MOS structure we used for our devices.

When a bias voltage V_b is applied between the inversion layer and the bulk of the semiconductor, we are not in thermal equilibrium conditions and we use quasi-Fermi levels E_{Fn} and E_{Fp} for electrons and holes. The difference between E_{Fn} in the inversion layer, where the electrons are the majority carriers,

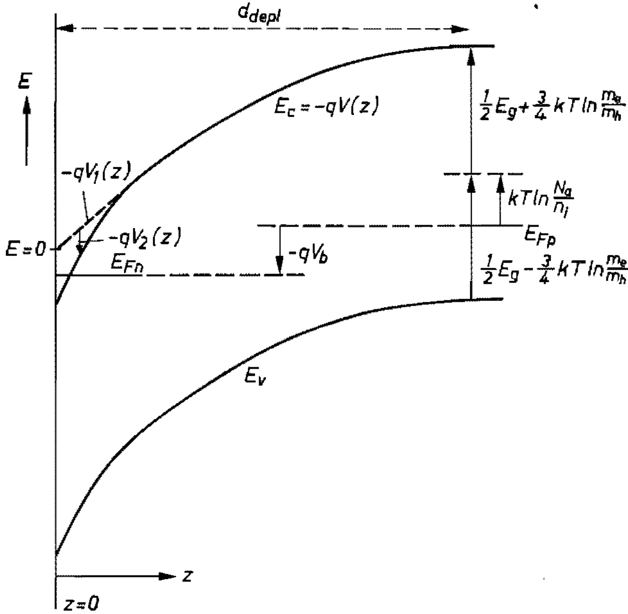


Fig. 2.1. Schematic drawing of the band bending of a *p*-type semiconductor with an *n*-type inversion layer.

and E_{Fp} in the bulk, where the holes are the majority carriers, is determined by V_b:

$$E_{Fp} - E_{Fn} = -q\, V_b. \tag{2.2}$$

We have supposed that the applied voltage V_b is sufficiently low to cause only a small current density. In that case E_{Fn} is constant in the inversion-layer region and E_{Fp} is constant in the bulk of the semiconductor. When $V_b = 0$ the semiconductor is in thermal equilibrium and E_{Fn} and E_{Fp} are equal to the constant Fermi level E_F.

Instead of the electric-field strength at the surface, which is directly determined by the external conditions, we may choose another independent variable which has a one-to-one relation to the electric-field strength. We choose as independent variable the band bending due to the charge in the depletion layer. This is easier to handle for the calculations than the electric-field strength. After the introduction of eqs (2.3) – (2.10) we are able to define this variable more precisely.

The potential in the semiconductor is determined by Poisson's equation

$$\frac{d^2 V}{dz^2} = -\frac{\varrho}{\varepsilon_s}, \tag{2.3}$$

where ε_s is the permittivity of the semiconductor material. The space-charge density ϱ consists of contributions of the hole concentration $q\,p(z)$, the electron concentration $-q\,n(z)$ and the net concentration of ionized acceptor impurities $-q\,N_a$. We suppose the acceptor concentration of the p-type bulk material to be independent of z.

In the bulk of the semiconductor ($z \rightarrow \infty$) the space-charge density is zero and the distance W_b from the conduction-band edge to the quasi-Fermi level for majority carriers (holes) E_{Fp} is determined by the semiconductor material under investigation and the temperature T. For a case in which we may apply Boltzmann statistics and in which all impurities are ionized we have in the bulk of the semiconductor:

$$W_b = E_c - E_{Fp} = \tfrac{1}{2}\,E_g + kT \ln \frac{N_a}{n_i} + \tfrac{3}{4}\,kT \ln \frac{m_e}{m_h}, \qquad (2.4)$$

where m_e and m_h are the effective density-of-states masses of electrons and holes and where n_i is the intrinsic carrier concentration at the temperature T. When the above conditions for eq. (2.4) to be valid are not fulfilled, W_b is nevertheless completely determined by the semiconductor material, the dopant concentration N_a and the temperature T.

To calculate the band bending at the surface we make the so-called depletion-layer approximation. At a certain value of $z = d_{\mathrm{depl}}$ there is an abrupt change from the depletion layer, where the hole concentration $p(z)$ is ignored, to the space-charge-neutral bulk where eq. (2.4) is valid. This approximation is allowed as we are mainly interested in the inversion-layer properties at small values of z, where the depletion-layer approximation for the potential is very good [84].

For reasons of convenience we divide the potential $V(z)$ into two parts:

$$V(z) = V_1(z) + V_2(z); \qquad (2.5)$$

$V_1(z)$ is only determined by the space-charge density due to the ionized acceptors in the depletion layer and $V_2(z)$ is determined by the electron concentration in the inversion layer:

$$\frac{\mathrm{d}^2 V_1}{\mathrm{d}z^2} = \frac{q}{\varepsilon_s}\,N_a, \qquad 0 < z < d_{\mathrm{depl}}, \qquad (2.6)$$

$$\frac{\mathrm{d}^2 V_2}{\mathrm{d}z^2} = \frac{q}{\varepsilon_s}\,n(z). \qquad (2.7)$$

The electron concentration has only to be taken into account in the inversion

layer at very small values of z; outside the inversion layer the minority-carrier concentration $n(z)$ can be neglected: $n(z) \ll N_a$. Therefore $V_2(z)$ will only give a contribution to the total potential in the inversion-layer region. We may therefore take as boundary conditions for V_2 at the edge of the depletion layer:

$$z = d_{\text{depl}}: \qquad V_2 = 0, \quad \frac{\mathrm{d}V_2}{\mathrm{d}z} = 0. \tag{2.8}$$

At the edge of the depletion layer at $z = d_{\text{depl}}$ the potential $V(z)$ and the electric field have to be continuous with the bulk values. This gives the following boundary conditions with the help of eq. (2.4):

$$z = d_{\text{depl}}: \qquad \begin{cases} -q\,V_1 = E_{\mathrm{F}p} + W_b, \\ \dfrac{\mathrm{d}V_1}{\mathrm{d}z} = 0. \end{cases} \tag{2.9}$$

We arbitrarily choose the zero of the potential level by the condition

$$V_1(0) = 0. \tag{2.10}$$

We are now able to formulate more precisely which independent variable we have chosen instead of the electric field at the surface. The total band bending due to the depletion-layer charge is the difference between the value of V_1 at $z = d_{\text{depl}}$ and $z = 0$. This difference is completely determined if we ascribe a value to $E_{\mathrm{F}n}$ at the surface with respect to the chosen zero level of potential as we can see by combining eqs (2.2) and (2.9). We may therefore take the values of V_b and $E_{\mathrm{F}n}$ as the independent variables of the inversion-layer problem. With the conditions (2.2), (2.9) and (2.10) eq. (2.6) can be integrated giving the following results:

$$V_1(z) = -\frac{q\,N_a}{\varepsilon_s}\,(d_{\text{depl}}\,z - \tfrac{1}{2}\,z^2), \tag{2.11}$$

$$d_{\text{depl}} = \left[\frac{2\,\varepsilon_s}{q\,N_a}\left(\frac{E_{\mathrm{F}n}}{q} + \frac{W_b}{q} - V_b\right)\right]^{1/2}. \tag{2.12}$$

To calculate the properties of the inversion layer, which is found to be very thin compared with the depletion-layer thickness for most practical cases, it is possible to approximate $V_1(z)$ by the linear term, giving

$$V_1(z) = -\frac{q\,N_a}{\varepsilon_s}\,d_{\text{depl}}\,z = -F\,z; \tag{2.13}$$

F is the electric field at the surface due to the total ionized impurity charge in the depletion layer.

To determine $V_2(z)$ out of eq. (2.7) we must have an expression for the electron-charge density $n(z)$ in the inversion layer. This expression will be derived in the following sections.

2.1.2. *Continuum model for the motion of charge carriers in an inversion layer*

To investigate the importance of the quantization of the motion in the z direction it is important to compare the results of a calculation with a quantized motion perpendicular to the surface with the results of a conventional calculation where the quantization is neglected [85]. We will first give the equations for an inversion layer with a conventional continuum motion perpendicular to the surface. The potential $V_1(z)$ is already known and given in eqs (2.11) and (2.12); $V_2(z)$ is determined by Poisson's equation (2.7) with boundary conditions (2.8). We only have to find an expression for $n(z)$.

Neglecting the quantization, the electron concentration $n(z)$ is a function of the difference between the edge of the conduction band $-q\,V(z)$ and the quasi-Fermi level for electrons E_{Fn}. In the effective-mass approximation the valley j of the conduction band has a density of states per unit of energy interval and per unit of volume, given by

$$D^j(E) = \frac{1}{2\pi^2}\left(\frac{2\,m_d^{\,j}}{\hbar^2}\right)^{3/2}(E - E_c)^{1/2}. \tag{2.14}$$

The effective density-of-states mass $m_d^{\,j}$ is given by

$$m_d^{\,j} = (m_{x'x'}\,m_{y'y'}\,m_{z'z'})^{1/3}, \tag{2.15}$$

where x',y',z' are the axes in the main directions of the constant-energy ellipsoids for the valley j. In sec. 2.1.3 the use of the effective-mass approximation and the band structure, which may have a multi-valley minimum, will be discussed. The electron concentration is given with the help of Fermi–Dirac statistics by

$$n(z) = \sum_j \int_{E_c}^{\infty} \frac{D^j(E)}{1 + \exp\,[(E - E_{Fn})/kT]}\,\mathrm{d}E. \tag{2.16}$$

The summation over j is the summation over the different valleys of the band. If $E_c - E_F \gg kT$ we may apply Boltzmann statistics and eq. (2.16) can be integrated, giving

$$n(z) = \sum_j \frac{1}{4\pi^3}\left(\frac{2\pi\,m_d^{\,j}\,kT}{\hbar^2}\right)^{3/2} \exp\frac{E_{Fn} - E_c}{kT}. \tag{2.17}$$

For strong inversion, however, the condition for Boltzmann statistics to be valid is not fulfilled and we have to use eq. (2.16) [40]. To obtain a solution of Poisson's equation with Fermi–Dirac statistics we have to use numerical methods, which will be described in sec. 2.2.1.

2.1.3. *Charge carriers in an inversion layer with a quantized motion perpendicular to the surface*

The electrons in an inversion layer are bound to the surface in a potential well $V(z)$. Schrieffer had already realized in 1957 that the motion of the electrons perpendicular to the surface should therefore be quantized [1]. To estimate the importance of this quantization we may make the following observations. For electrons with effective mass m in the order of the free-electron mass and with thermal energy of room temperature the wavelength is about 100 Å, which is of the same order or even larger than the thickness of an inversion layer calculated in the conventional way of the previous section. The energy separation between the energy levels for the motion in the z direction turns out to be in the order of 10 to 100 meV, as the calculations in sec. 2.2.2 will show. This difference is of the same order as the thermal energy at 300 K which is 26 meV. The quantization will therefore be important for temperatures up to room temperature.

For the description of the electron waves in the inversion layer we shall use the effective-mass approximation, which is allowed if the macroscopic potential $V(z)$ in the crystal has only a slight variation over an elementary cell of the crystal compared with the microscopic potential changes within an elementary cell. To see whether this condition is fulfilled for a silicon inversion layer we remark the following.

The maximum field strength which is possible in the silicon surface layer is about 2.10^8 V/m. It is determined by the breakdown field strength of the oxide layer on the silicon. This field causes a potential variation of about 0·1 V over an elementary cell with lattice constant 5·4 Å. The pseudo-potential within an elementary cell can be expanded in Fourier components. It has been shown [86,87] that three terms in this expansion, with reciprocal lattice vectors [1 1 1], [2 0 0] and [3 1 1] and with form factors equal to −0·21, 0·04 and 0·08 Ry respectively, give a good approximation of the pseudo-potential in an elementary cell. The magnitude of the form factors may vary with 0·01 Ry without changing appreciably the calculated band-gap and effective-mass values. As the Rydberg constant Ry is equal to 13·6 V the potential variations within an elementary cell are larger than the external variation of the potential which can be at most 0·1 V. This external potential variation is of the same order as the magnitude of 0·01 Ry within which the form factors may vary. The application of the effective-mass approximation with the bulk values of the

effective-mass tensor may therefore only be questionable for high values of the electric-field strength at the surface.

In the effective-mass approximation the constant-energy surfaces in k space are ellipsoids. The minimum of the conduction band or valence band may be degenerate and there may be several valleys in the band structure each with its effective-mass tensor. We shall treat the different valleys to be independent of each other, which is again an approximation. When we have a degenerate minimum of the band at a certain k value, the degeneracy may be removed by the potential $V(z)$ in the same way that a uniaxial mechanical stress removes the degeneracy and the two valleys interact with each other, giving different constant-energy surfaces in k space compared to the case where the degeneracy is not removed [88].

For silicon we have the following band structure. The conduction band has six equivalent minima for values of the wave vector k in the [1 0 0], [—1 0 0], [0 1 0], [0 —1 0], [0 0 1] and [0 0 —1] directions at a distance $0{\cdot}85.2\pi/a$ from the origin, where a is the lattice constant. The constant-energy surfaces are ellipsoids of revolution with symmetry axes along the corresponding k direction of the minimum and with a longitudinal effective mass m_l in the corresponding k direction, a transversal effective mass m_t perpendicular to that direction. For silicon, the values of m_l and m_t are respectively $m_l = 0{\cdot}98\ m$, $m_t = 0{\cdot}19\ m$, with m the free-electron mass [89]. The valence band of silicon has a two-fold degenerate minimum at the origin of k space. For small values of energy the constant-energy surfaces are spheres with an effective mass $m_h = 0{\cdot}5\ m$ for the heavy holes and an effective mass $m_l = 0{\cdot}16\ m$ for the light holes [90]. We shall label the different valleys with an index j. For the silicon conduction band j runs through values 1 to 6 and for the valence band j only has values 1 and 2.

In the effective-mass approximation the wave function $\Psi^j(x,y,z)$ for electrons in valley j is determined by the Schrödinger equation [12,15]:

$$\left[-\sum_{k,l=x,y,z} \frac{\hbar^2}{2\,m_{kl}{}^j}\frac{\partial^2}{\partial k\,\partial l} - qV(z)\right]\Psi^j = E^j\,\Psi^j, \tag{2.18}$$

where $1/m_{kl}{}^j$ is the reciprocal effective-mass tensor for valley j in the x,y,z coordinates whose z direction is determined by the surface orientation. As the potential $V(z)$ is only a function of z this equation can be separated. We put

$$\Psi^j(x,y,z) = \exp\,[i\,(k_x\,x + k_y\,y)]\,\exp\left[-i\,z\left(\frac{k_x}{m_{xz}} + \frac{k_y}{m_{yz}}\right)m_{zz}\right]\psi^j(z). \tag{2.19}$$

The extra phase term with z dependence is introduced to obtain an equation

for $\psi^j(z)$ which does not have a first derivative with respect to z [12]). By putting eq. (2.19) into eq. (2.18) we get a reduced Schrödinger equation for $\psi^j{}_i(z)$:

$$-\frac{\hbar^2}{2\,m_{zz}}\frac{\mathrm{d}^2}{\mathrm{d}z^2}\,\psi^j{}_i(z)-q\,V(z)\,\psi^j{}_i(z)=E_z{}^j{}_i\,\psi^j{}_i(z); \tag{2.20}$$

$E_z{}^j{}_i$ is the ith energy level for the motion perpendicular to the surface for electrons in valley j. The corresponding wave function $\psi^j{}_i(z)$ is supposed to be normalized:

$$\langle\psi^j{}_i\mid\psi^j{}_i\rangle=\int_0^\infty|\psi^j{}_i(z)|^2\,\mathrm{d}z=1. \tag{2.21}$$

As boundary conditions for eq. (2.20) we take

$$\psi^j{}_i(z)=0 \quad \text{for} \quad z=0 \quad \text{and} \quad z\to\infty. \tag{2.22}$$

The condition of a vanishing wave function at the surface is not completely correct; it requires an infinitely high potential step at the surface. The approximation will nevertheless be good as the potential step at the surface is of the order of several electronvolts and is about a hundred times larger than the energy-level separation of the quantized motion perpendicular to the surface. The total energy of an electron in valley j at the ith energy level for the motion perpendicular to the surface and with wave vector (k_x, k_y) parallel to the surface is

$$E^j{}_i(k_x,k_y)=E_z{}^j{}_i+\frac{\hbar^2}{2}\left[\left(\frac{1}{m_{xx}}-\frac{m_{zz}}{m_{xz}{}^2}\right)k_x{}^2+\right.$$
$$\left.+2\left(\frac{1}{m_{xy}}-\frac{m_{zz}}{m_{xz}\,m_{yz}}\right)k_x\,k_y+\left(\frac{1}{m_{yy}}-\frac{m_{zz}}{m_{yz}{}^2}\right)k_y{}^2\right]. \tag{2.23}$$

All possible electron states out of valley j at the ith energy level for the motion in the z direction form a two-dimensional parabolic subband, the energy parallel to the surface being a quadratic function of k_x and k_y. This subband is schematically given in fig. 2.2. For these subbands the two-dimensional density of states, which means the number of states per unit energy interval and per square unit of the surface, is derived from eq. (2.23). The number of states $D^j{}_i(E)\,\mathrm{d}E$ for subband j between E and $E+\mathrm{d}E$ is equal to the area of the projection of the part of the energy surface between E and $E+\mathrm{d}E$ on the k_x, k_y plane multiplied by the density of states $1/4\pi^2$ in the k_x, k_y plane. By including a factor 2 for spin degeneracy we find:

$$D^j{}_i(E)=\frac{m_d{}^j}{\pi\,\hbar^2}\,H(E-E_z{}^j{}_i), \tag{2.24}$$

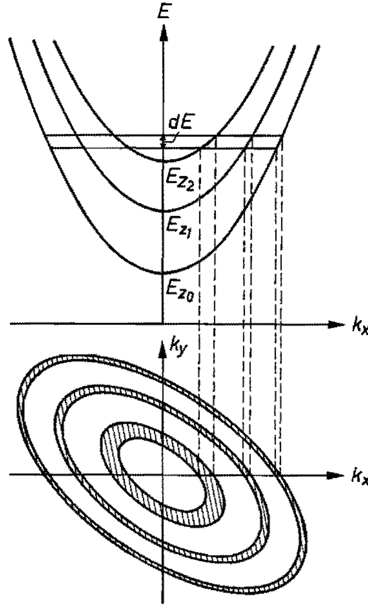


Fig. 2.2. Sketch of the two-dimensional electric subbands for carriers in an inversion layer with a quantized motion perpendicular to the surface. The upper part gives the intersection of the energy surface with the E,k_x plane, the lower part is the projection on the k_x,k_y plane of those parts of the energy surface which have an energy between E and $E + \mathrm{d}E$.

where $H(x)$ is the unit-step function

$$H(x) = 0 \quad \text{for} \quad x < 0,$$
$$H(x) = 1 \quad \text{for} \quad x \geqslant 0,$$

and where the density-of-states mass m_d for the motion parallel to the surface is given by (the index j is omitted)

$$\frac{1}{m_d} = \left[\left(\frac{1}{m_{xx}} - \frac{m_{zz}}{m_{xz}^2}\right)\left(\frac{1}{m_{yy}} - \frac{m_{zz}}{m_{yz}^2}\right) - \left(\frac{1}{m_{xy}} - \frac{m_{zz}}{m_{xz}\, m_{yz}}\right)^2\right]^{1/2}. \tag{2.25}$$

For a given quasi-Fermi level $E_{\mathrm{F}n}$ we are now able to calculate the total number of electrons per unit square $N^j{}_i$ in valley j with energy $E_z{}^j{}_i$ in the z direction. With the use of Fermi–Dirac statistics we find:

$$N^j{}_i = \int\limits_{E_z{}^j{}_i}^{\infty} \frac{D^j{}_i(E)}{1 + \exp\,[(E - E_{\mathrm{F}n})/kT]}\,\mathrm{d}E = \frac{kT}{\pi\,\hbar^2}\, m_d{}^j \ln\left(1 + \exp\frac{E_{\mathrm{F}n} - E_z{}^j{}_i}{kT}\right). \tag{2.26}$$

In the case $E_{Fn} - E_z{}^j{}_i \ll -kT$ we see that $N^j{}_i$ depends exponentially on the difference $E_{Fn} - E_z{}^j{}_i$, which is the Boltzmann approximation. If $E_{Fn} - E_z{}^j{}_i \gg kT$ we see that $N^j{}_i$ is linearly dependent on $E_{Fn} - E_z{}^j{}_i$. The probability of finding these electrons at a distance z from the surface is proportional to the squared modulus of the normalized wave function $|\psi^j{}_i(z)|^2$. The total electron concentration per unit volume in the inversion layer is then given by

$$n(z) = \sum_{i,j} N^j{}_i \, |\psi^j{}_i(z)|^2. \tag{2.27}$$

To describe the inversion-layer properties we want to find a self-consistent solution of the equations (2.20), (2.21) and (2.22). The part $V_1(z)$ of the total potential in the Schrödinger equation is given by eqs (2.11) and (2.12) and $V_2(z)$ is determined by the Poisson equation (2.7) with an electron concentration $n(z)$, which is now no longer an explicit function of $V(z)$ as in the conventional case (eq. (2.16)). The electron concentration is determined by the solution of the Schrödinger equation and by the values of temperature T and quasi-Fermi level E_{Fn} via eqs (2.26) and (2.27). A self-consistent solution of this set of equations can only be obtained with a numerical calculation, which will be described in sec. 2.2.2.

2.2. Numerical solution of the equations for an inversion layer

2.2.1. *Calculation of free-carrier density and potential with the continuum model*

The calculation of an inversion layer neglecting the quantization is straightforward. After choosing a value for the independent parameter E_{Fn} the potential $V_1(z)$ is calculated with eqs (2.11) and (2.12) in which the parameters ε_s, N_a, E_g, V_b, T and n_i are given for the special case we want to calculate. For numerical reasons we choose a quantity z_{max}, so that for $z > z_{max}$ we can neglect the electron concentration. With $n(z) = 0$ for $z_{max} < z < d_{depl}$ we immediately see with the aid of eqs (2.7) and (2.8) that the boundary conditions for V_2 at $z = d_{depl}$ are also valid at $z = z_{max}$. If we choose a suitable value for z_{max} the influence of this approximation is completely negligible. At the end of the calculation we can check whether a correct value of z_{max} has been chosen by comparing the calculated electron concentration at $z = z_{max}$ with the acceptor concentration N_a and the electron concentration at the surface to see if the condition $n(z_{max}) \ll N_a$, $n(0)$ is fulfilled.

If we put $\mathrm{d}V_2/\mathrm{d}z = -F_2$ we have to solve the following set of equations obtained from eq. (2.7):

$$\frac{dV_2}{dz} = -F_2(z),$$
$$\frac{dF_2}{dz} = -\frac{q}{\varepsilon_s} n(z), \tag{2.28}$$

with boundary conditions

$$F_2 = 0 \quad \text{and} \quad V_2 = 0 \quad \text{at} \quad z = z_{\max}; \tag{2.29}$$

$n(z)$ is a function of $V_2(z)$ given by the integral expression (2.16). This integral can be solved by standard numerical integration procedures for each value of V_2 and $n(z)$ can be treated numerically as a known function of V_2. The equations (2.28) and (2.29) can then be solved by a step-by-step Runge–Kutta integration procedure from the initial values of V_2 and F_2 at $z = z_{\max}$ to $z = 0$, giving, at each point, the values of $V_2(z)$ and $n(z)$ [91]. Two results of the calculation for silicon at room temperature are given in figs 2.3 and 2.4. Figure 2.3 gives the electron concentration and potential for an n-type inversion layer with a total number of electrons of $1{\cdot}25.10^{16}\ \text{m}^{-2}$ on a p-type substrate with a dope $N_a = 1{\cdot}5.10^{22}\ \text{m}^{-3}$. Figure 2.4 is for a p-type inversion layer with hole density $5{\cdot}8.10^{15}\ \text{m}^{-2}$ and bulk dope $N_d = 1{\cdot}2.10^{22}\ \text{m}^{-3}$. Both cases are for a {100}-oriented surface. The applied bulk bias V_b is equal to zero.

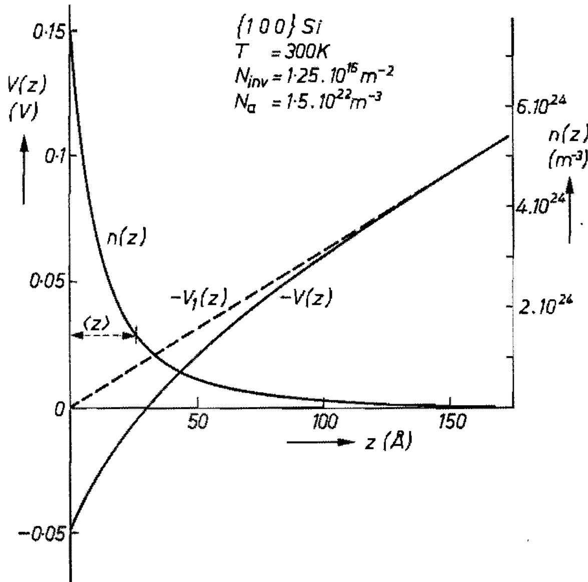


Fig. 2.3. Electron concentration and potential for an n-type inversion layer obtained with a conventional calculation.

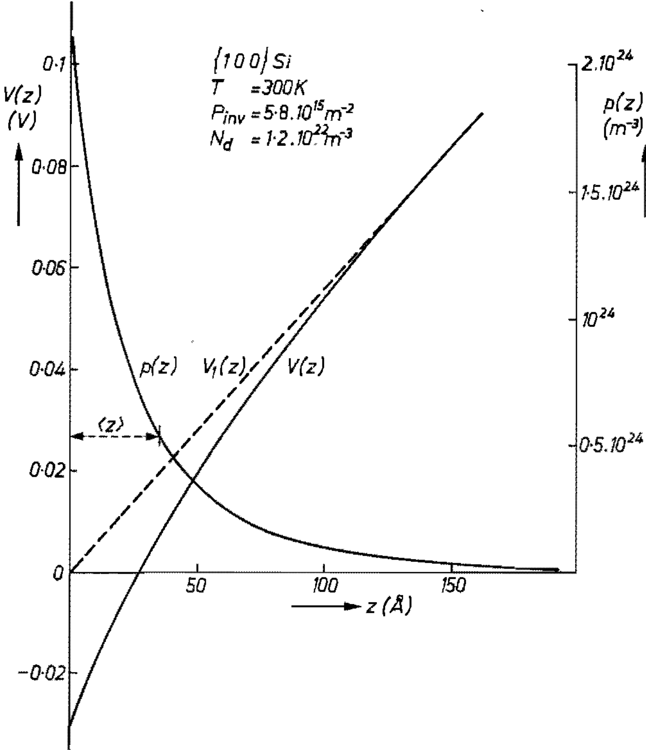


Fig. 2.4. Hole concentration and potential for a p-type inversion layer obtained with a conventional calculation.

2.2.2. *Calculation with the quantum model for the motion of charge carriers*

We want to find a numerical self-consistent solution of the equations (2.2)–(2.10), (2.20)–(2.22) and (2.25)–(2.27). The values of the effective-mass tensor for the different valleys are given by the semiconductor material and the orientation of the surface. The concentration of impurities in the bulk of the semiconductor, the applied bulk bias V_b and the temperature T are known. The only independent parameter is the position of the quasi-Fermi level E_{Fn} at the surface. For a chosen value of E_{Fn}, $V_1(z)$ can be calculated with eqs (2.11) and (2.12). We now start by solving the Schrödinger equation with $V_2(z)$ equal to zero. The Schrödinger equation (2.20) has to be solved in the interval $0 < z < \infty$. For a numerical calculation we solve the differential equation in the interval $0 < z < z_{max}$, with boundary condition instead of eq. (2.22):

$$\psi^j{}_i(z) = 0 \quad \text{for} \quad z = 0 \quad \text{and} \quad z = z_{max}. \tag{2.30}$$

The value of z_{max} has to be carefully chosen; it has to be so large that the boundary condition (2.30) is practically the same as (2.22) and, for numerical

reasons, it has to be as small as possible. When the calculations are finished we have to check, by inspecting the resulting wave functions, whether z_{max} has been given a good value. The criterion for z_{max} to be large enough, is that the wave functions of the energy levels, for which the occupancy cannot be neglected, are already zero within the numerical accuracy for values of z smaller than z_{max}.

We now divide the interval $0 < z < z_{max}$ into N equal subintervals of length h. The value of the wave function $\psi^j{}_i(z)$ at the point $z = nh$ is indicated as $\psi^j{}_{in}$. The Schrödinger equation is now replaced by the following set of $N-1$ linear difference equations for the $N-1$ unknown values of $\psi^j{}_{in}$:

$$\frac{-\hbar^2}{2\,m^j{}_{zz}} \frac{\psi^j{}_{i(n+1)} - 2\,\psi^j{}_{in} + \psi^j{}_{i(n-1)}}{h^2} - q\,V_n\,\psi^j{}_{in} = E_z{}^j{}_i\,\psi^j{}_{in},$$

$$(n = 1, 2, \ldots, N-1), \tag{2.31}$$

$$\psi^j{}_{i0} = \psi^j{}_{iN} = 0.$$

We now want to find the eigenvalues $E_z{}^j{}_i$ with corresponding eigenvectors $\psi^j{}_{in}$ of the tridiagonal $N-1$-by-$N-1$ matrix of the set of linear equations (2.31).

The set of equations (2.31) has a finite number of $N-1$ eigenvalues for the energy levels $E_z{}^j{}_i$. The Schrödinger differential equation has an infinite number of eigenvalues. The difference, of course, is due to the discretization of the Schrödinger equation. For the difference equation it is impossible to describe wave functions whose nodes are at a smaller distance than the step size h. The difference equation only gives a good approximation to the differential equation for the lowest eigenvalues where a sufficiently large number of interval points lie between the nodes of the wave function.

We have solved the set of equations (2.31) by a numerical method known as QR transformation method [92]). For each valley j of the conduction band such an equation has to be solved. We then know the eigenvalues for the energy $E_z{}^j{}_i$ with corresponding wave functions. With eq. (2.26) we then calculate the total number of electrons on each energy level $N^j{}_i$ and with eq. (2.27) we calculate the resulting electron concentration $n(z)$ in the $N-1$ discrete points of the interval. Numerically it is not possible to calculate the infinite number of energy levels $E_z{}^j{}_i$. We have only calculated the ten lowest energy levels, which turns out to be sufficient for most practical cases since the number of electrons at higher energy levels can then be neglected. This is illustrated by the calculated examples given in table 2-I. With the known electron concentration we now solve Poisson's equation (2.7) for $V_2(z)$. As we have taken the wave functions to be zero at $z \geqslant z_{max}$, the electron concentration $n(z) = 0$ for $z \geqslant z_{max}$ and the boundary conditions for V_2 become instead of eq. (2.8):

TABLE 2-I

Some data of the examples of an inversion layer calculated with the quantum model

	n-channel		*p*-channel	
surface orientation	{100}		{100}	
temperature (K)	300		300	
dope of the bulk material (m^{-3})	$1{\cdot}5.10^{22}$		$1{\cdot}2.10^{22}$	
position of Fermi level (eV)	—0·07		—0·06	
depletion-layer charge (m^{-2})	$4{\cdot}1.10^{15}$		$3{\cdot}7.10^{15}$	
inversion-layer charge (m^{-2})	$1{\cdot}25.10^{16}$		$5{\cdot}8.10^{15}$	
number of valleys with different eff. masses	2		2	
	valley 1	valley 2	valley 1	valley 2
number of equivalent valleys n_v	2	4	1	1
effective mass perpendicular to surface m_{zz}	0·98 *m*	0·19 *m*	0·5 *m*	0·16 *m*
density of states eff. mass parallel to surface m_d	0·19 *m*	0·43 *m*	0·5 *m*	0·16 *m*
lowest energy levels (eV)	—0·11 0·0245 0·0476 0·0654 0·0805 0·0939 0·106	0·0236 0·0689 0·100	0·0121 0·0431 0·0651 0·0834 0·0995 0·114 0·128	0·0324 0·0726 0·102
number of carriers per unit square on the corresponding energy level (m^{-2})	$4{\cdot}20.10^{15}$ $1{\cdot}10.10^{15}$ $4{\cdot}54.10^{14}$ $2{\cdot}29.10^{14}$ $1{\cdot}28.10^{14}$ $7{\cdot}59.10^{13}$ $4{\cdot}72.10^{13}$	$5{\cdot}11.10^{15}$ $8{\cdot}92.10^{14}$ $2{\cdot}68.10^{14}$	$3{\cdot}23.10^{15}$ $9{\cdot}92.10^{14}$ $4{\cdot}24.10^{14}$ $2{\cdot}10.10^{14}$ $1{\cdot}13.10^{14}$ $6{\cdot}10.10^{13}$ $3{\cdot}78.10^{13}$	$4{\cdot}77.10^{14}$ $1{\cdot}02.10^{14}$ $3{\cdot}23.10^{13}$

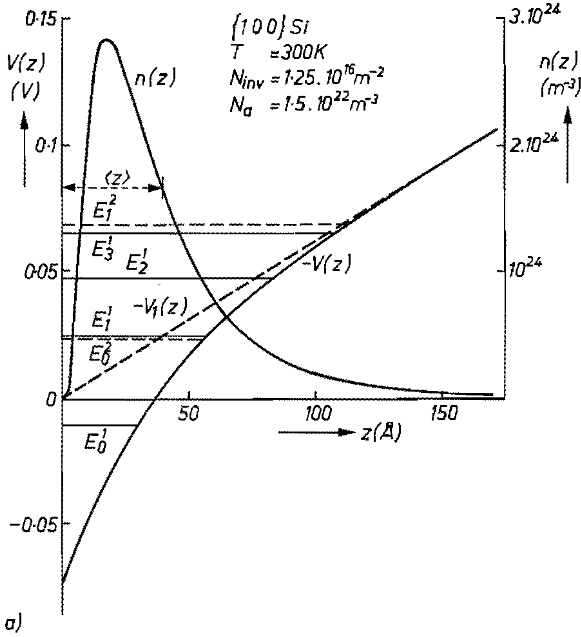
$$V_2 = 0, \quad \frac{dV_2}{dz} = 0 \quad \text{at} \quad z = z_{max}. \tag{2.32}$$

Numerical solution of the Poisson equation can be obtained with a discrete step-by-step integration starting at $z = z_{max}$ [91]. After this calculation of $V_2(z)$ we go back to the Schrödinger equation which we now solve with a potential $V(z) = V_1(z) + V_2(z)$. We find new energy levels with corresponding wave functions, a new electron concentration $n(z)$ and a new $V_2(z)$, and we then go back again to the Schrödinger equation, etc. This iterative procedure is stopped if the results of an iteration differ sufficiently little from the preceding iteration. This numerical procedure turns out to converge to a stable solution for all practical cases we have calculated. This iterative process is far more time-consuming than the process for the conventional calculation of an inversion layer which has no iteration steps. The number of discrete steps N and the value of z_{max} have therefore to be chosen carefully to find a good compromise between computer time and accuracy.

With this numerical procedure we calculate, for a certain value of the independent parameter E_{Fn}, the self-consistent potential $V(z)$, the energy levels $E_z{}^j{}_i$ for the motion perpendicular to the surface, the number of electrons $N^j{}_i$ on each level, the electron concentration $n(z)$ and the total number of charge carriers in the inversion layer N_{inv}. All possible related quantities can then easily be calculated, as for instance the average distance of the electrons from the surface $\langle z \rangle$. By repeating the calculations for a number of different values of E_{Fn}, which results in different values of N_{inv}, we may numerically determine the dependence of all possible quantities, as for instance $\langle z \rangle$, as a function of N_{inv}.

To illustrate the result of the calculations we give two examples, for an n-channel and a p-channel, under the same conditions as given in the preceding section with a conventional calculation. The various parameters of both cases and a number of results of the calculations are represented in table 2-I. The resulting potential, electron concentration and the five wave functions for the five lowest energy levels are given in figs 2.5 and 2.6.

To calculate an n-type inversion layer of silicon we have to consider the six valleys of the conduction band. However, it is not necessary to solve a Schrödinger equation for each valley separately. A certain number of valleys have the same effective-mass tensor in the axes system determined by the surface orientation. These valleys thus have exactly the same Schrödinger equation (2.18) and give the same energy levels with the same number of electrons on these energy levels. The index j may therefore be allowed to run through the number of valleys with a different effective-mass tensor. The summation over the number $n_v{}^j$ of valleys with the same reciprocal effective-mass tensor may be taken into account by multiplying the number of electrons $N^j{}_i$ out of eq.



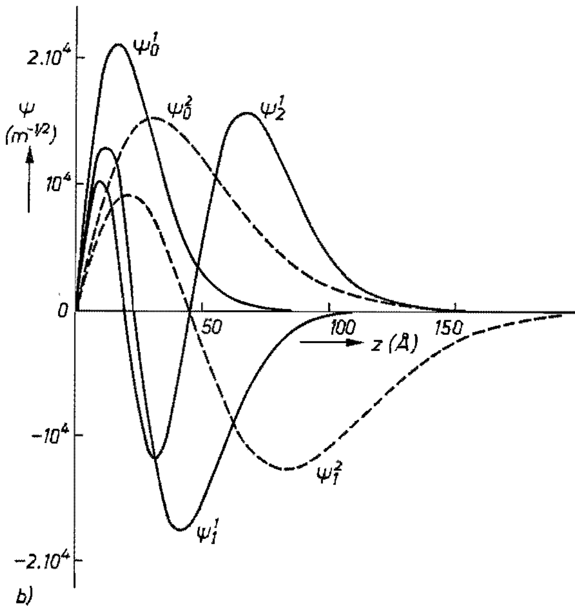


Fig. 2.5. (*a*) Electron concentration and potential for an *n*-type inversion layer with a quantized motion perpendicular to the surface. The energy levels are indicated by horizontal lines. (*b*) The corresponding wave functions for the three lowest energy levels of valley 1 and for the two lowest energy levels of valley 2.

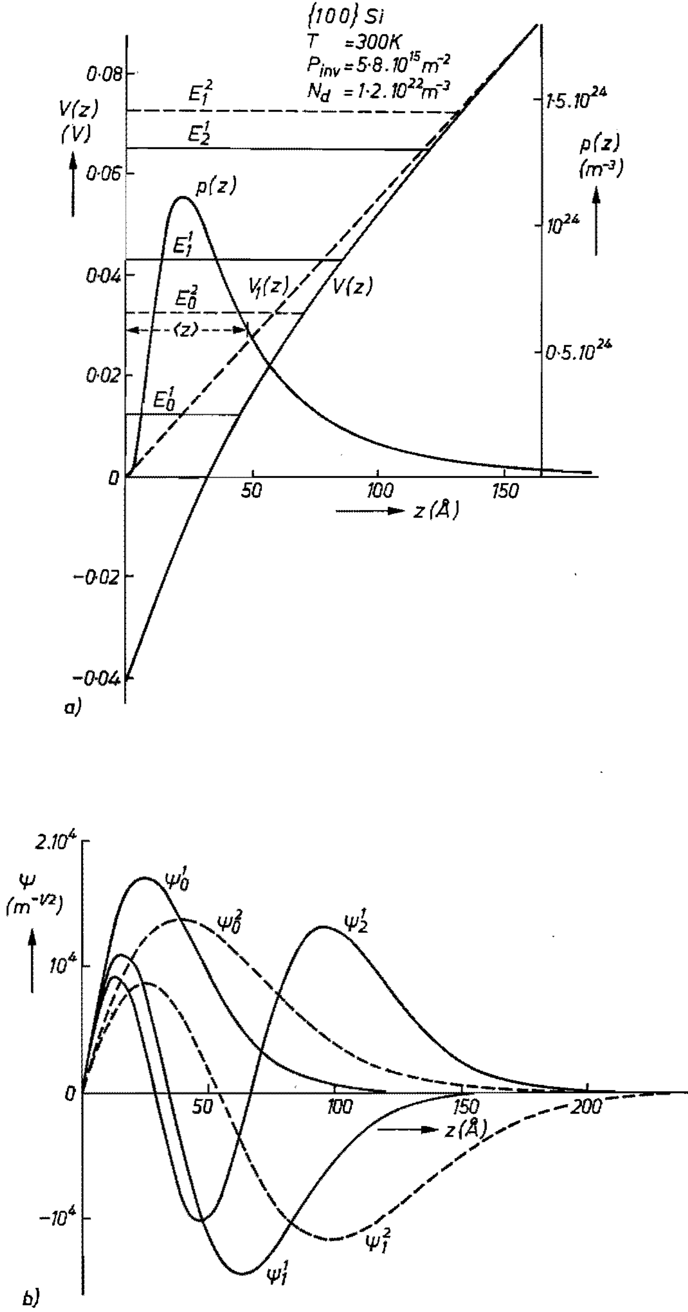


Fig. 2.6. (*a*) Hole concentration and potential for a *p*-type inversion layer with a quantized motion perpendicular to the surface. The energy levels are indicated by horizontal lines. (*b*) The corresponding wave functions for the three lowest energy levels for heavy holes and the two lowest energy levels for light holes.

(2.26) by the degeneracy factor n_v^j [12]). For a {111} surface orientation for the silicon conduction band all six valleys are equivalent and j only has the value 1 with $n_v = 6$. For a {100} surface orientation the two valleys with axis of revolution in the k_z direction are equivalent. The four remaining valleys are also equivalent. In this case j has the values 1 and 2 with n_v - values of 2 and 4. The number of carriers on the energy levels of table 2-I is the total number of electrons on that level including the degeneracy factor n_v.

2.3. Inversion and accumulation layers in the electric quantum limit

With decreasing temperature the lowest energy level for the quantized motion in the z direction is increasingly occupied with respect to the higher energy levels. At a sufficiently low temperature we may neglect the electrons at higher energy levels and only take the occupation of the lowest level into account. This situation has become known as the electric quantum limit [12]).

2.3.1. *A general solution in dimensionless variables for inversion layers in the electric quantum limit*

In the electric quantum limit the calculation is appreciably simplified. We now only have to solve one Schrödinger equation for electrons in the valley with the highest effective mass in the z direction, giving the lowest energy level for the motion perpendicular to the surface. The other valleys have higher energies and are not occupied. In this section we therefore omit the index j labelling the different valleys of the conduction band. We approximate $V_1(z)$ by its linear term for small values of z (eq. (2.13)) and we then have to solve the following set of equations:

$$\left[-\frac{\hbar^2}{2\,m_{zz}}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + q\,F\,z - q\,V_2(z)\right]\psi_0 = E_0\,\psi_0, \qquad \frac{\mathrm{d}^2 V_2}{\mathrm{d}z^2} = \frac{q\,N_0}{\varepsilon_s}\,|\psi_0|^2, \tag{2.33}$$

with boundary conditions

$$\psi_0 = 0 \quad \text{for} \quad z = 0 \quad \text{and} \quad z \to \infty, \qquad V_2 = 0, \quad \frac{\mathrm{d}V_2}{\mathrm{d}z} = 0 \quad \text{for} \quad z \to \infty; \tag{2.34}$$

ψ_0 is the normalized wave function for the lowest eigenvalue E_0 of the Schrödinger equation; N_0 is the total number of electrons in the inversion layer since only the lowest level E_0 is occupied. The linear approximation for $V_1(z)$ is allowed because the wave function ψ_0 is already practically zero before the

quadratic term in $V_1(z)$ becomes important. The solution of eqs (2.33) with boundary conditions at $z \to \infty$ and $q\,V_1 = -q\,F\,z$ for all values of z is therefore identical with the solution of the set of equations with V_1 given by eq. (2.11) and with boundary conditions at $z = d_{\mathrm{depl}}$. We may now choose N_0 to be the only independent parameter in this set of equations for a semiconductor with given m_{zz} and electric field F due to the depletion-layer charge; the equations have become independent of the temperature.

We have shown that we can get a set of equations which are also independent of the parameters m_{zz} and F by choosing suitable dimensionless variables [93]. We introduce the following dimensionless quantities:

$$\begin{aligned}
\xi &= z\left(\frac{q\,m_{zz}\,F}{\hbar^2}\right)^{1/3}; \quad \varphi(\xi) = V_2(z)\left(\frac{q\,m_{zz}}{\hbar^2\,F^2}\right)^{1/3}; \\
\varepsilon_0 &= E_0\left(\frac{m_{zz}}{q^2\,\hbar^2\,F^2}\right)^{1/3}; \quad \vartheta_0(\xi) = \psi_0(z)\left(\frac{\hbar^2}{q\,m_{zz}\,F}\right)^{1/6}; \\
\alpha &= \frac{q\,N_0}{\varepsilon_s\,F} = \frac{N_0}{N_a\,d_{\mathrm{depl}}}.
\end{aligned} \tag{2.35}$$

With these new variables the equations (2.33) and (2.34) change into

$$\begin{aligned}
-\tfrac{1}{2}\frac{\mathrm{d}^2\vartheta_0}{\mathrm{d}\xi^2} + [\xi - \varphi(\xi)]\,\vartheta_0 &= \varepsilon_0\,\vartheta_0, \\
\frac{\mathrm{d}^2\varphi}{\mathrm{d}\xi^2} &= \alpha\,\vartheta_0{}^2,
\end{aligned} \tag{2.36}$$

with boundary conditions

$$\begin{aligned}
&\vartheta_0 = 0 \quad \text{for} \quad \xi = 0 \quad \text{and} \quad \xi \to \infty, \\
&\varphi = 0, \quad \frac{\mathrm{d}\varphi}{\mathrm{d}\xi} = 0 \quad \text{for} \quad \xi \to \infty;
\end{aligned} \tag{2.37}$$

α is the only parameter in these equations and is the ratio of the charge in the inversion layer to the charge in the depletion layer.

If we have a solution of this set of equations with α as a parameter we have solved the problem of an inversion layer in the electric quantum limit for all semiconductors with all possible bulk dopes and surface orientations. The influence of the quantities m_{zz} and F, given with the semiconductor under investigation, comes in through the relations (2.35).

The set of equations (2.36) and (2.37) can be numerically solved along the

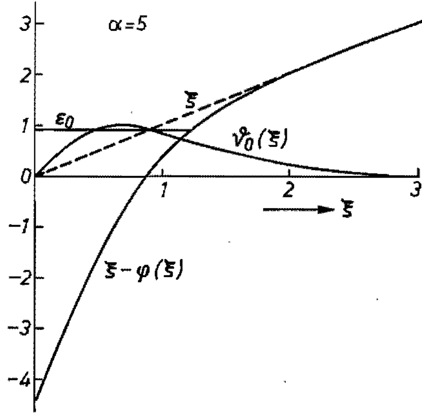


Fig. 2.7. Normalized potential $\xi - \varphi(\xi)$ and wave function ϑ_0 as a function of the normalized distance from the surface for a normalized number of carriers in the inversion layer $\alpha = 5$ in the electric quantum limit.

same lines as described in sec. 2.2.2. As a typical example, the wave function ϑ_0 and potential $\varphi(\xi)$ are given for a solution with the parameter $\alpha = 5$ in fig. 2.7. For our measurements the practical values of α are from 0·1 up to about 20.

2.3.2. *An analytical solution with variational calculus*

Although we now have a numerical solution of the equations (2.36) and (2.37) for all values of α, the drawback to this kind of solution is the lack of an analytical expression which gives insight into the general behaviour of the solution. As an exact analytical solution is impossible, it is desirable to have an approximate one. Such an approximate solution can be obtained with variational calculus [12,93]. We therefore choose a normalised function $\eta_1(\xi)$ with an adjustable parameter a, which has roughly the same shape as the exact solution of the wave function $\vartheta_0(\xi)$. We take for the trial function η_1:

$$\eta_1(\xi) = 2\, a^{3/2}\, \xi \exp(-a\, \xi). \tag{2.38}$$

For a given value of α the Poisson equation for the potential φ (eq. (2.36)) can immediately be integrated if ϑ_0 is replaced by η_1. With the boundary conditions (2.37) we find:

$$\varphi(\xi) = \frac{\alpha}{a}\,(a^2\, \xi^2 + 2\, a\, \xi + \tfrac{3}{2}) \exp(-2\, a\, \xi). \tag{2.39}$$

The adjustable parameter a has to be chosen so that the total mean energy ε_m is minimized. The contribution to the mean energy of the kinetic-energy term $-\frac{1}{2}\, \mathrm{d}^2/\mathrm{d}\xi^2$ and the external potential ξ in the normalized Hamiltonian of eq. (2.36) is $\langle \eta_1 | -\frac{1}{2}\, \mathrm{d}^2/\mathrm{d}\xi^2 + \xi | \eta_1 \rangle$. The part $\varphi(\xi)$ in the Hamiltonian is a self-

consistent potential and is caused by the electrons themselves in the inversion layer. To calculate its contribution per electron ε_φ we consider the total energy in the electric field due to the potential $\varphi(\xi)$ per unit square divided by the total number of electrons α in the inversion layer. Realizing that the energy density in the field given by $\varphi(\xi)$ is equal to $\frac{1}{2}(\mathrm{d}\varphi/\mathrm{d}\xi)^2$ in the dimensionless variables we obtain:

$$\varepsilon_\varphi = \frac{1}{\alpha}\int_0^\infty \tfrac{1}{2}\left(\frac{\mathrm{d}\varphi}{\mathrm{d}\xi}\right)^2 \mathrm{d}\xi.$$

With partial integration and using eqs (2.36) and (2.37) we find:

$$\varepsilon_\varphi = \frac{1}{\alpha}\left[-\tfrac{1}{2}\,\varphi(0)\,\frac{\mathrm{d}\varphi}{\mathrm{d}\xi}(0) - \tfrac{1}{2}\int_0^\infty \varphi\,\frac{\mathrm{d}^2\varphi}{\mathrm{d}\xi^2}\,\mathrm{d}\xi\right] =$$

$$= \tfrac{1}{2}\int_0^\infty [\varphi(0) - \varphi(\xi)]\,\eta_1{}^2\,\mathrm{d}\xi = \tfrac{1}{2}\,\langle \eta_1\,|\varphi(0) - \varphi(\xi)|\,\eta_1\rangle.$$

For the total mean energy per electron we therefore obtain:

$$\varepsilon_m = \left\langle \eta_1 \left| -\tfrac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \xi + \tfrac{1}{2}\,\varphi(0) - \tfrac{1}{2}\,\varphi(\xi) \right| \eta_1 \right\rangle. \tag{2.40}$$

Using eqs (2.38) and (2.39) we find:

$$\varepsilon_m = \frac{a^2}{2} + \frac{3}{2a}\left(1 + \frac{11}{32}\,\alpha\right). \tag{2.41}$$

ε_m is minimum if the parameter a has the value

$$a_{\min} = \left(\frac{3}{2}\right)^{1/3}\left(1 + \frac{11}{32}\,\alpha\right)^{1/3}. \tag{2.42}$$

For this value of $a_{\min}$ we find for the energy level ε_0:

$$\varepsilon_0 = \left\langle \eta_1 \left| -\tfrac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \xi - \varphi(\xi) \right| \eta_1 \right\rangle = \frac{9}{4a_{\min}}\left(1 - \frac{3}{32}\,\alpha\right). \tag{2.43}$$

The average distance from the surface is given by

$$\langle\xi\rangle = \langle \eta_1\,|\xi|\,\eta_1\rangle = \frac{3}{2a_{\min}}. \tag{2.44}$$

The potential $\varphi(0)$ at the surface turns out to be

$$\varphi(0) = \alpha \langle \xi \rangle = \frac{3\alpha}{2a_{\min}}. \tag{2.45}$$

At $\xi = 0$ the boundary condition (2.37) states that the wave function $\vartheta_0 = 0$. With the Schrödinger equation (2.36) we see that $d^2\vartheta_0/d\xi^2$ also has to be zero at the surface $\xi = 0$. Our trial function η_1, however, does not fulfil this last condition and we hope to find a better approximation for eqs (2.36) and (2.37) by choosing a trial function which also fulfils the condition for the second derivative at the surface. As a normalized trial function for this case we have tried:

$$\eta_2(\xi) = \left(\frac{4a}{7}\right)^{1/2} (a\,\xi + a^2\,\xi^2) \exp(-a\,\xi). \tag{2.46}$$

After a tedious but straightforward calculation carried out in the same way as for η_1 we find in this case:

$$a_{\min} = 5^{1/3}\left(1 + \frac{2581}{7168}\,\alpha\right)^{1/3}, \tag{2.47}$$

$$\varepsilon_0 = \frac{45}{14\,a_{\min}}\left(1 - \frac{1431}{21504}\,\alpha\right), \tag{2.48}$$

$$\langle \xi \rangle = \frac{15}{7\,a_{\min}}, \tag{2.49}$$

$$\varphi(0) = \frac{15\,\alpha}{7\,a_{\min}}. \tag{2.50}$$

For $\alpha = 0$ the solution of the Schrödinger equation (2.36) can be given in the form of Airy functions. The energy levels are then given by the condition $\mathrm{Ai}\,(2^{1/3}\,\varepsilon_i) = 0$ [8,9]). This gives for the energy level $\varepsilon_0 = 1{\cdot}855$ and for the average distance from the surface $\langle \xi \rangle = 1{\cdot}241$. These values are, of course, also obtained with the numerical solution described in sec. 2.3.1. For the variational calculation we find by using η_1:

$$\varepsilon_0 = 1{\cdot}97, \quad \langle \xi \rangle = 1{\cdot}31,$$

and by using η_2:

$$\varepsilon_0 = 1{\cdot}88, \quad \langle \xi \rangle = 1{\cdot}25.$$

We thus see that the trial function η_2 gives really a better approximation than the function η_1. In fig. 2.8 the values of ε_0, $\langle\xi\rangle$ and $\varphi(0)$ are given as a function of α as they result both from numerical calculations and from the approximate analytical solutions of eqs (2.43)–(2.45) and eqs (2.48)–(2.50). We see a fairly

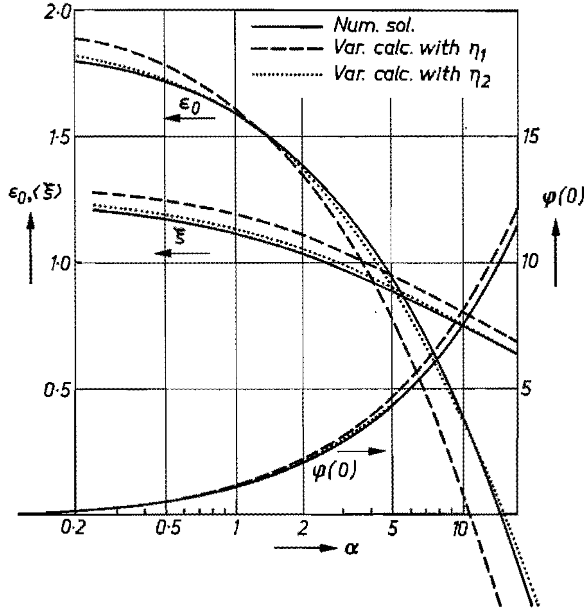


Fig. 2.8. The normalized value of energy level ε_0, potential at the surface $\varphi(0)$ and the average distance $\langle\xi\rangle$ of the carriers from the surface as a function of the normalized number of carriers in the inversion layer α in the electric quantum limit.

good agreement between the different curves and for most practical purposes the variational solution obtained by using η_1 as a trial function is sufficiently accurate. The advantage of the analytical solutions is that they are much easier to handle than a numerical solution. In chapter 4, when we need the derivative of ε_0 with respect to α we shall for instance use the analytical solution, for which the derivative can be easily calculated.

2.3.3. *A solution for accumulation layers for a non-degenerate semiconductor*

The calculation of the self-consistent potential of an accumulation layer of a semiconductor is, in general, far more difficult than for an inversion layer. In an inversion layer we only have to consider the electrons on the quantized energy levels as mobile charge carriers. For an accumulation layer, however, we have, besides the electrons bound to the surface in the energy levels for the quantized motion perpendicular to the surface, the electrons whose motion perpendicular to the surface is not quantized and whose wave functions extend

as travelling waves into the bulk. These electrons form the negative space charge in the bulk compensating for the positive charge of the ionized donors. In the accumulation layer the travelling waves are distorted by the varying potential and these electrons also make a contribution to the self-consistent potential in the accumulation layer. Recently Baraff and Appelbaum have investigated the influence of these electrons on the potential of an accumulation layer for a degenerate semiconductor [23-25]. This investigation was stimulated by Tsui's tunnel experiments through n-type InAs–oxide–Pb junctions with a quantized surface accumulation layer on the degenerate InAs bulk material [64].

We shall only investigate the electric quantum limit for an accumulation layer. For the electric quantum limit all electrons in the accumulation layer must be on the lowest energy level for the motion perpendicular to the surface. We therefore have to take a non-degenerate semiconductor at such a low temperature that the number of ionized impurities is negligible.

In the electric quantum limit there is a very important difference between an inversion layer and an accumulation layer. For an inversion layer we have a potential well $V_1(z)$ due to the depletion-layer charge and for an accumulation layer $V_1(z)$ equals zero. The potential well for the quantized motion perpendicular to the surface in an accumulation layer is fully self-consistent and is only caused by the electrons themselves on the quantized energy level.

For an accumulation layer we have to solve the same set of equations (2.33) and (2.34) as for an inversion layer if we take the electric field F to be zero. The use of the dimensionless variables of eq. (2.35) is impossible in this case with the electric field F equal to zero. We find, however, that a new set of dimensionless variables can be introduced to get a set of equations independent of the particular semiconductor under study. They are given by

$$\xi = z \left(\frac{q^2 m_{zz} N_0}{\varepsilon_s \hbar^2}\right)^{1/3}; \quad \varphi(\xi) = V_2(z) \left(\frac{\varepsilon_s^2 m_{zz}}{q \hbar^2 N_0^2}\right)^{1/3};$$
$$\varepsilon_0 = E_0 \left(\frac{\varepsilon_s^2 m_{zz}}{q^4 \hbar^2 N_0^2}\right)^{1/3}; \quad \vartheta_0(\xi) = \psi_0(z) \left(\frac{\varepsilon_s \hbar^2}{q^2 m_{zz} N_0}\right)^{1/6}. \tag{2.51}$$

With these equations we get the following set of equations for an accumulation layer at the electric quantum limit:

$$-\tfrac{1}{2} \frac{d^2 \vartheta_0}{d\xi^2} - \varphi(\xi)\, \vartheta_0 = \varepsilon_0\, \vartheta_0,$$
$$\frac{d^2 \varphi}{d\xi^2} = \vartheta_0^2, \tag{2.52}$$

with boundary conditions

$$\begin{aligned} &\vartheta_0 = 0 \quad \text{for} \quad \xi = 0 \quad \text{and} \quad \xi \to \infty, \\ &\varphi(\xi) = 0 \quad \text{and} \quad \frac{\mathrm{d}\varphi}{\mathrm{d}\xi} = 0 \quad \text{for} \quad \xi \to \infty. \end{aligned} \tag{2.53}$$

We first try to find a variational solution. As trial function we again take $\eta = 2\, a^{3/2}\, \xi \exp(-a\, \xi)$. Minimizing the mean energy

$$\left\langle \eta \left| -\tfrac{1}{2} \frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + \tfrac{1}{2}\, \varphi(0) - \tfrac{1}{2}\, \varphi(\xi) \right| \eta \right\rangle = \frac{a^2}{2} + \frac{33}{64\, a},$$

we find in the same way as in sec. 2.3.2:

$$\begin{aligned} a_{\min} &= \tfrac{1}{4} \,.\, 33^{1/3}, \\ \varepsilon_0 &= -\frac{27}{32 \,.\, 33^{1/3}} = -0{\cdot}26, \\ \langle \xi \rangle &= \varphi(0) = \frac{6}{33^{1/3}} = 1{\cdot}88. \end{aligned} \tag{2.54}$$

To find a numerical self-consistent solution of eqs (2.52) and (2.53) we cannot start the iteration procedure described in sec. 2.2.2 as we have no potential well at the surface due to the ionized impurities in the depletion layer. For an accumulation layer we therefore start the numerical iteration procedure with a potential well $\varphi(\xi)$ given by the variational calculus:

$$\varphi(\xi) = \frac{1}{a_{\min}} \left(a_{\min}{}^2\, \xi^2 + 2\, a_{\min}\, \xi + \tfrac{3}{2} \right) \exp(-2\, a_{\min}\, \xi). \tag{2.55}$$

In this case, where there is no external potential well, the numerical iteration procedure described in sec. 2.2.2 is found to be divergent. It is indeed expected that the iteration procedure for a fully self-consistent potential well has a tendency to be less stable than the iteration procedure with an extra external potential well, which is not altered after each iteration step. With an under-relaxation factor [94]) half the iteration procedure to solve eqs (2.52) and (2.53) became stable. The results of the numerical calculation for the wave function and potential are given in fig. 2.9, together with the variational solution. For the numerical calculation we find for the corresponding quantities of (2.54):

$$\begin{aligned} \varepsilon_0 &= -0{\cdot}208, \\ \langle \xi \rangle &= \varphi(0) = 1{\cdot}766. \end{aligned} \tag{2.56}$$

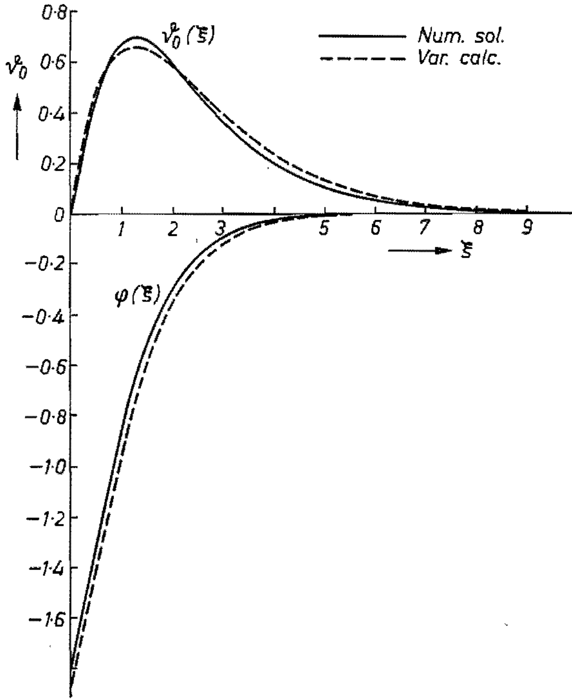


Fig. 2.9. The normalized potential φ and wave function ϑ_0 as a function of the normalized distance from the surface ξ for an accumulation layer in the electric quantum limit.

The variational calculation again gives a fairly good approximation which is suitable for most practical purposes.

2.3.4. *Range of validity of the electric quantum limit at* $T = 0$ K

In the preceding sections (2.3.1, 2, 3) we have supposed that only the lowest energy level for the motion in the z direction is occupied. In this section we shall see that even at zero absolute temperature this supposition may be wrong in the case of high surface concentrations of carriers in the inversion or accumulation layer. We shall only investigate the low-temperature limit $T = 0$ K. The number of electrons N_0 on the lowest energy level E_0 for the valleys with the highest value of m_{zz} is given by

$$N_0 = \frac{n_v\, m_d}{\pi\, \hbar^2} (E_{Fn} - E_0). \tag{2.57}$$

This equation is derived from eq. (2.26) by letting $T \to 0$ K. A factor n_v is included because the energy level E_0 may be n_v-fold degenerate due to the n_v valleys with the same effective-mass tensor giving the same energy level E_0.

(a) Inversion layers

We shall first investigate the inversion layers. By introducing the dimensionless quantities of eq. (2.35) and making E_{Fn} also dimensionless with the same factor as E_0 we can transform eq. (2.57) into

$$\varepsilon_F - \varepsilon_0 = \alpha \frac{\pi \hbar^2 \varepsilon_s}{n_v m_d q} \left(\frac{m_{zz} F}{q^2 \hbar^2} \right)^{1/3} . \tag{2.58}$$

With increasing inversion-layer concentration α the difference $\varepsilon_F - \varepsilon_0$ increases and ε_F may pass the next energy level for the motion perpendicular to the surface. This energy level then also becomes populated and we are no longer in the situation of the electric quantum limit in the sense that only the lowest energy level is occupied. To see when this situation occurs we have calculated the two lowest energy levels ε_0 and ε_1 for the valleys with the highest m_{zz} as a function of α. Due to the introduction of the dimensionless quantities these values ε_0 and ε_1 are independent of m_{zz}. We have also calculated the lowest energy level ε_0' as a function of α for carriers in valleys with a lower m_{zz} than the carriers on the level ε_0. To determine ε_0' the ratio between the two values of m_{zz} in the two different valleys, m_r, comes in. We have calculated ε_0' for two cases. The first for a {100} silicon n-type inversion layer where the ratio of effective masses m_{zz} in the two different valleys is $m_r = 0{\cdot}19/0{\cdot}98 = 0{\cdot}194$ and the second for a silicon p-type inversion layer with a ratio $m_r = 0{\cdot}16/0{\cdot}5 = 0{\cdot}32$. The results of the calculations under the condition that only the lowest energy level ε_0 is occupied are given in fig. 2.10. For electrons in a

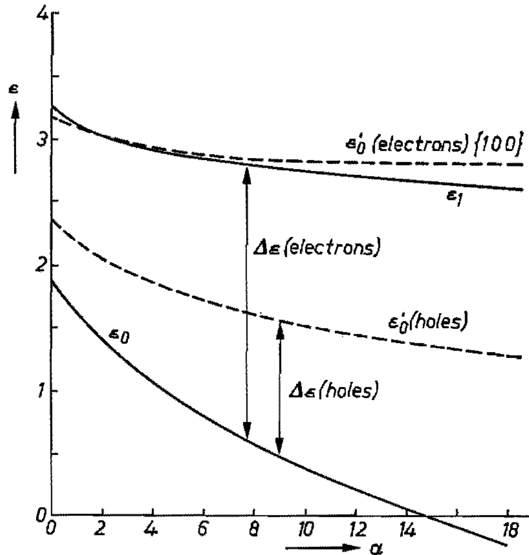


Fig. 2.10. The two lowest energy levels ε_0 and ε_1 as a function of α for charge carriers in the valleys with the highest m_{zz} value. The lowest energy level ε_0' for charge carriers with a lower m_{zz} value is also given for electrons and holes in a {100} silicon inversion layer.

{100}-oriented surface the lowest energy level ε_0 is for the two valleys with $m_{zz} = 0{\cdot}98\ m$. The next energy level is the lowest level ε_0' for the four valleys with $m_{zz} = 0{\cdot}19\ m$ for values of $\alpha < 2{\cdot}5$. For values of $\alpha > 2{\cdot}5$ the second level ε_1 for electrons with $m_{zz} = 0{\cdot}98\ m$ becomes lower than ε_0'. For holes in a silicon surface inversion layer the lowest energy level is for the heavy holes, while the next energy level is the lowest level ε_0' for light holes.

We have calculated the distance $\Delta\varepsilon(\alpha)$ between the two lowest energy levels as a function of α. Only the lowest energy level is then occupied, when the quasi-Fermi level is smaller than the second energy level. This condition is expressed by the following relation:

$$\alpha \frac{\pi \hbar^2 \varepsilon_s}{n_v m_d q} \left(\frac{m_{zz} F}{q^2 \hbar^2} \right)^{1/3} \leqslant \Delta\varepsilon(\alpha). \tag{2.59}$$

In fig. 2.11 this condition is illustrated in the α,F plane for a {100} silicon n-type and p-type inversion layer. For condition (2.59) to be valid we have to be in the region at the left-hand side of the plotted curves 1 and 2 for n- and p-type layers respectively. To see whether we can violate condition (2.59) in practice we have also plotted the curve which bounds the region in the α,F plane, where the field strength in the oxide layer covering the semiconductor is lower than the breakdown field strength of the oxide $F_{B,\mathrm{ox}}$. This region below curve 3 is given by the condition

$$\frac{\varepsilon_s}{\varepsilon_{\mathrm{ox}}} (1 + \alpha) F < F_{B,\mathrm{ox}}.$$

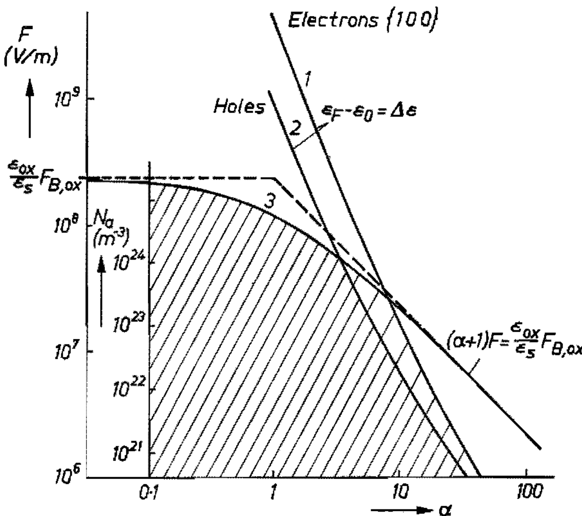


Fig. 2.11. The shaded area is the area in the α,F plane where the electric-quantum-limit condition is valid and where the field strength in the oxide covering the semiconductor is lower than the breakdown field strength.

The region where both conditions are fulfilled is shaded. The dopant concentration of the bulk material, which is required to have the corresponding value of F as given by eqs (2.12) and (2.13) with an applied bulk bias $V_b = 0$, is also indicated in the vertical axis.

We see that for concentrations $N_a < 2{\cdot}9.10^{23}$ m^{-3} for n-type inversion layers and $N_d < 1{\cdot}2.10^{24}$ m^{-3} for p-type inversion layers we may violate the electric-quantum-limit condition before the breakdown field strength of the oxide is reached. When the electric-quantum-limit condition is violated a calculation with two occupied energy levels has to be made along the same lines as described in sec. 2.2.2.

(b) Accumulation layers

The situation in an accumulation layer is different from that in an inversion layer. With the introduction of the dimensionless quantities given in eq. (2.51) we have found a bound energy level $\varepsilon_0 = -0{\cdot}208$ in the self-consistent potential well caused by the charge carriers on that level.

Our numerical calculations have shown that in that potential well there are no other bound levels for the motion perpendicular to the surface for silicon n- or p-type accumulation layers. There is only the continuous energy spectrum $\varepsilon > 0$ for the charge carriers which are not bound to the surface. The electric quantum limit is therefore valid at $T = 0$ K as long as the Fermi level is negative. By making the Fermi level dimensionless with the same factor as the energy level E_0 in eq. (2.51) we find for the dimensionless ε_F in the limit $T = 0$ K with the help of eq. (2.26):

$$\varepsilon_F = \varepsilon_0 + \frac{\pi}{n_v\, m_d} \left(\frac{\varepsilon_s^2\, \hbar^4\, m_{zz}\, N_0}{q^4} \right)^{1/3}; \qquad (2.60)$$

n_v is again the number of equivalent valleys with the highest m_{zz} value. By substituting the values of ε_0, m_{zz}, m_d and n_v we find that $\varepsilon_F < 0$, that is, the electric quantum limit is valid up to the following values of N_0:
for a {100} n-type silicon accumulation layer: $7{\cdot}4.10^{15}$ m^{-2},
for a p-type silicon accumulation layer: $2{\cdot}9.10^{16}$ m^{-2}.
In practice we may attain surface concentrations up to $1{\cdot}5.10^{17}$ m^{-2} before breakdown of the oxide occurs, so we may violate the electric-quantum-limit condition.

When we are in the situation that the lowest energy level cannot contain all charge carriers, the next energy level becomes occupied. The charge carriers on this level change the self-consistent potential well in such a way that this occupied level also becomes a bound state. It is the same situation as with the

lowest energy level ε_0 which is also only a bound state if it is occupied with charge carriers which form the self-consistent potential well.

We shall give some results of the calculations when more energy levels are occupied. These calculations are needed to interpret the measurements on accumulation layers in chapter 3. We shall only investigate the zero temperature limit $T = 0$ K. We have found that we have to consider three energy levels for silicon accumulation layers for concentrations N_0 below the maximum value $1{\cdot}5.10^{17}$ m^{-2} which can be reached in practice. These three levels are: the two lowest levels ε_0 and ε_1 for charge carriers in the valleys with the highest effective mass m_{zz}, valley degeneracy n_v, density-of-states mass m_d, and the lowest level ε_0' for carriers in the other valleys with effective mass $m_r\, m_{zz}$, valley degeneracy n_v' and density-of-states mass m_d'. The number of electrons on the levels ε_0, ε_1, ε_0' may be N_0, N_1, N_0' respectively and with the introduction of the dimensionless quantities of eq. (2.51) together with $n_1 = N_1/N_0$ and $n_0' = N_0'/N_0$ we have to solve the following set of normalized equations:

$$\begin{aligned} \tfrac{1}{2}\frac{\mathrm{d}^2\vartheta_i}{\mathrm{d}\xi^2} + (\varepsilon_i + \varphi)\,\vartheta_i &= 0, \qquad i = 0, 1, \\ \frac{1}{2\,m_r}\frac{\mathrm{d}^2\vartheta_0'}{\mathrm{d}\xi^2} + (\varepsilon_0' + \varphi)\,\vartheta_0' &= 0, \\ \frac{\mathrm{d}^2\varphi}{\mathrm{d}\xi^2} &= \vartheta_0^2 + n_1\,\vartheta_1^2 + n_0'\,\vartheta_0'^2. \end{aligned} \tag{2.61}$$

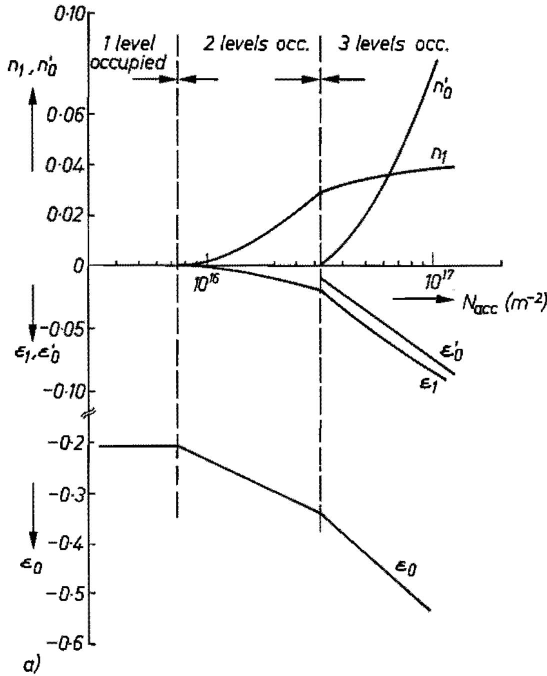
The boundary conditions are

$$\begin{aligned} \xi = 0\ :&\quad \vartheta_0, \vartheta_1, \vartheta_0' = 0, \\ \xi \to \infty:&\quad \vartheta_0, \vartheta_1, \vartheta_0', \varphi, \frac{\mathrm{d}\varphi}{\mathrm{d}\xi} = 0. \end{aligned} \tag{2.62}$$

For the relative occupation of the levels ε_1 and ε_0' we easily derive from eq. (2.26) for $T \to 0$ K in dimensionless parameters:

$$\begin{aligned} n_1 &= 0 & &(\varepsilon_\mathrm{F} \leqslant \varepsilon_1), \\ n_1 &= \frac{\varepsilon_\mathrm{F} - \varepsilon_1}{\varepsilon_\mathrm{F} - \varepsilon_0} & &(\varepsilon_\mathrm{F} > \varepsilon_1); \end{aligned} \tag{2.63}$$

$$\begin{aligned} n_0' &= 0 & &(\varepsilon_\mathrm{F} \leqslant \varepsilon_0'), \\ n_0' &= \frac{n_v'\,m_d'}{n_v\,m_d}\,\frac{\varepsilon_\mathrm{F} - \varepsilon_0'}{\varepsilon_\mathrm{F} - \varepsilon_0} & &(\varepsilon_\mathrm{F} > \varepsilon_0'). \end{aligned} \tag{2.64}$$



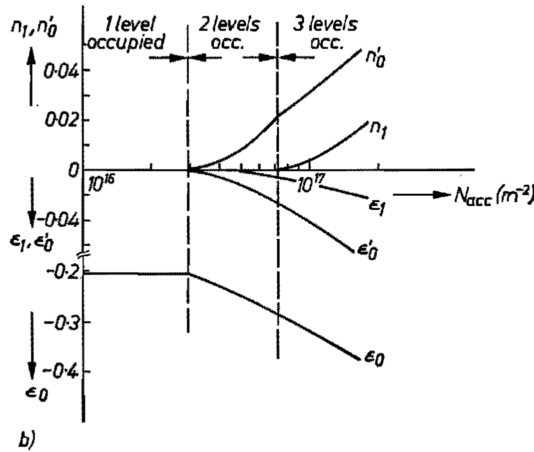


Fig. 2.12. The normalized energy levels ε_0, ε_1 and ε_0' and the relative occupation n_1 and n_0' of ε_1 and ε_0' with respect to the occupation of ε_0 as a function of the total surface concentration of charge carriers in the accumulation layer. (*a*) For a silicon {100} *n*-type accumulation layer; (*b*) for a silicon *p*-type accumulation layer.

The set of equations (2.61)–(2.64) is numerically solved with an iteration procedure.

The iteration is started with the solution of the two Schrödinger equations out of the set of equations (2.61) with a suitable first guess for the potential function φ. We then find the energy levels ε_0, ε_1 and ε_0' with the corresponding wave functions ϑ_0, ϑ_1 and ϑ_0'. If $\varepsilon_1 < \varepsilon_0'$ we choose n_1 as an independent parameter, which is kept constant after each iteration step. The occupation number n_0' and the Fermi level ε_F can then be calculated with eqs (2.63) and (2.64), as ε_0, ε_1, ε_0' and n_1 are known. If $\varepsilon_0' < \varepsilon_1$ the role of ε_0' and ε_1 is exchanged. A new potential function is now calculated by integration of the Poisson equation with the calculated functions ϑ_0, ϑ_1 and ϑ_0' and the values n_1 and n_0'. With this new potential function the procedure is again started. The iteration procedure is stopped when the result of an iteration differs sufficiently little from the preceding iteration. We stopped the iteration procedure when the relative difference for the calculated energy levels after two succeeding iteration steps was less than 10^{-4}. For each chosen value of the independent parameter n_1 we find the corresponding value of n_0', the energy levels and the wave functions.

For electrons in a {100} silicon surface we have $m_r = 0{\cdot}194$ and $n_v'm_d'/n_v m_d = 4.0{\cdot}43/2.0{\cdot}19 = 4{\cdot}54$ and for holes both values are equal to 0·32 (see table 2-I). For electrons we then find $\varepsilon_1 < \varepsilon_0'$ and $\varepsilon_0' < \varepsilon_1$ for holes. The values of ε_0, ε_1 and ε_0' are given in fig. 2.12 as a function of the total unnormalized number of charge carriers in the accumulation layer for electrons and holes in a {100} silicon surface. The relative occupation numbers n_1 and n_0' are also given in this figure. Figure 2.13 gives an example of the wave func-

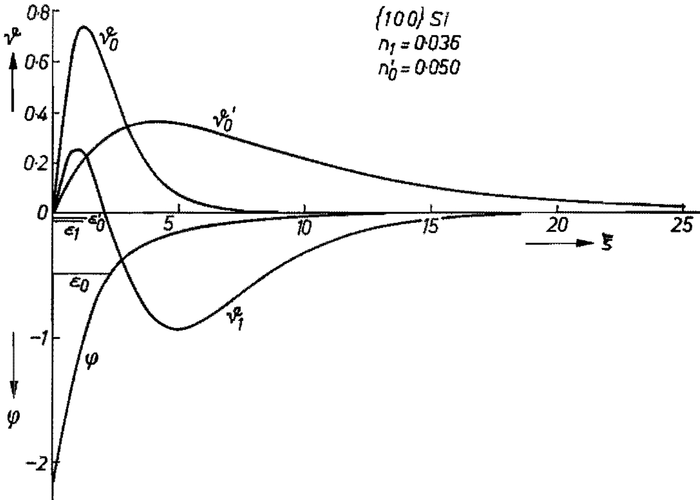


Fig. 2.13. The normalized wave functions ϑ_0, ϑ_1 and ϑ_0' together with the self-consistent potential φ for an n-type {100} silicon inversion layer with $n_1 = 0{\cdot}036$ and $n_0' = 0{\cdot}050$.

tions and self-consistent potential well for a $\{100\}$-silicon n-type accumulation layer beyond the electric quantum limit at $T = 0$ K.

We may note that at high surface concentrations the occupation of the level ε_1 may be smaller than the occupation of the level ε_0' for electrons, although ε_1 is lower than ε_0. This is due to the higher density-of-states mass and the degeneracy factor of the level ε_0'.

3. EXPERIMENTAL VERIFICATION OF QUANTIZATION BY MEASURING GATE-CAPACITANCE VARIATIONS OF AN MOS TRANSISTOR

One of the quantities on which quantization of the motion perpendicular to the surface has an appreciable influence is the average distance of the charge carriers from the oxide–semiconductor interface. If quantization is taken into account, this average distance for carriers in an inversion or accumulation layer may be appreciably larger than can be expected if we neglect quantization. This can be seen by comparing the numerical calculations of which examples are given in secs 2.2.1 and 2.2.2 with and without quantization taken into account, the other conditions remaining the same. Measurements of the average distance of the charge carriers from the interface would therefore give an experimental verification of the quantization. A straightforward way to find this average distance would be to measure a capacitance in which this average distance figures as an extra distance between the two "plates" of the capacitor [95]).

In the first section of this chapter we will investigate the relation between the average distance and the gate capacitance of a metal–oxide–semiconductor transistor (MOST). The method of measurement will be described in sec. 3.2, the results of the measurements compared with the calculations in sec. 3.3. This chapter will conclude with a discussion of the measurements.

3.1. Calculation of the gate capacitance of an MOS transistor

For the measurements we use a simple metal–oxide–semiconductor transistor shown schematically in fig. 3.1. In an n-channel MOST, the only kind which we will describe here, we have a p-type substrate with a dopant concentration of acceptors N_a. A silicon-oxide layer with a thickness d_{ox} is grown on the substrate. The oxide is covered with a metal gate. Two n^+ regions, the source and the drain, which are only just overlapped by the gate metal are made in

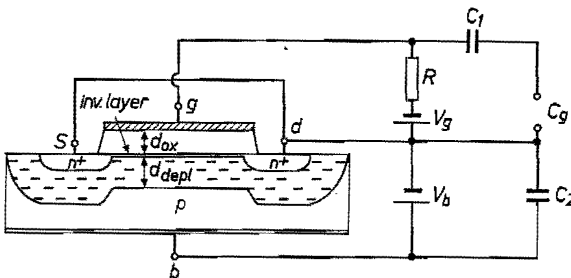


Fig. 3.1. Schematic diagram of an MOST and the configuration in which the gate capacitance is measured.

the substrate. The source and drain regions make electrical contact with the inversion layer under the gate metal, which exists when a sufficiently high applied positive gate bias V_g is applied. The inversion layer and the n^+ source and drain regions are isolated from the bulk by a depletion region. We measure the gate capacitance under the conditions shown schematically in fig. 3.1. The a.c. and d.c. supplies to the gate are separated by a sufficiently large resistor R and capacitor C_1. The source and drain contacts are short-circuited. The bulk contact is short-circuited to the source and drain for the measuring frequency by a large capacitor C_2 but may have a d.c. bulk bias V_b applied between the source and bulk contacts. This configuration has the following consequences for the physical parameters in the device. The quasi-Fermi level E_{Fn} for majority carriers in the n^+ regions and in the inversion layer is controlled by the source and drain contacts, while the quasi-Fermi level E_{Fp} for majority carriers in the bulk is controlled by the bulk contact. The difference between E_{Fn} and E_{Fp} is determined by the d.c. voltage V_b (eq. (2.2)).

To determine the gate capacitance we have to calculate the derivative of the total charge in the semiconductor or on the gate with respect to the total voltage between the two contacts of the capacitor. The total variable charge Q in the semiconductor per unit square is readily seen to be

$$Q = -q\,(N_{\text{inv}} + N_a\, d_{\text{depl}}); \tag{3.1}$$

N_{inv} is the total number of electrons in the inversion layer per unit square, while $N_a\, d_{\text{depl}}$ is the total number of the ionized acceptors in the depletion layer per unit square.

As we need only the variation in the voltage V_g between the gate and source–drain–bulk contacts to calculate the capacitance we calculate only the variable contributions to that voltage. We have two varying contributions, namely the voltage drop across the inversion and depletion layers, and the voltage drop across the oxide. The voltage drop across the depletion and inversion layers can be separated into a contribution due to the potential V_1 and another due to V_2 (eq. (2.5)). From eq. (2.11) we find for the contribution of the ionized acceptor charge density:

$$V_1(d_{\text{depl}}) - V_1(0) = -\frac{q\,N_a}{2\,\varepsilon_s}\, d_{\text{depl}}^{\,2}. \tag{3.2}$$

We find the contribution of the charge density of the electrons in the inversion layer by twice integrating Poisson's equation (2.7) and using the boundary conditions (2.8):

$$V_2(d_{\text{depl}}) - V_2(0) = -\frac{q}{\varepsilon_s} \int_0^{d_{\text{depl}}} \mathrm{d}z' \int_{z'}^{d_{\text{depl}}} n(z)\,\mathrm{d}z.$$

A change in the integration order of the variables gives

$$V_2(d_{\mathrm{depl}}) - V_2(0) = -\frac{q}{\varepsilon_s} N_{\mathrm{inv}} \langle z \rangle . \tag{3.3}$$

Here we have introduced the average distance of the electrons from the interface at $z = 0$:

$$\langle z \rangle = \frac{1}{N_{\mathrm{inv}}} \int_0^{d_{\mathrm{depl}}} z\, n(z)\, \mathrm{d}z . \tag{3.4}$$

The contribution of the voltage drop V_{ox} across the oxide, which is variable, is given by

$$V_{\mathrm{ox}} = -\frac{d_{\mathrm{ox}}}{\varepsilon_{\mathrm{ox}}} Q . \tag{3.5}$$

We therefore find for the gate voltage V_g:

$$V_g = -\frac{d_{\mathrm{ox}}}{\varepsilon_{\mathrm{ox}}} Q + \frac{q N_a}{2\, \varepsilon_s} d_{\mathrm{depl}}^2 + \frac{q}{\varepsilon_s} N_{\mathrm{inv}} \langle z \rangle + \text{const.} \tag{3.6}$$

The constant in this equation includes the work-function difference between the gate metal and the semiconductor, and the contribution of the voltage drop due to the fixed oxide charge at the oxide–semiconductor interface [96]. The value of these constant contributions is not explicitly given because they are not involved in the differential capacitance. For the gate capacitance we find:

$$C_g = -\frac{\mathrm{d}Q}{\mathrm{d}V_g} = -\frac{\mathrm{d}Q}{\mathrm{d}N_{\mathrm{inv}}} \left(\frac{\mathrm{d}V_g}{\mathrm{d}N_{\mathrm{inv}}} \right)^{-1} = \left[1 + N_a \frac{\mathrm{d}d_{\mathrm{depl}}}{\mathrm{d}N_{\mathrm{inv}}} \right] \times$$

$$\times \left[\frac{d_{\mathrm{ox}}}{\varepsilon_{\mathrm{ox}}} \left(1 + N_a \frac{\mathrm{d}d_{\mathrm{depl}}}{\mathrm{d}N_{\mathrm{inv}}} \right) + \frac{N_a}{2\, \varepsilon_s} \frac{\mathrm{d}d_{\mathrm{depl}}^2}{\mathrm{d}N_{\mathrm{inv}}} + \frac{\langle z \rangle}{\varepsilon_s} + \frac{1}{\varepsilon_s} \frac{\mathrm{d}\langle z \rangle}{\mathrm{d} \ln N_{\mathrm{inv}}} \right]^{-1} . \tag{3.7}$$

The term $N_a\, \mathrm{d}d_{\mathrm{depl}}/\mathrm{d}N_{\mathrm{inv}}$ is the ratio of the charge variation due to the ionized acceptors in the depletion layer to the charge variation of the electrons in the inversion channel. With a strong inversion layer at the surface, to which the electrons in the channel can be sufficiently rapidly supplied by the source and drain contacts, we know that this ratio is small compared to unity when the gate voltage is varied:

$$N_a \frac{\mathrm{d}d_{\mathrm{depl}}}{\mathrm{d}N_{\mathrm{inv}}} \ll 1 . \tag{3.8}$$

The average distance of the electrons $\langle z \rangle$ is found to be of the order of 100 Å or less as a result of the calculations in sec. 2.2.2. The oxide thickness d_{ox} and the depletion-layer thickness d_{depl} are usually an order of magnitude larger than $\langle z \rangle$ for common MOS transistors. We may therefore put:

$$\frac{\varepsilon_{ox}}{\varepsilon_s} \frac{\langle z \rangle}{d_{ox}}, \quad \frac{\langle z \rangle}{d_{depl}} \ll 1. \tag{3.9}$$

We introduce the difference ΔC between the oxide and gate capacitance per unit square:

$$\Delta C = C_{ox} - C_g = \frac{\varepsilon_{ox}}{d_{ox}} - C_g. \tag{3.10}$$

Using eqs (3.8) and (3.9) we can find $\langle z \rangle$ from eq. (3.7):

$$\langle z \rangle = \frac{\varepsilon_s \, d_{ox}^2}{\varepsilon_{ox}^2} \Delta C - \frac{\mathrm{d}\langle z \rangle}{\mathrm{d} \ln N_{inv}} - N_a \, d_{depl} \frac{\mathrm{d}d_{depl}}{\mathrm{d}N_{inv}}. \tag{3.11}$$

As a result of numerical calculations described in sec. 2.2 we obtain the value of $\langle z \rangle$ as a function of N_{inv}. The terms $\mathrm{d}\langle z \rangle/\mathrm{d} \ln N_{inv}$ and $N_a \, d_{depl} \, \mathrm{d}d_{depl}/\mathrm{d}N_{inv}$ are also found from these numerical calculations, since the second term may, with the aid of eq. (2.12), be rewritten in terms of the quasi-Fermi level E_{Fn}, which can be obtained directly from the numerical calculation:

$$N_a \, d_{depl} \frac{\mathrm{d}d_{depl}}{\mathrm{d}N_{inv}} = \frac{\varepsilon_s}{q^2} \frac{\mathrm{d}E_{Fn}}{\mathrm{d}N_{inv}}. \tag{3.12}$$

The experiments consist in measuring ΔC as a function of the gate voltage V_g [95]). If we use the approximation that the depletion-layer thickness is independent of the gate voltage above the threshold voltage V_{th} [2]), the total number of electrons per unit square is given by

$$N_{inv} = \frac{\varepsilon_{ox}}{q \, d_{ox}} (V_g - V_{th}). \tag{3.13}$$

With eq. (3.11) we may then investigate whether the experimentally determined ΔC as a function of V_g agrees with the numerical calculations. The reasons why we use eq. (3.11) in the form that at the right-hand side of the equation there are measured as well as calculated terms are the following. The calculated value of $\langle z \rangle$ at the left-hand side of the equation is a simple quantity which is physically relevant for investigation of the influence of quantization. The

variations in $d\langle z\rangle/d \ln N_{inv}$ only give a small correction for the range of N_{inv} values used in our experiments. The last term at the right-hand side of eq. (3.11) is only a small correction term for large values of N_{inv}, but has an appreciable influence at small values of N_{inv}. This term originates from the charge variation due to the variation in the depletion-layer width and it is not essentially different in the quantum or continuum model for the inversion layer. We therefore still put this term at the right-hand side of eq. (3.11) so that it will not obscure the influence of quantization on the calculated values of $\langle z\rangle$.

For accumulation layers the situation is somewhat different. There is no depletion layer under the accumulation layer, so there is no contribution from the dopant impurities in the bulk to the total charge in the semiconductor. We therefore see that, instead of eq. (3.11) for an inversion layer, the corresponding equation for an accumulation layer is

$$\langle z\rangle = \frac{\varepsilon_s d_{ox}^2}{\varepsilon_{ox}^2} \Delta C - \frac{d\langle z\rangle}{d \ln N_{acc}}, \tag{3.14}$$

where N_{acc} is now the total concentration per unit square of the electrons in the accumulation layer. N_{acc} depends on the gate voltage in exactly the same way as N_{inv} in eq. (3.13). The threshold voltage in this case is defined as the voltage for which $N_{acc} = 0$, and as there is no depletion-layer charge this situation occurs with flat bands at the surface. The threshold voltage in this case equals therefore the flat-band voltage [96]). At the electric-quantum-limit condition we know $\langle z\rangle$ as a function of N_{acc}. It is found with the help of eqs (2.51) and (2.54) that

$$\langle z\rangle = \frac{6}{33^{1/3}} \left(\frac{\varepsilon_s \hbar^2}{q^2 m_{zz} N_{acc}} \right)^{1/3}. \tag{3.15}$$

The term $d\langle z\rangle/d \ln N_{acc}$ may be calculated in this case, giving

$$\frac{d\langle z\rangle}{d \ln N_{acc}} = -\tfrac{1}{3} \langle z\rangle. \tag{3.16}$$

These equations can also be experimentally verified by measuring ΔC as a function of V_g. We have to bear in mind however that eq. (3.15) is only valid when the lowest energy level is occupied. The range of validity of this condition was investigated in sec. 2.3.4. For high values of N_{acc} when more levels are occupied we have to go back to eq. (3.14) for a comparison between measurements and calculations.

3.2. Description of the measurements

All measurements were made with metal–oxide–semiconductor devices having the same circular geometry. The configuration of such an MOS transistor is illustrated in fig. 3.2. The channel length, the distance between the source and drain diffusion, is 22 μm. The inner radius of the channel is 157 μm, giving a total channel area of $2{\cdot}33.10^{-8}$ m^2. The rest of the total gate-metal area of $3{\cdot}63.10^{-8}$ m^2 is the overlap of the gate over the source and drain regions. The devices were made by a planar process [97]).

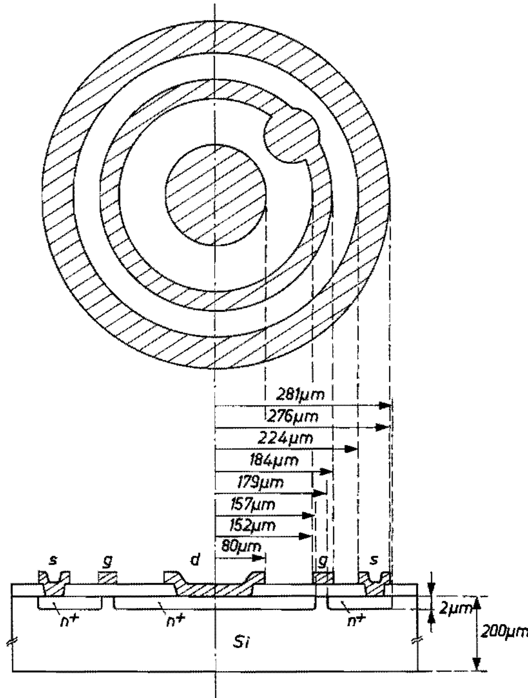


Fig. 3.2. Top view and cross-section of the geometry of the devices with which the measurements were made.

The starting material is n- or p-type {100}-oriented silicon with a homogeneous dopant concentration. The following processing steps are done.

The source and drain diffusions are made through windows made with a photo-lithographic process in the masking thermal-oxide layer grown on the silicon. For the n^+ regions we use a phosphorus diffusion; for the p^+ regions a boron diffusion. Both diffusions consist of a deposition of the impurities followed by a drive-in step. They result in n^+ or p^+ regions with a diffusion depth of about 2 μm and with such a high dopant concentration that the regions are degenerate and still conductive at very low temperatures.

After the source and drain diffusions all oxide is removed and a new gate

oxide is grown in wet nitrogen at 1000 °C for about 30 minutes followed by 5 minutes in dry nitrogen.

After a heavy phosphorus deposition for gettering purposes and removal of the resultant phosphorus-glass layer a light phosphorus deposition is done to provide a stabilizing phosphorus-silicate-glass layer.

The device is then heated for 10 minutes in dry nitrogen for gettering and stabilization, followed by a 30-minute heat treatment in wet nitrogen at 450 °C for removal of the surface states.

After the etching of contact windows aluminium is deposited as contact for the source and drain; it is heated to 300 °C to ensure good contact with the source and drain.

A number of parameters have to be known before the measurements can be compared with the numerical calculations. We will discuss the way in which we have determined the oxide thickness, the threshold voltage and the dopant concentration of the bulk material.

The determination of the oxide thickness consists in measuring the oxide capacitance between gate contact and the source–drain and bulk contacts with a d.c. gate bias causing a strong inversion layer. The area of the gate being known, the value of d_{ox}/ε_{ox} can be immediately calculated. A complication is caused by the fact that the oxide layer above the source and drain regions may be somewhat thicker than that above the channel of the device, especially in the case of devices with n^+ source and drain regions. The growth rate of the thermal oxide on an n^+ region is somewhat higher than on the low-doped bulk region. We have estimated the difference in oxide thickness of the two regions by the interference-colour difference. Combining this difference with the average oxide thickness following from the capacitance measurement, we have calculated the oxide thickness above the inversion layer. Although the determination of the oxide-thickness difference is rather crude, the final result for the oxide thickness d_{ox} is still sufficiently accurate due to the relatively small correction which is necessary due to the oxide-thickness difference.

The threshold voltage of the device at the wanted temperature and with the bulk bias voltage applied is determined by measuring the conductance between the source and drain contacts as a function of the applied gate voltage V_g with a type B201 Wayne–Kerr bridge. The value of V_g for which the conductance is zero is the threshold voltage and it can be determined with an accuracy of a few tenths of a volt by extrapolating the measured points to the conductance value zero.

Another important parameter is the dopant concentration of the bulk material of the device. This dopant concentration was determined experimentally in two ways.

Firstly we measured the capacitance of the drain–bulk diode as a function of the d.c. bias voltage applied to this diode. The applied gate bias has a value such that no inversion layer exists. Plotting the squared inverse value of this capacitance as a function of the bias voltage we obtain a straight line. The slope of this line determines the bulk dopant concentration under the drain region, if we assume that the drain region is heavily doped compared with the bulk region and that the bulk region has a constant dopant concentration N_a [98]):

$$\frac{\mathrm{d}(1/C^2)}{\mathrm{d}V} = \frac{1}{\varepsilon_s\, S_d^{\,2}}\,\frac{2}{q\, N_a}\,; \tag{3.17}$$

S_d is the area of the drain region.

In all equations used in chapter 2, where the dopant concentration is dealt with, we assumed that this concentration was a constant. There may be some doubt about this assumption especially in the surface region of the bulk under the gate oxide. During the thermal growth of the oxide layer a redistribution of impurities between the semiconductor material and the growing oxide occurs [99]) and this may cause a varying dopant concentration at the surface.

We therefore also measured the dopant concentration in a second manner by determining the threshold voltage of the MOST as a function of the bias voltage applied between the bulk and source contacts [100]). At the threshold voltage the charge in the inversion layer is zero (eq. (3.13)). Using eqs (3.1), (3.5) and (2.12) we therefore find for the threshold voltage measured between the gate and source contact:

$$V_{\mathrm{th}} = q\,\frac{d_{\mathrm{ox}}}{\varepsilon_{\mathrm{ox}}}\,N_a\, d_{\mathrm{depl}} + \mathrm{const} = \frac{d_{\mathrm{ox}}}{\varepsilon_{\mathrm{ox}}}\,[2\,q\,\varepsilon_s\,N_a\,(V_d - V_b)]^{1/2} + \mathrm{const}, \tag{3.18}$$

in which we have introduced the diffusion voltage V_d for the voltage drop across the depletion layer with a bulk voltage $V_b = 0$; V_d depends on the dopant concentration and on the position of the Fermi level at the surface, as can be seen in eq. (2.12). In practice it varies only slightly and has a value of about 0·7 V for silicon. Equation (3.18) is also derived on the assumption that N_a is constant. With this assumption the threshold voltage is linearly dependent on $(V_d - V_b)^{1/2}$. The slope of this straight line determines the dopant concentration. With all our samples we in fact find a straight line for the measured values of V_{th} as a function of $(V_d - V_b)^{1/2}$, as illustrated in fig. 3.3. For the determination of the threshold voltage we are therefore justified if we take N_a as a constant under the inversion layer.

Closer inspection of the way in which the dopant concentration occurs in the equations of chapters 2 and 3 shows that the only important parameter is the total charge of ionized impurities in the depletion layer. This total charge

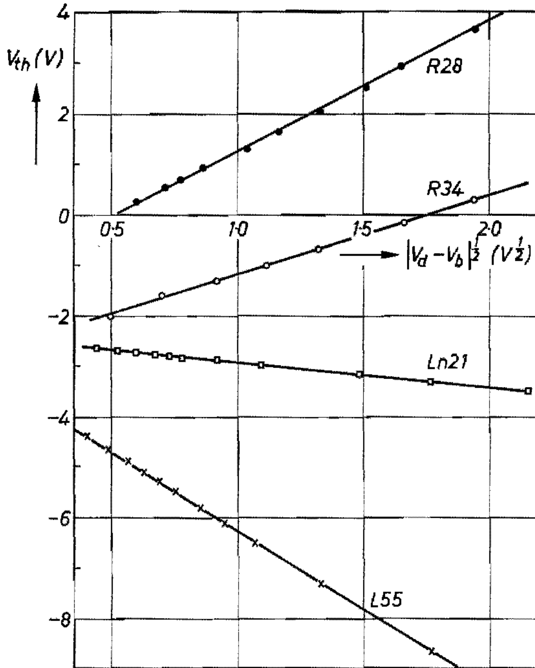


Fig. 3.3. The measured threshold voltage plotted versus the square root of the difference between bulk bias and diffusion voltage at room temperature.

is almost independent of a varying dopant concentration in a small region near the surface. By measuring the variation of the threshold voltage we just determined the total charge in the depletion layer as a function of the bulk bias and found the same dependence, within experimental error, as for a constant dopant concentration under the gate oxide for all the devices we used in our experiments. The equations derived with the assumption of a constant dopant concentration N_a may therefore be used for the devices we have measured. The values of N_a which we found with these two different manners agreed very well with each other.

In the case of accumulation-layer devices the described methods for determining the dopant concentration cannot be used because of the absence of a depletion layer. For these devices, however, the dopant concentration need not be known for our measurements at 4·2 K, because the impurities are not ionized at this temperature. We nevertheless roughly know the bulk-dopant concentration from the specific resistivity of the material used for the device.

The experimentally determined parameters of the devices on which the measurements were made are listed in table 3-I. The type of the devices is characterized by giving successively the type of the material in the source region, the bulk region and the drain region.

TABLE 3-I

Parameters of the devices used for the measurements

device number	type of device	d_{ox} (Å)	dopant conc. (m^{-3})	surface orientation
R28	n^+pn^+	1450	$N_a = 1{\cdot}5.10^{22}$	{100}
R34	n^+pn^+	2200	$N_a = 3.10^{21}$	{100}
L55	p^+np^+	1550	$N_d = 1{\cdot}2.10^{22}$	{100}
Ln21	p^+np^+	1250	$N_d = 8{\cdot}4.10^{20}$	{100}
H73-1	n^+nn^+	1250	$N_d \approx 2{\cdot}8.10^{20}$	{100}
H73-2	p^+pp^+	1100	$N_a \approx 1{\cdot}6.10^{21}$	{100}

When all these parameters have been determined the actual measurements of the gate-capacitance variations as a function of the gate voltage can be compared with the results of numerical calculations by checking the validity of eq. (3.11) or eq. (3.14).

The gate capacitance is measured on a type B201 Wayne–Kerr capacitance bridge at a measuring frequency of 500 kHz. The a.c. voltage on the bridge is supplied externally by a 500-kHz generator with an output voltage of 100 mV. The output of the bridge is amplified by a low-noise selective amplifier, whose output is detected with a communication receiver. The bridge is thus made sensitive to capacitance variations of 0·001 pF in a total gate capacitance C_g of the order of 10 pF [101]). The absolute value of the gate capacitance, however, cannot be measured with such accuracy, because the parasitic capacitance of the header on which the device is mounted is not known within sufficiently narrow limits. As we can measure only variations in C_g with the desired accuracy, and as, moreover, we cannot determine the value of ε_{ox}/d_{ox} with the same relative accuracy, we can only check the variations of $\langle z \rangle$ in eq. (3.11) or eq. (3.14) with V_g and not the absolute value. We effected these measurements at room temperature and liquid-nitrogen temperature for the inversion-layer devices and at liquid-helium temperature for the accumulation-layer devices. The low temperatures were achieved by simply immersing the device in the liquid nitrogen or helium.

3.3. Comparison between measured and calculated variations in the average inversion-layer thickness

As remarked earlier in sec. 2.2 we can numerically solve the equations determining the inversion-layer properties of a device with a known dopant con-

centration and a given applied bulk bias. By carrying out the calculations for a sufficiently large number of values of the independent variable E_{Fn}, we can find the total number of inversion-layer carriers N_{inv} and the average distance $\langle z \rangle$ of these carriers from the interface, both as functions of E_{Fn}. These functions, of course, also yield $\langle z \rangle$ and E_{Fn} as functions of N_{inv}. The correction terms in eqs (3.11), (3.12) and (3.14), $\mathrm{d}\langle z \rangle/\mathrm{d} \ln N_{inv}$ and $(\varepsilon_s/q^2)\, \mathrm{d}E_{Fn}/\mathrm{d}N_{inv}$ can be obtained from the latter functions by differentiation.

For all the inversion-layer devices whose gate-capacitance variations we measured as a function of the gate voltage, we also calculated $\langle z \rangle$ as a function of N_{inv} with a quantum-mechanical calculation described in sec. 2.2.2, as well as with a conventional continuum calculation described in sec. 2.2.1. The results are given in figs 3.4 to 3.7. The results of the quantum-mechanical calculations are indicated by the drawn curves, the "classical" calculations with the continuum model for motion perpendicular to the surface are indicated by the dashed curves. The values of $\langle z \rangle$ calculated with the conventional continuum model are increased by a constant value for easier comparison of these curves with the curves for the quantum-mechanical calculation. This constant value is indicated in the figure.

The correction terms $(\varepsilon_s/q^2)\, \mathrm{d}E_{Fn}/\mathrm{d}N_{inv}$ and $\mathrm{d}\langle z \rangle/\mathrm{d} \ln N_{inv}$ for the quantum-mechanical calculation are also given.

The measured values indicated were obtained as follows. For each value of the gate voltage the value of N_{inv} was calculated with eq. (3.13). The corresponding capacitance value was subtracted from a constant value and divided by the area of the gate to obtain the ΔC per unit square in eq. (3.11). The constant value from which all measured capacitances were subtracted was adapted to obtain values of $\langle z \rangle$ which accorded with the calculations. This is allowed if we keep in mind that only the variations of $\langle z \rangle$ with varying V_g can be determined with sufficient accuracy. The values of $\langle z \rangle$ were calculated with eq. (3.11) and the calculated correction terms; they are indicated by crosses in the figures.

It will be noted that the correction term $(\varepsilon_s/q^2)\, \mathrm{d}E_{Fn}/\mathrm{d}N_{inv}$ becomes very important at the low-inversion side of the curves. This term varies strongly with N_{inv} for small values of N_{inv}. The smaller the value of N_{inv}, the less accurately it can be determined, because the measuring inaccuracy of the threshold voltage V_{th}, which is of the order of a few tenths of a volt, becomes relatively marked. The combination of a steep rise of the correction term and the decreasing accuracy of N_{inv} with a decreasing value of N_{inv} means that $\langle z \rangle$ can only be determined with reasonable accuracy as long as this correction term is smaller than $\langle z \rangle$.

For accumulation-layer devices we have measured the variation of the gate capacitance with the gate voltage only at liquid-helium temperature. The acceptor or donor impurities in the bulk are not ionized at that temperature and we

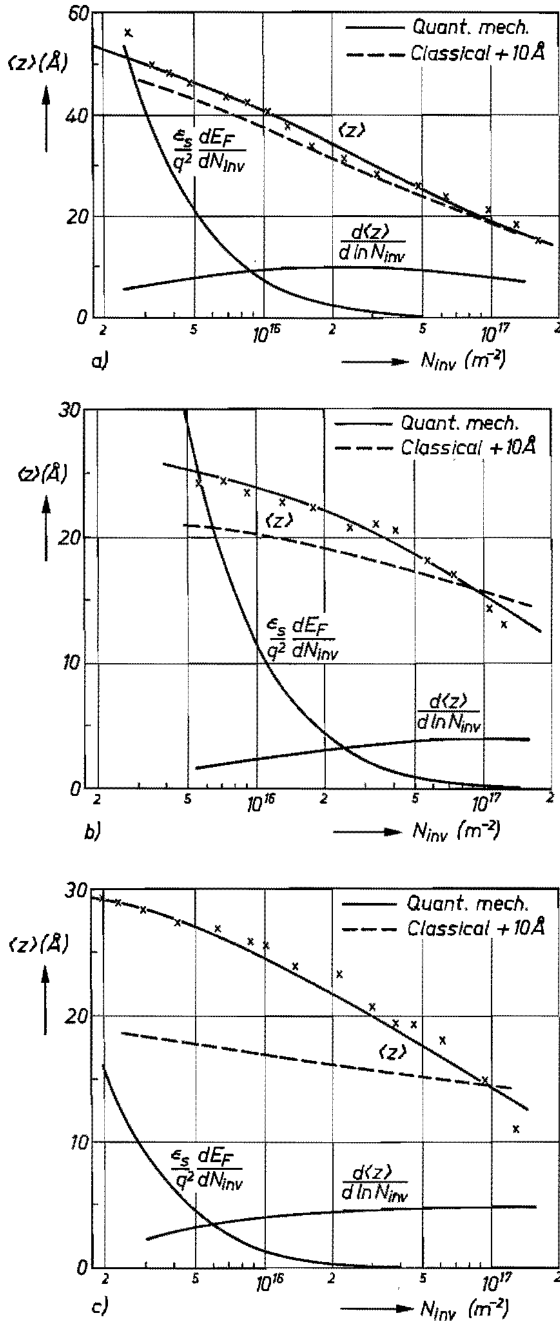


Fig. 3.4. Calculations and measurements of $\langle z \rangle$ as a function of N_{inv} for device R28 with $N_a = 1{\cdot}5.10^{22}$ m^{-3}.
(*a*) At $T = 300$ K with $V_b = 0$; (*b*) at $T = 300$ K with $V_b = -9{\cdot}7$ V; (*c*) at $T = 77$ K with $V_b = 0$.

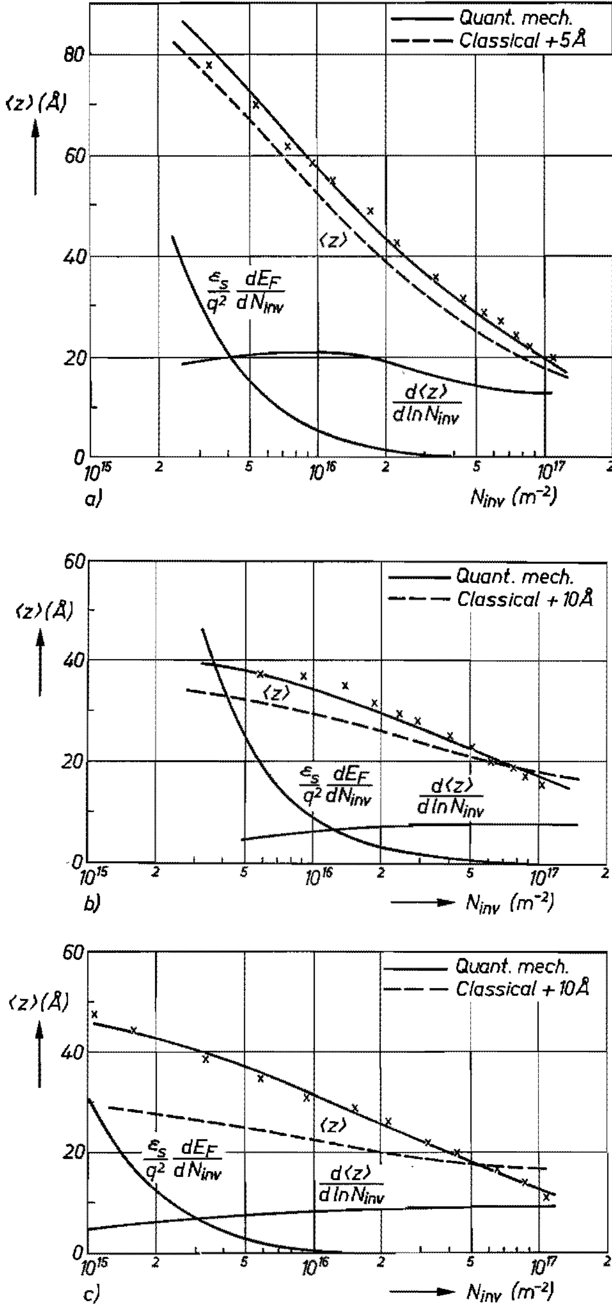


Fig. 3.5. Calculations and measurements of $\langle z \rangle$ as a function of N_{inv} for device R34 with $N_a = 3.10^{21}\ m^{-3}$.
(*a*) At $T = 300$ K with $V_b = 0$; (*b*) at $T = 300$ K with $V_b = -9{\cdot}7$ V; (*c*) at $T = 77$ K with $V_b = 0$.

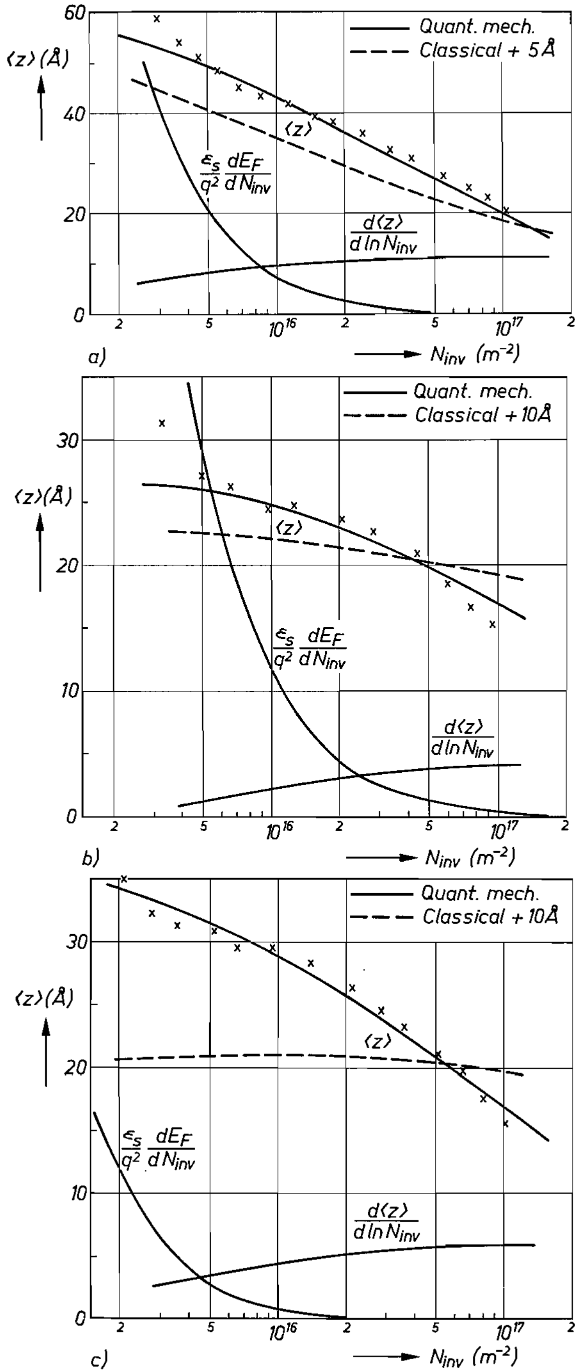


Fig. 3.6. Calculations and measurements of $\langle z \rangle$ as a function of N_{inv} for device L55 with $N_d = 1{\cdot}2.10^{22}\ m^{-3}$. (*a*) At $T = 300$ K with $V_b = 0$; (*b*) at $T = 300$ K with $V_b = 9{\cdot}7$ V; (*c*) at $T = 77$ K with $V_b = 0$.

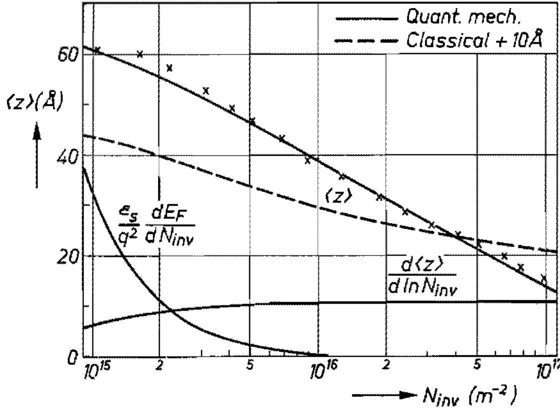


Fig. 3.7. Calculations and measurements of $\langle z \rangle$ as a function of N_{inv} for device Ln21 with $N_d = 8{\cdot}4.10^{20}$ m^{-3} at $T = 77$ K with $V_b = 0$.

may compare the measurements with the calculations described in sec. 2.3.3 and 2.3.4. As long as the electric quantum limit is valid eq. (3.15) gives $\langle z \rangle$ as a function of N_{acc}. In figs 3.8 and 3.9 this curve is shown for an n-type and p-type silicon accumulation layer by a drawn curve in the region where the electric quantum limit is valid and by a dashed curve beyond that region. The drawn curve in the last region is the result of calculations in which the occupation of the higher levels is taken into account as described in sec. 2.3.4. The measured values were obtained in the same way as for the inversion layers, except that eq. (3.14) was used with only one correction term instead of the two correction terms employed for the inversion layers.

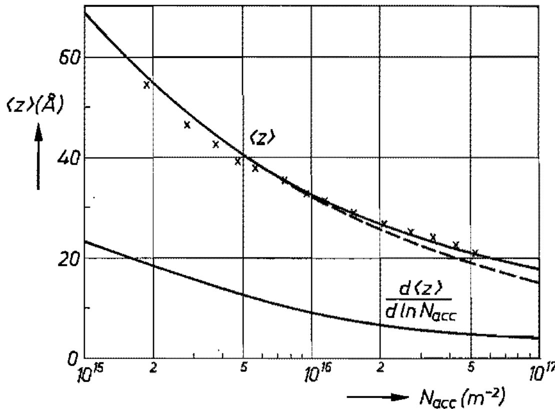


Fig. 3.8. Calculations and measurements of $\langle z \rangle$ as a function of N_{acc} for device H73-1 at $T = 4{\cdot}2$ K.

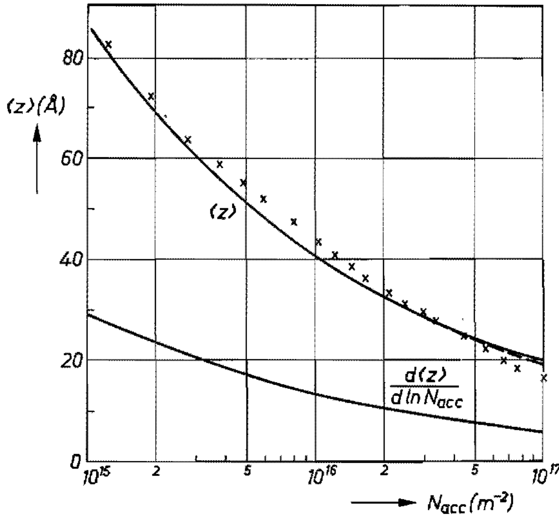


Fig. 3.9. Calculations and measurements of $\langle z \rangle$ as a function of N_{acc} for device H73-2 at 4·2 K.

3.4. Discussion

If we examine the results of the capacitance measurements as a function of the gate voltage we see overall agreement between the measurements and the calculations when the quantized motion perpendicular to the surface is taken into account. This overall agreement cannot be obtained with calculations when the conventional continuum model is used for the inversion layers. The conventional continuum calculation is not given for the accumulation layers, the measurement temperature being so low that a quantized calculation is known in advance to be necessary.

Further investigation of the results given in the preceding section for inversion layers gives rise to the following observations.

A lower dopant concentration of the bulk material causes a larger depletion-layer width and a smaller value of the electric field F, which characterises the steepness of the potential well V_1 owing to the ionized impurities in the depletion layer.

The average distance $\langle z \rangle$ and the variation of $\langle z \rangle$ with N_{inv} increase with decreasing steepness of the potential well. This can be seen by comparing the results of the calculations as well as the measurements in fig. 3.4 with those in fig. 3.5, and fig. 3.6 with 3.7. The steepness of the potential well can also be changed by applying a bulk bias. A reverse bias increases the steepness and decreases $\langle z \rangle$ and the variation of $\langle z \rangle$ with N_{inv}, as can be seen by comparing fig. 3.4*a* with 3.4*b*, 3.5*a* with 3.5*b*, and 3.6*a* with 3.6*b*.

With increasing temperature or decreasing steepness of the potential well the difference between the calculations with and without quantization of the motion perpendicular to the surface becomes less pronounced. This is readily under-

standable. When the separation of the energy levels for the motion perpendicular to the surface becomes smaller with respect to the thermal energy kT, the higher energy levels become increasingly occupied and we may approach a quasi-classical situation. In figs 3.4*a* and 3.5*a* we cannot distinguish between the calculations with and without quantization on the basis of the measurements. The energy-level separation of the two lowest levels varies from 25 meV for device R28 and 15 meV for device R34 at $N_{inv} = 5.10^{15}$ m^{-2} to 80 meV for device R28 and 60 meV for device R34 at $N_{inv} = 10^{17}$ m^{-2}. These differences have to be compared with the 26 meV for the thermal energy at room temperature. On the low-inversion side there is practically no difference between the slope of the curves calculated with and without quantization. At higher values of N_{inv}, where the energy separation between the levels is larger owing to the extra self-consistent potential well of the electrons themselves, a difference in slope again occurs. But this difference is too small to allow us to distinguish between the two curves on the basis of the few measurements in this region of N_{inv}.

Application of a negative bulk bias of $-9{\cdot}7$ V increases the energy separation to 50 meV for device R28 and to 30 meV for device R34 at $N_{inv} = 5.10^{15}$ m^{-2} and a distinction between the two kinds of calculation can clearly be made on the basis of the measurements (figs 3.4*b* and 3.5*b*).

Lowering the thermal energy to 6·5 meV at liquid-nitrogen temperature also brings us to a situation where the two calculations give significantly different results (figs 3.4*c* and 3.5*c*).

We may also remark that *n*-type channels for {100}-oriented silicon surfaces have lower values of $\langle z \rangle$ and a smaller energy separation than the corresponding *p*-type channels. This is due to the higher value of m_{zz} for *n*-type channels on {100}-oriented silicon compared with the value for the *p*-type channels for the lowest energy level ($m_{zz} = 0{\cdot}98\ m$ resp. $0{\cdot}5\ m$). We therefore see a significant difference between the calculation with and without quantization for $T = 300$ K and $V_b = 0$ V for the *p*-type channel device L55, whereas the *n*-type channel device R28 with an even somewhat higher dopant concentration does not show this significant difference (figs 3.4*a* and 3.6*a*).

In fig. 3.6*c* we see that the behaviour of $\langle z \rangle$ is qualitatively different for the conventional model calculation where $\langle z \rangle$ increases with increasing N_{inv} compared with the quantum-mechanical calculation which always gives a decreasing $\langle z \rangle$ with increasing N_{inv}. The increase of $\langle z \rangle$ with N_{inv} for the continuum calculation can be understood by realizing that in this case the Fermi level enters the valence band. The hole concentration is a function of the difference between the band edge E_v and the Fermi level E_F. At the surface where $E_F < E_v$ the hole concentration increases roughly as $(E_v - E_F)^{3/2}$ with decreasing E_F, while further on in the regions where $E_F > E_v$ the hole concentration increases as $\exp\,[(E_v - E_F)/kT]$ with decreasing E_F. The hole concentration therefore

increases relatively faster at a larger distance from the surface than near the surface with increasing N_{inv}. This effect is opposed by the increase of E_v at the surface due to the band bending caused by the space charge in the inversion layer. Nevertheless the first tendency may dominate at low temperatures, causing an increasing value of $\langle z \rangle$ with increasing N_{inv} for the calculations with the conventional continuum model.

As we are measuring relatively small changes in the gate capacitance we have to be sure that the measured variations are indeed due to the variation of $\langle z \rangle$ and not to other effects which are almost always neglected but may be important in this case. Such effects might be due to changes in the oxide properties, e.g. a change of the oxide thickness due to the forces of attraction between the two plates of the capacitor, or a variation in ε_{ox} with the electric-field strength. For our devices it is reasonable to neglect variations in oxide properties with a varying gate voltage V_g. Calculation of the influence on the capacitance of a variation in d_{ox} with the known compressibility of the oxide shows that this effect is at least an order of magnitude smaller than the influence of $\langle z \rangle$ on the capacitance. Moreover, all the variations of the oxide properties during the measurement with varying V_g would be independent of the dopant concentration of the bulk and the applied bulk bias voltage, while our measurements clearly illustrate dependence on these parameters.

The influence of surface states is also neglected in our calculations. The devices we used are devices with a low surface-state density. Moreover, the influence of surface states on the gate capacitance in the inversion region is small. The position of the Fermi level at the surface changes only slightly for large variations in the inversion-layer charge density. The change in occupation of the surface states with varying N_{inv} may therefore be neglected for calculation of the gate capacitance.

The resistance of the inversion layer may give rise to ohmic losses which influence the capacitance measurements. Owing to the favourable geometrical shape of the inversion channel and the relatively low measuring frequency we have calculated that this influence can be neglected ($\omega R C \lesssim 2.10^{-2}$). This is also experimentally shown by the negligible ohmic losses measured on the bridge with which the capacitance value is determined.

To conclude, we may remark that the measurements described in this chapter show the practical importance of quantization of the motion perpendicular to the surface of the carriers bound in a surface channel for temperatures up to room temperature. Although such striking effects as quantum oscillations proving quantization very directly at very low temperatures [51,59]) ($< 4{\cdot}2$K) do not occur in these capacitance measurements, the measurements have the remarkable result of demonstrating quantization in inversion layers at very much higher temperatures up to room temperature.

4. EXPERIMENTAL VERIFICATION OF QUANTIZATION BY MEASURING THE ANOMALOUS BEHAVIOUR IN THE GATE–BULK TRANSFER CAPACITANCE OF AN MOS TRANSISTOR

In chapter 3 we verified quantization by capacitance measurements on the MOST considered to be a two-terminal device, one terminal being the gate contact, the other the a.c. short-circuited source, drain and bulk contacts. In this chapter we shall deal with the MOST as a two-port device. The input terminal is the gate contact and the output terminal the bulk contact, while the short-circuited source and drain contacts form the common terminal for input and output. In this configuration the MOST is still a passive device.

In sec. 4.1 we shall show that the behaviour of the MOST under these circumstances can be represented by an equivalent circuit containing only capacitances. The values of these equivalent-circuit elements of the device will be calculated. The transfer capacitance between gate and bulk contacts, which is shown to be proportional to the derivative of the Fermi-level position with respect to the inversion-layer charge density, is found to behave anomalously as a function of the applied d.c. gate voltage owing to quantization of the motion of the charge carriers in the inversion layer. In sec. 4.2 the experimental method to determine this transfer capacitance as a function of the gate voltage will be described, whereas in sec. 4.3 the experimentally determined values will be compared with the calculated values. In sec. 4.4 the influence of surface states on the results of the measurements will be investigated by developing a simple model and by presenting some additional measurements with a varying oxide charge and surface-state density. The chapter will conclude with a discussion of the results.

4.1. Calculation of the two-port capacitances of an MOS system with an inversion layer

For the purpose of our calculation we consider an *n*-channel inversion layer with a surface area equal to unity and with external contacts to the gate, the bulk and the inversion layer. The energy-level diagram for an electron in this metal–oxide–semiconductor system is shown schematically in fig. 4.1 [85]). The electron-affinity difference $\Delta\chi_1$ is the difference between the electron energy in the oxide and in the semiconductor at the interface. The difference at the gate–oxide interface between the electron energy in the oxide and the Fermi level in the gate E_{Fg} is $\Delta\chi_2$; $\Delta\chi_1$ and $\Delta\chi_2$ depend only on the materials which make contact at the interface and not on the applied biases. The position of the quasi-Fermi levels E_{Fn} and E_{Fp} in the semiconductor are measured with respect to the zero level, for which the value chosen is that at which the potential V_1 reaches the surface. The potential V_1 is the solution of the Poisson

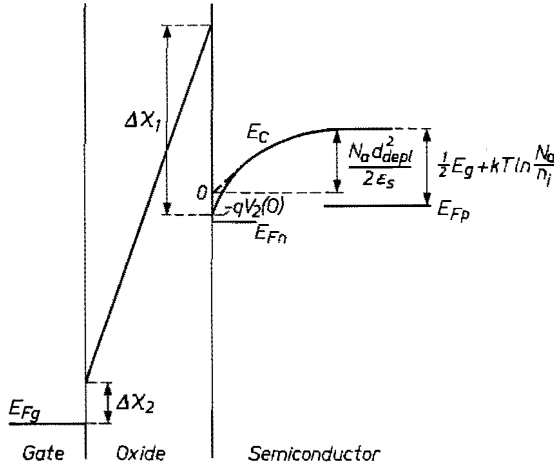


Fig. 4.1. Electron-energy diagram in an MOS structure with an inverted semiconductor surface.

equation if only the ionized impurities in the depletion layer are taken into account and the inversion-layer charge is neglected (sec. 2.1.1). The gate voltage V_g is defined as the voltage difference between the gate contact and the inversion-layer contact, and the bulk voltage V_b as the voltage difference between the bulk contact and the inversion-layer contact. The constant contribution to the voltages not giving explicitly we find the following relations for V_g and V_b:

$$V_g = -\frac{E_{Fg} - E_{Fn}}{q} = V_2(0) + \frac{q\,d_{ox}}{\varepsilon_{ox}}(N_{inv} + N_a\,d_{depl}) + \frac{E_{Fn}}{q} + \text{const}, \tag{4.1}$$

$$V_b = -\frac{E_{Fp} - E_{Fn}}{q} = -\frac{q\,N_a}{2\,\varepsilon_s}\,d_{depl}{}^2 + \frac{E_{Fn}}{q} + \text{const}. \tag{4.2}$$

The quasi-Fermi level E_{Fn} is a function of N_{inv}. We can only give an expression for this function at the electric quantum limit. Further on we shall make the approximation to use the electric-quantum-limit equations. Fortunately this approximation is allowable for the temperatures we used in measuring our samples as we shall show later on in this section. The lowest energy level for the motion perpendicular to the surface E_0 may be n_v-fold degenerate due to the n_v equivalent valleys with the maximum value of m_{zz}. Omitting the valley index j and taking n_v into account we obtain with the aid of eq. (2.26):

$$E_{Fn} = E_0 + kT \ln\left[\exp\left(\frac{\pi\,\hbar^2\,N_0}{n_v\,m_d\,kT}\right) - 1\right]. \tag{4.3}$$

At the electric quantum limit N_0 is the total number of carriers in the inversion layer:

$$N_0 = N_{\text{inv}}. \tag{4.4}$$

E_0 and $V_2(0)$ are related to the dimensionless functions ε_0 and $\varphi(0)$ by eqs (2.35); ε_0 and $\varphi(0)$ are known functions of the dimensionless inversion-layer charge, which is the ratio of the charge in the inversion layer to the charge in the depletion layer $q\, N_0/\varepsilon_s F$ (fig. 2.8). We therefore know E_0 and $V_2(0)$ as a function of N_0 and the electric field F.

Small changes $\varDelta V_g$ and $\varDelta V_b$ in the gate and bulk voltages cause a small change $\varDelta N_0$ in the inversion-layer carrier density and a small change $\varDelta d_{\text{depl}}$ in the depletion-layer thickness. Because of these changes a charge $\varDelta Q_g$ enters through the gate contact and a charge $\varDelta Q_b$ through the bulk contact:

$$\varDelta Q_g = q\,\varDelta N_0 + q\, N_a\,\varDelta d_{\text{depl}}, \tag{4.5}$$

$$\varDelta Q_b = -q\, N_a\,\varDelta d_{\text{depl}}. \tag{4.6}$$

We further introduce the following quantities which have the dimensions of a reciprocal capacitance per unit square:

$$\begin{aligned} \frac{1}{q^2}\frac{\partial E_{Fn}}{\partial N_0} &= \frac{1}{C_1}, & \frac{1}{q\,\varepsilon_s}\frac{\partial E_0}{\partial F} &= \frac{1}{C_2}, \\ \frac{1}{q}\frac{\partial V_2(0)}{\partial N_0} &= \frac{1}{C_3}, & \frac{1}{\varepsilon_s}\frac{\partial V_2(0)}{\partial F} &= \frac{1}{C_4}. \end{aligned} \tag{4.7}$$

It is now a straightforward matter to find the first-order disturbances in eqs (4.1) and (4.2), using eqs (4.3)–(4.7) and the relationship between the electric field F and the depletion-layer thickness d_{depl} (eq. (2.13)):

$$\begin{aligned} \varDelta V_g &= \left(\frac{d_{\text{ox}}}{\varepsilon_{\text{ox}}} + \frac{1}{C_1} + \frac{1}{C_3}\right)\varDelta Q_g + \left(\frac{1}{C_1} - \frac{1}{C_2} + \frac{1}{C_3} - \frac{1}{C_4}\right)\varDelta Q_b, \\ \varDelta V_b &= \frac{1}{C_1}\varDelta Q_g + \left(\frac{d_{\text{depl}}}{\varepsilon_s} + \frac{1}{C_1} - \frac{1}{C_2}\right)\varDelta Q_b. \end{aligned} \tag{4.8}$$

This set of two-port equations, which give a linear relationship between the charges entering through the contacts and the voltages on these contacts, represents a lossless passive circuit. The equations therefore have to be reciprocal because of the law of conservation of energy and the following relation holds:

$$\frac{1}{C_2} - \frac{1}{C_3} + \frac{1}{C_4} = 0. \tag{4.9}$$

We were also able to prove this relation at the electric-quantum-limit condition starting from the definition of C_2, C_3 and C_4 of eq. (4.7). We used the dimensionless formulation of the electric-quantum-limit problem as given in sec. 2.3.1. Without solving the Schrödinger equation the relation (4.9) can be proven by making use of the fact that the mean energy ε_m is minimal for the solution of the Schrödinger equation. We will not give this proof here; it only illustrates the consistency of the electric-quantum-limit equations.

For the small-signal a.c. voltages and currents v_g, v_b, i_g, i_b at angular frequency ω we find the following relations by differentiating eqs (4.8) with respect to time:

$$
\begin{aligned}
v_g &= \frac{1}{j\omega}\left(\frac{d_{\text{ox}}}{\varepsilon_{\text{ox}}} + \frac{1}{C_1} + \frac{1}{C_3}\right) i_g + \frac{1}{j\omega C_1}\, i_b, \\
v_b &= \frac{1}{j\omega C_1}\, i_g + \frac{1}{j\omega}\left(\frac{d_{\text{depl}}}{\varepsilon_s} + \frac{1}{C_1} - \frac{1}{C_2}\right) i_b.
\end{aligned}
\tag{4.10}
$$

From these equations we see that the MOS inversion-layer system can be represented by an equivalent circuit given in fig. 4.2*a*, with the following values for the capacitances per unit square:

$$
\begin{aligned}
\frac{1}{C_g} &= \frac{d_{\text{ox}}}{\varepsilon_{\text{ox}}} + \frac{1}{C_3}, \\
\frac{1}{C_b} &= \frac{d_{\text{depl}}}{\varepsilon_s} - \frac{1}{C_2}.
\end{aligned}
\tag{4.11}
$$

In practice the reciprocal capacitances $1/C_1$ to $1/C_3$ turn out to be small compared with the reciprocal oxide capacitance $d_{\text{ox}}/\varepsilon_{\text{ox}}$ and the reciprocal depletion-layer capacitance $d_{\text{depl}}/\varepsilon_s$ for a unit square of the MOS surface:

$$
\frac{1}{C_1}, \frac{1}{C_2}, \frac{1}{C_3} \ll \frac{d_{\text{ox}}}{\varepsilon_{\text{ox}}}, \frac{d_{\text{depl}}}{\varepsilon_s}.
\tag{4.12}
$$

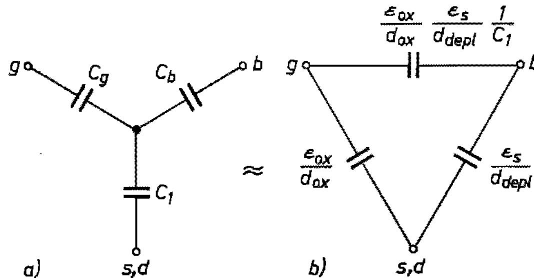


Fig. 4.2. Equivalent circuit of a unit surface area of an MOS structure with an inverted semiconductor surface for small a.c. signals.

With these relations we see that we can approximate the equivalent circuit of fig. 4.2*a* by the equivalent circuit of fig. 4.2*b*.

If we want to investigate the influence of quantization of the motion perpendicular to the surface the variation of $1/C_1$ is especially interesting. At low temperatures $1/C_1$ as a function of the d.c. gate voltage has a minimum due to quantization, and in sec. 4.2 we shall describe an experimental method for measuring this varying function of V_g. From eqs (4.3) and (4.7) we find:

$$\frac{1}{C_1} = \frac{1}{q^2}\frac{\partial E_{Fn}}{\partial N_0} = \frac{1}{q^2}\frac{\partial E_0}{\partial N_0} + \frac{\pi \hbar^2}{q^2\, n_v\, m_d}\, \frac{1}{1 - \exp(-\pi \hbar^2 N_0/n_v\, m_d\, kT)}. \tag{4.13}$$

We consider it preferable to use a dimensionless formulation of this equation and introduce the dimensionless quantities given in eq. (2.35), which made possible a general solution of the electric-quantum-limit problem. We then find:

$$\frac{d\varepsilon_F}{d\alpha} = \frac{\varepsilon_s}{C_1}\left(\frac{q\, m_{zz}\, F}{\hbar^2}\right)^{1/3} = \frac{d\varepsilon_0}{d\alpha} + \frac{\gamma}{1 - \exp(-\alpha/\beta)}; \tag{4.14}$$

ε_F is the Fermi level E_{Fn} normalized with the same factor as the energy level E_0 (eq. (2.35)).

The normalized temperature β is given by

$$\beta = \frac{q\, n_v\, m_d\, kT}{\pi \hbar^2\, \varepsilon_s\, F}, \tag{4.15}$$

and the normalized quantity γ by

$$\gamma = \varepsilon_s \left(\frac{q\, m_{zz}\, F}{\hbar^2}\right)^{1/3} \frac{\pi \hbar^2}{q^2\, n_v\, m_d}. \tag{4.16}$$

The first term on the right-hand side of eq. (4.14) is a function known from the calculations of secs 2.3.1 and 2.3.2; it is a negative increasing function of α and independent of the semiconductor material, surface orientation, bulk dope and applied bulk bias. The second term on the right-hand side of eq. (4.14) is a decreasing function of α and since it contains the effective masses, the electric field and the temperature, it depends on the material we use. This term becomes steeper with decreasing temperature and the sum of the two terms may show a minimum as a function of α as illustrated in fig. 4.3, where $d\varepsilon_0/d\alpha$ and $d\varepsilon_F/d\alpha$ are given as functions of α for practical values of β and γ.

In these calculations we make the approximation that the electric-quantum-limit equations may be used. To investigate the validity of the assumption we numerically calculated the variation of the Fermi level E_{Fn} with the total number

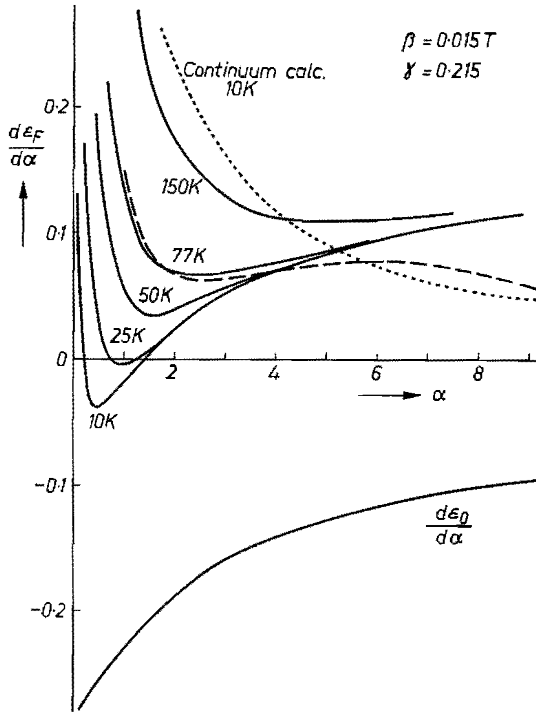


Fig. 4.3. $d\varepsilon_0/d\alpha$ and $d\varepsilon_F/d\alpha$ as a function of α for different temperatures. The drawn curves are the electric-quantum-limit approximations; the dashed curve is the result of a numerical calculation without this approximation. The dotted curve is a result for the continuum model.

of electrons in the inversion channel for a case corresponding to that in fig. 4.3 at 77 K, taking the occupation of the higher energy levels into account. The dashed curve is the result of this numerical calculation in normalized variables. For low values of α there is good agreement; for high values of α the numerically calculated value of $d\varepsilon_F/d\alpha$ becomes lower. Occupation of the higher energy levels causes a steeper decrease of the energy level E_0 with increasing N_{inv} than we would expect if only the lowest energy level was occupied. This can be understood by realizing that the occupation of a higher energy level has a relatively greater influence on the self-consistent potential well than the occupation of a lower level, owing to the larger average distance from the surface of the carriers at a higher level. As the characteristic minimum of $d\varepsilon_F/d\alpha$ in particular occurs at temperatures lower than 77 K, calculating the function in the neighbourhood of its minimum using the electric-quantum-limit approximation involves only a small error and is completely justified.

For completeness we have also given in fig. 4.3 the results of a calculation of $d\varepsilon_F/d\alpha$ in which the quantization was not taken into account. We see that the characteristic minimum in $d\varepsilon_F/d\alpha$ does not occur for the conventional continuum model.

4.2. Description of the measurements

The most straightforward way to determine the quantity $1/C_1 = (1/q^2) \times \partial E_{Fn}/\partial N_0$, whose behaviour is anomalous owing to quantization, would be to apply a small-signal gate voltage v_g and to measure the resulting bulk voltage v_b under the condition $i_b = 0$ (eq. (4.10)). We then have up to the first order:

$$\frac{v_b}{v_g} = \frac{1}{C_1} \frac{\varepsilon_{ox}}{d_{ox}}. \tag{4.17}$$

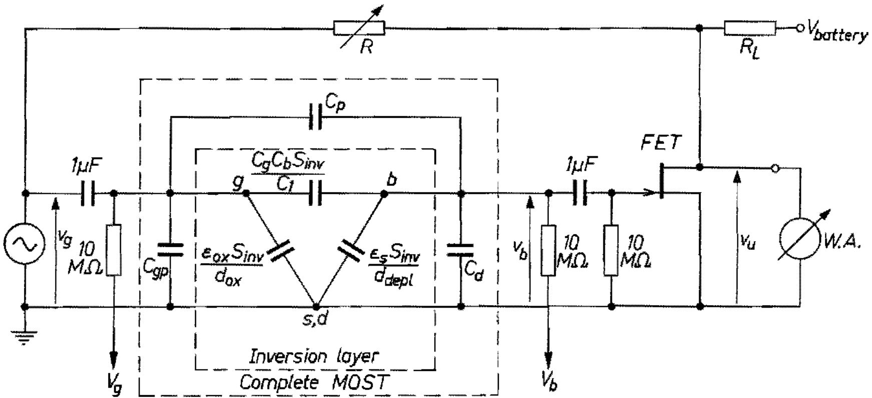


Fig. 4.4. Measurement circuit for determining $d\varepsilon_F/d\alpha$ as a function of α.

In practice, however, this equation is incomplete. In fig. 4.4 the equivalent circuit of the inversion layer is given in the inner dashed rectangle. The capacitances per unit square of fig. 4.2 are multiplied with the surface area S_{inv} of the inversion layer. Our practical devices have diffused source and drain regions to make contact with the inversion layer. We therefore have in the interior of the device an extra depletion-layer capacitance C_d between the bulk and source–drain contacts. In addition to this capacitance C_d, the complete equivalent circuit comprises a capacitance C_{gp} for the parasitic overlap of the gate contact over the source and drain regions and a parasitic capacitance C_p between the gate and bulk contacts because of the header on which the device is mounted. The total equivalent circuit is indicated in the outer dashed rectangle in fig. 4.4.

The devices on which our measurements were carried out were the same as those described in sec. 3.2. The capacitance values were approximately

$$C_d \approx 50 \text{ pF}, \ C_{gp} \approx 5 \text{ pF}, \ C_p \approx 0{\cdot}5 \text{ pF},$$

$$C_b\, S_{inv} \approx 5 \text{ pF}, \ C_g\, S_{inv} \approx 5 \text{ pF}, \text{ and } \frac{C_g\, C_b\, S_{inv}}{C_1} \approx 5.10^{-3} \text{ pF}.$$

Using these approximate values and using eq. (4.12) we find that the ratio of the resulting bulk voltage v_b to the applied gate voltage v_g, subject to the condition that the external bulk current is zero, can be expressed as

$$\frac{v_b}{v_g} = \frac{C_p}{C_t} + \frac{1}{C_1} \frac{\varepsilon_{ox}}{d_{ox}} \frac{\varepsilon_s}{d_{depl}} \frac{S_{inv}}{C_t}, \tag{4.18}$$

where C_t is the total capacitance between bulk and source–drain contacts. It is the sum of the depletion-layer capacitances $C_b\, S_{inv}$ and C_d and the parasitic capacitance due to the external leads to the device. The second term in this equation which we want to measure proves to be very small compared to the first one. We are, however, able to measure the variation in the second term, which is proportional to the quantity $1/C_1$ in which we are interested, if the constant first term owing to the parasitic capacitance C_p is compensated.

The measurement circuit is also shown in fig. 4.4. An a.c. gate voltage v_g of about 100 mV at 50 kHz is applied to the gate contact of the MOST. The resulting bulk voltage v_b is amplified by a junction-field-effect transistor with a very high input impedance and with a transconductance g_d. Owing to the high input impedance of this JFET the current through the bulk contact is practically zero. To compensate the constant contribution of the parasitic capacitance C_p in the ratio v_b/v_g a variable resistor R is connected between the input and output of the circuit. It will be readily seen that the following relation applies:

$$\frac{v_u}{v_g} = \frac{R_L}{R_L + R} - \frac{g_d R R_L}{R_L + R} \frac{v_b}{v_g} =$$

$$= \frac{g_d R R_L}{R_L + R} \left(\frac{1}{g_d R} - \frac{C_p}{C_t} - \frac{1}{C_1} \frac{\varepsilon_{ox}}{d_{ox}} \frac{\varepsilon_s}{d_{depl}} \frac{S_{inv}}{C_t} \right). \tag{4.19}$$

The 50-kHz output voltage v_u is measured with a wave analyser. The d.c. gate voltage V_g and the bulk voltage V_b are supplied via 10-MΩ resistors and do not influence the a.c. circuit.

The measuring procedure is now as follows. The resistor R is adjusted to make v_u very small at a high gate voltage V_g. In this situation the dominating contribution of C_p/C_t in the ratio v_b/v_g (eq. (4.18)) is practically compensated by the term $1/g_d R$ in the ratio v_u/v_g (eq. (4.19)). By varying V_g continuously and sufficiently slowly we can measure the output voltage v_u as a function of V_g. This function is plotted with an X–Y recorder. The variation of v_u with V_g is proportional to the variation in $1/C_1$ with V_g. Note that we are thus only able to measure the variations in $1/C_1$ and not its absolute value; the difference between the two large contributions $1/g_d R$ and C_p/C_t in eq. (4.19), which

practically cancel each other, cannot be measured with the accuracy needed to determine the absolute value of $1/C_1$.

The measurements are made at different temperatures in an experimental set-up in which all temperatures between 4·2 K and room temperature can be achieved for the device. The device is in thermal contact with a copper block in a Dewar vessel partly filled with liquid helium. The block can be heated by a resistive heating wire wound around the block and cooled by making the liquid helium boil with a heating resistor in the liquid helium, so that the cold helium gas flows over the copper block. The temperature of the copper block is measured with a thermocouple element. A control unit holds the thermoelectric voltage at a constant but adjustable value by alternately cooling and heating the copper block when the temperature becomes too high or too low. This arrangement enables the device temperature to be kept constant within $\pm$ 0·2 K [102]), which is quite sufficient for our measurements.

4.3. Comparison between measurements and calculations

From a measurement at a given temperature described in the preceding section we obtain the output voltage v_u as a function of the applied d.c. gate voltage V_g. These measurements are effected with devices R28, R34 and L55, which have already been described in sec. 3.2 and for which the various data have been given in table 3-I. The variations in the output voltage v_u and the variations in the derivative of the normalized Fermi level with respect to the normalized inversion-layer charge density, $d\varepsilon_F/d\alpha$, are both proportional to the variations in $1/C_1$ as follows from eqs (4.14) and (4.19). The normalized inversion-layer charge density α is related to the gate voltage V_g by the combination of eqs (2.35) and (3.13). We can therefore easily calculate the $d\varepsilon_F/d\alpha$ curve as a function of α by multiplying the measured curve of v_u as a function of V_g by the appropriate proportionality factors. All experimentally determined curves are given in the form of $d\varepsilon_F/d\alpha$ as a function of α.

The experimental curves in figs 4.5 to 4.9 are represented by drawn curves while the corresponding theoretical curves for the electric-quantum-limit approximation are indicated by dashed curves. These theoretical curves were calculated from eqs (4.14), (4.15) and (4.16). The electric field F which is needed for the calculations is given by eqs (2.12) and (2.13) while the various effective masses and the degeneracy factors are given by the surface orientation and the type of the inversion channel. Figure 4.5 to 4.9 show the results for devices R28, R34 and L55 at different temperatures and with different applied d.c. bulk biases V_b. All these devices have a {100}-oriented surface.

From the same silicon bar from which the {100} slices for the device L55 were cut, we also had {110}- and {111}-oriented slices at our disposal. Simultaneously with the fabrication of the L55 devices from the {100}-oriented slice we also made devices from the {110}- and {111}-oriented slices. The only

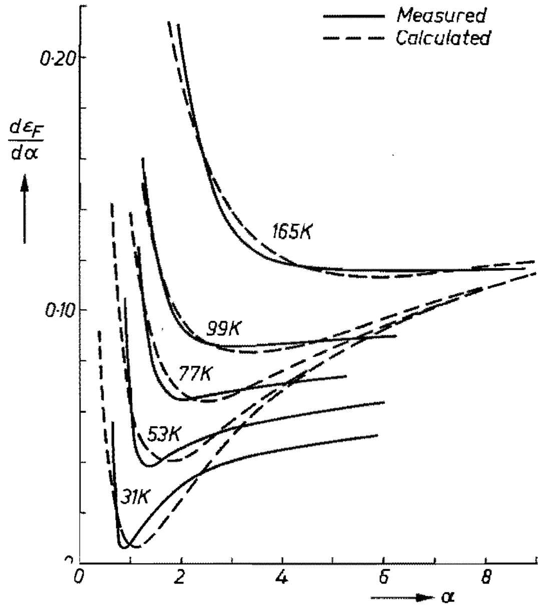


Fig. 4.5. Measured and calculated curves of $d\varepsilon_F/d\alpha$ as a function of α for device R28 with $V_b = 0$ at various temperatures.

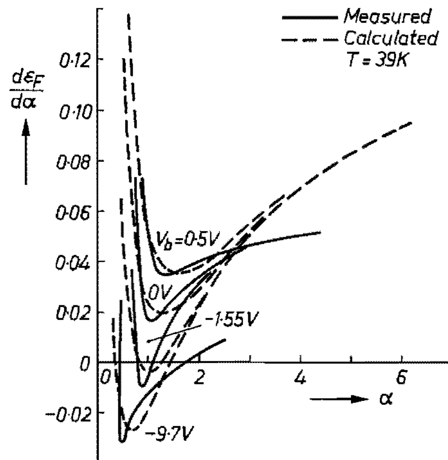


Fig. 4.6. Measured and calculated curves of $d\varepsilon_F/d\alpha$ as a function of α for device R28 at $T = 39$ K and with various bulk biases.

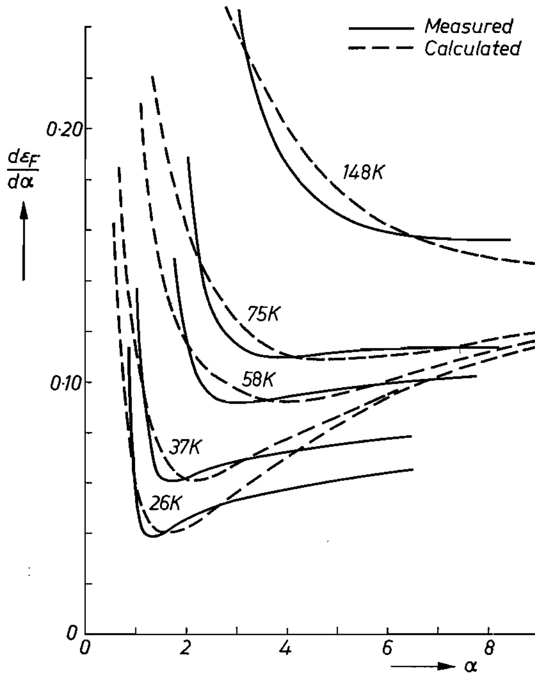


Fig. 4.7. Measured and calculated curves of $d\varepsilon_F/d\alpha$ as a function of α for device R34 with $V_b = 0$ at various temperatures.

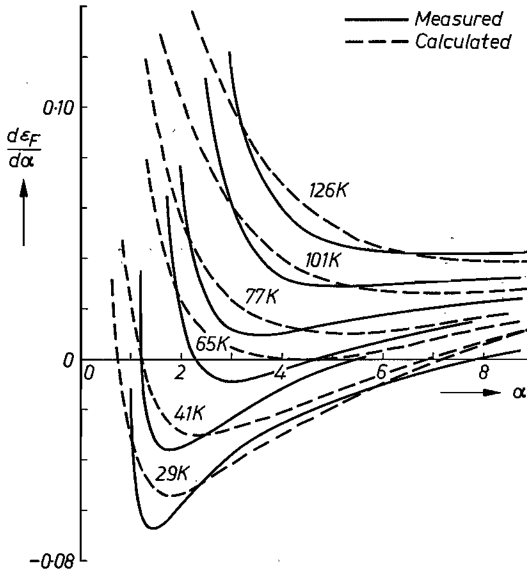


Fig. 4.8. Measured and calculated curves of $d\varepsilon_F/d\alpha$ as a function of α for device L55 with $V_b = 0$ at various temperatures.

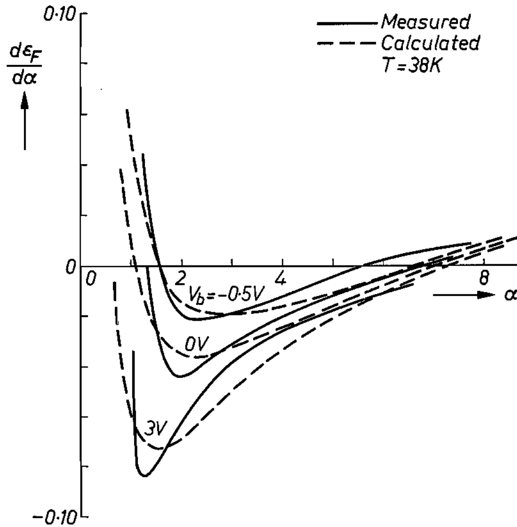


Fig. 4.9. Measured and calculated curves of $d\varepsilon_F/d\alpha$ as a function of α for device L55 at $T = 38$ K with different bulk biases.

important difference between these devices is their surface orientation; the dopant concentration and the fabrication conditions are the same. As the effective-mass tensor of holes is isotropic we would expect the behaviour of these devices to be the same, irrespective of the surface orientation. In fig. 4.10 we show the experimentally determined curves of $d\varepsilon_F/d\alpha$ as a function of α for the three different surface orientations at 38 K and with $V_b = 0$. We see that the minimum in the measured curves is much more pronounced for the $\{100\}$ orientation than for the other surface orientations. This characteristic difference is found for all temperatures and bulk biases, but we have not given these measurements. As the differences cannot be caused by the effective-mass

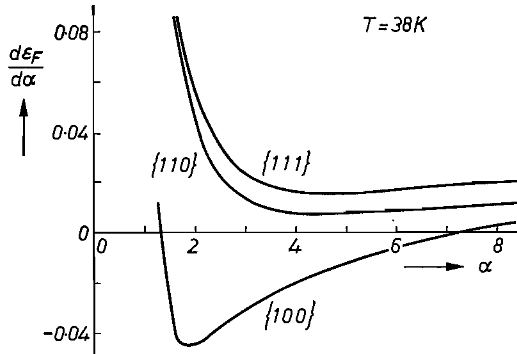


Fig. 4.10. Measured curves of $d\varepsilon_F/d\alpha$ as a function of α for p-channel devices differing only in their surface orientation at $T = 38$ K and with $V_b = 0$.

values of the holes there must be another reason. It is known that {100}-oriented oxidized silicon surfaces have a lower surface-state density and fixed oxide charge than surfaces with a different surface orientation [103,104]). The difference in the measurements may therefore be due to the different surface-state densities and oxide charges. The influence of surface states and oxide charges on the measurements will be further investigated in the next section.

4.4. The influence of surface states and oxide charges

To investigate whether surface states and fixed oxide charges indeed cause a flattening of the minimum of the $d\varepsilon_F/d\alpha$ curve we carried out the following experiment. A device L55 with a {100}-oriented surface was first measured at 38 K with $V_b = 0$ V. After this measurement the device was exposed for a certain time to X-rays from an X-ray tube with a tungsten anode having an anode voltage of 150 kV and with an intensity of 10^4 röntgens per minute. After this irradiation the device was again measured under the same experimental conditions as before exposure to the X-rays. The experimental results for $d\varepsilon_F/d\alpha$ are given in fig. 4.11 with the total duration of exposure to the X-rays

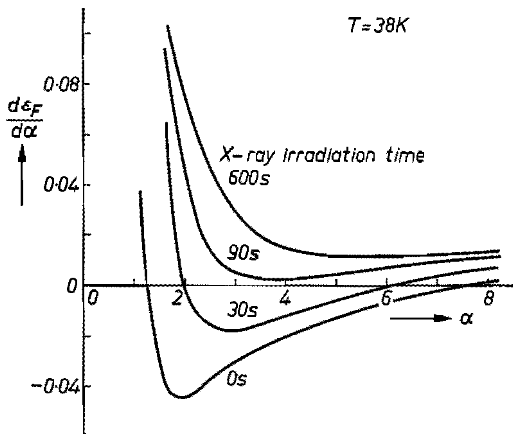


Fig. 4.11. Measured curves of $d\varepsilon_F/d\alpha$ as a function of α for device L55 after X-ray irradiation for various durations at $T = 38$ K and with $V_b = 0$.

as a parameter. It is well known that irradiation with X-rays causes an increase in surface-state density and in fixed oxide charge [105]). The total extra positive charge at the semiconductor–oxide interface due to the irradiation can be calculated from the resulting change in the threshold voltage. For our case we find an extra number of positive elementary charges of $3{\cdot}5.10^{15}$, $6{\cdot}5.10^{15}$ and $1{\cdot}5.10^{16}$ m^{-2} for exposure times of 30, 90 and 600 s, respectively. We in fact observe that these extra surface states and oxide charges have a marked flattening effect on the minimum of the curve.

This experiment gives a strong indication that the results on the {110} and {111} surfaces can in fact be explained by the extra charge in the interface for these devices compared with device L55 with a {100} surface orientation. The extra charges at the oxide–semiconductor interface for the {110} and {111} surfaces compared with the {100} surface can be found from the threshold-voltage differences and are $4{\cdot}5.10^{15}$ and $5{\cdot}5.10^{15}$ m^{-2}, respectively. These differences are of the same order as the difference for the X-ray-irradiated {100} device with a similar curve for the measured $d\varepsilon_F/d\alpha$ values.

The original behaviour of the irradiated devices can be restored by heat treatment at 300 °C for 20 minutes. The extra surface states and oxide charges are annealed by this heat treatment.

To understand the effect of surface states and oxide charges on the measured curves we consider the following influences.

As the Fermi level at the surface changes with varying inversion-layer charge density, the occupation of the surface states in the neighbourhood of the Fermi level also changes. In fact, we are measuring not the difference in the Fermi level divided by the corresponding difference in inversion-layer charge, but the difference in the Fermi level divided by the corresponding difference in the sum of the inversion-layer charge and the charge in the surface states. However, this direct influence of the charge in the surface states can be neglected in our practical situation. The two-dimensional state density of the inversion-layer carriers at the lowest energy level for the motion perpendicular to the surface is equal to $n_v m_d/\pi\hbar^2$ (eq. (2.24)) and therefore of the order of 10^{18}–10^{19} m^{-2} eV^{-1}. For oxidized silicon surfaces the surface-state density of is the order of 10^{15}–10^{16} m^{-2} eV^{-1} [106–110]. With a varying Fermi level for an inverted silicon surface the variation of the charge in the surface states is very small compared with the corresponding change in the inversion-layer charge. This is affirmed by measurements on annealed oxidized silicon surfaces [67,111].

The surface states may however have another effect, because they may be positively or negatively charged. In this respect the fixed oxide charges located at the interface play the same part as the charged surface states. A charged centre at the interface causes a local variation in the potential at the surface. These local potential variations make it impossible to separate the motion of the charge carriers in a free motion along the surface and a quantized motion perpendicular to the surface as was done by formulating the Schrödinger equation in sec. 2.1.3. The problem becomes very complicated, but we may formalize the influence of the charges at the interface by supposing that the lowest energy level E_0 for motion perpendicular to the surface is no longer a sharp level but a broadened level. We consider only the electric-quantum-limit condition, and replace the discontinuous state density for the carriers in the inversion layer with a step $n_v m_d/\pi\hbar^2$ at $E = E_0$ (eq. (2.24)) by a continuous state density given by

$$D'(E) = \frac{n_v\, m_d}{\pi\, \hbar^2} \frac{1}{1 + \exp\,[(E_0 - E)/E_s]}\,; \tag{4.20}$$

E_s is a measure for the broadening of the level E_0. For $E_s \rightarrow 0$ we get the old situation with a sharp level E_0. With this state-density function introduced a priori the inversion-layer carrier density can be calculated with the integral in eq. (2.26) as a function of the difference between E_{Fn} and the lowest energy level E_0. If we assume that E_0 is the same function of the inversion-layer carrier density as for the case without interface charges, the normalized function $d\varepsilon_F/d\alpha$ may be straightforwardly calculated by numerically evaluating the integral in eq. (2.26) and its derivative to E_{Fn}.

The results of these calculations for a p-type inversion layer with the same data as for device L55 at $T = 38$ K with $V_b = 0$ V are given in fig. 4.12 with the spreading energy E_s as a parameter. We see that a spread of about 5 meV in the energy level E_0 can explain the different behaviour of the devices with a $\{110\}$- or $\{111\}$-oriented surface or for the X-ray-irradiated devices compared with the original $\{100\}$-oriented L55 device.

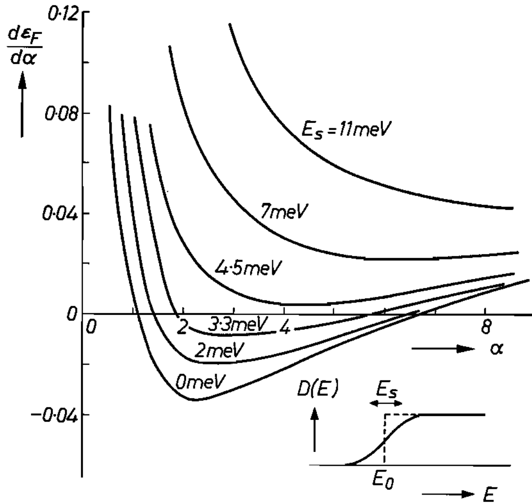


Fig. 4.12. Calculated curves of $d\varepsilon_F/d\alpha$ as a function of α for device L55 at $T = 38$ K and with $V_b = 0$ assuming a broadened energy level E_0, the broadening of which is characterized by the parameter E_s.

Stern and Howard [12]) have estimated the energy difference for an n-channel inversion layer with positive oxide charges at the interface. Owing to the attracting potential a bound state for the motion parallel to the surface is possible with a total energy lower than E_0. Including the effect of screening of the positive charge by the electrons in the inversion channel they estimate an energy

of about 0·1 Ry* below E_0, where Ry* is the Rydberg constant for two-dimensional motion:

$$\mathrm{Ry}^* = \frac{m_d\, q^4}{8\,(\varepsilon_s + \varepsilon_{\mathrm{ox}})^2\, \pi^2\, \hbar^2}\,.$$

For a silicon surface we find an energy of about 4 meV below E_0 for a charged centre exactly at the interface. In practice this sharp level is spread out owing to the overlap of the wave functions bound at neighbouring oxide charges and to the spread in the distance of the oxide charges from the interface. The energy spread which Stern and Howard find for theoretical reasons agrees very well with the spread we need to explain our measurements.

Wheeler and Ralston [65]) found for their experiments with infrared absorption a spread of about 3 meV in the energy levels for the motion perpendicular to the surface. This also agrees well with the spread of 5 meV, which we need to explain our measurements.

We may therefore conclude that the difference in the measured curves for {110}- and {111}-oriented samples compared with the {100}-oriented ones is caused by the broadening of the quantized energy levels owing to the potential variations of the charged centres in the interface.

4.5. Discussion

For all the devices with a {100} surface orientation we see a qualitative agreement between the measurements and the calculations of the derivative of the normalized Fermi level $\mathrm{d}\varepsilon_F/\mathrm{d}\alpha$ as a function of the normalized inversion-layer charge density α. For the p-channels the agreement is better than for the n-channels. The measurements for the n-channel devices R28 and R34 give a sharper but less deep minimum in the $\mathrm{d}\varepsilon_F/\mathrm{d}\alpha$ curve than the corresponding calculations. In all the devices measured the dependence of the measured curves on temperature and bulk bias is qualitatively the same as for the calculations. With increasing temperature the minimum becomes flatter and at sufficiently high temperatures the minimum disappears. Applying a forward bulk bias has the tendency to flatten the minimum, while a reverse bulk bias causes a sharper minimum in both the calculated and the measured curves. The dependence of the measured curves on the bulk bias is somewhat greater than expected on the basis of the calculations. The quantitative differences between measurements and calculations are very difficult to explain. They may be due to a number of simplifying assumptions we have made.

The influence of the surface states and oxide charges is neglected in all the calculations of sec. 4.3. The indirect effect of surface states and oxide charges on the minimum in the $\mathrm{d}\varepsilon_F/\mathrm{d}\alpha$ curve is to flatten this minimum, owing to the

broadening of the energy levels for motion perpendicular to the surface. This influence is experimentally demonstrated in sec. 4.4 and can also be understood from the model with a broadened energy level as calculated in that section. This effect has the opposite tendency to explain the sharper minima in the measured curves for the *n*-channel devices.

For practical values of surface-state densities the influence of the change in occupation of these states is shown to be negligible. At low temperatures however, the Fermi level enters the band in the inversion layer at the surface and values of surface-state densities in that energy region are hardly known. A slight effect of the change in occupation of the surface states therefore cannot completely be excluded.

In all our calculations we have used the effective-mass approximation for the carriers in the inversion layer. As the potential in the inversion-layer region varies appreciably over a unit cell, the effective-mass approximation may be invalid. To solve the problem for an inversion layer without an effective-mass approximation the periodic potential variations in the unit cell have to be used explicitly. This is a very complicated task and no known attempt has been made in the literature to solve this problem, although the importance of the possible influence of the effective-mass approximation on the final results has already been recognized [12]).

The interaction between the carriers in the inversion layer is approximated by taking into account the average space-charge density of the carriers for the solution of the Poisson equation. This is the same kind of approximation as the Hartree approximation used for the calculation of atomic wave functions. Alferieff and Duke [22]) have indicated that the density of states may be more discontinuous than we obtain with our approximation. This effect may also contribute to a sharper minimum in the measured curves compared with the calculated ones.

In conclusion we may remark that the measurements described in this chapter experimentally demonstrate quantization in inversion layers in a temperature range from about 25 K to 75 K. A characteristic minimum in the ratio of the change of the Fermi level to the corresponding change in inversion-layer charge density occurs. This minimum is due to the combination of two effects. On the one hand the lowest energy level for the motion perpendicular to the surface decreases with increasing inversion-layer charge density. On the other hand the inversion-layer charge density changes from an exponential to a linear dependence on the difference between the lowest energy level and the Fermi level, when the Fermi level passes the lowest energy level (eq. (4.3)). For sufficiently high temperatures this minimum disappears; for very low temperatures measurements are no longer possible owing to the freezing-out of the majority carriers in the bulk.

These measurements demonstrate the importance of quantization more

directly for inversion layers than the measurements of chapter 3. The temperature range in which quantization is demonstrated by measuring the minimum in the $d\varepsilon_F/d\alpha$ curve is smaller than the temperature range of the measurements of chapter 3, but it is still much larger than for previous experiments below 4·2 K with a direct demonstration of quantization in inversion layers.

5. CONCLUSIONS AND REMARKS

Surveying the results described in this thesis, we may draw the following conclusions and make certain observations.

In chapter 2 we give a method of numerically solving the inversion-layer problem. This method had to be at our disposal for the interpretation of the gate-capacitance measurements. The necessity of making a numerical program illustrates one of the serious drawbacks of a numerical solution of a certain problem. Although several numerical solutions were already known we could not use them because of the different cases for which they were calculated.

In the same chapter we show that the introduction of dimensionless variables for the equations governing the behaviour of charge carriers in an inversion or accumulation layer with a quantized motion perpendicular to the surface makes it possible to obtain a general solution of these equations at the electric quantum limit. This dimensionless solution is only valid for temperatures which are so low that only the lowest energy level is occupied. The solution is independent of the properties of the materials considered such as effective masses and the electric field due to the depletion layer as long as the general model is applicable. Material constants which may depend on the surface orientation and external parameters such as applied biases are only involved via the characteristic quantities used to make the variables dimensionless. Once we have obtained a solution to the dimensionless set of equations the influence of the value of the effective masses, the bulk dopant concentration and the externally applied biases on the various quantities of the surface layer can immediately be seen at the electric quantum limit. A numerical solution of the dimensionless equations is given, but we show that fairly good analytical approximations to the exact solution can be obtained with variational calculus. An analytical solution is very easy to use in practice compared with a numerical solution and particularly in chapter 4 the advantage of using the analytical solution becomes apparent.

With the gate-capacitance measurements described in chapter 3 we were able to determine the variations in the average thickness of the inversion layers or accumulation layers on p- or n-type silicon material. Even at room temperature the quantization in the inversion channel can be experimentally shown to have an appreciable influence on the measured capacitance variations. It is a somewhat indirect way of demonstrating the quantization but the temperature range for which quantization can be shown to be important is very large. These measurements constitute a strong indication that it is indeed necessary to take the quantization into account for inversion layers at room temperature to understand, for instance, the mobility in these layers.

The measurements described in chapter 4 give a more direct experimental verification of the quantization in silicon p- and n-type inversion channels. The temperature range for these measurements, namely from about 25 K to 75 K, is appreciably smaller than for the gate-capacitance measurements. Nevertheless, this temperature range is much larger than the temperature range of previous measurements directly demonstrating the quantization, which were effected below 4·2 K [51–66]. The measurements are based on anomalous behaviour of the externally measurable transfer capacitance between gate and bulk contacts when the Fermi level in the inversion layer passes through the lowest quantized energy level for the motion perpendicular to the surface. A dip in this capacitance occurs which becomes flatter for increasing temperatures and disappears at a sufficiently high temperature, namely about 75 K for the cases we investigated.

The measurements and calculations do not agree exactly. Quantitatively there is a relatively small discrepancy. This suggests that the simplifying assumptions made for the calculations influence the final results.

For instance, the effective-mass approximation with the bulk values for the effective masses may be an oversimplification of the problem. Particularly in the case of the effective mass perpendicular to the surface an error may be introduced. The potential changes rapidly in the direction perpendicular to the surface over the small thickness of the inversion layer. Moreover, the atoms at the surface may be somewhat displaced at the surface from the position they would occupy if the periodicity of the crystal continued right up to the surface. The strain which always exists in the oxide–silicon interface may also cause changes in the effective-mass values.

Another approximation is that the interaction between the carriers in the inversion layer is taken into account by the average space-charge density of the carriers when solving the Poisson equation, in analogy with the Hartree approximation for calculating wave functions of atoms with more than one electron around the nucleus. In reality the density of states may be more discontinuous than we find with our approximation [22]. This may be an indication as to why the measured curves in chapter 4 show a sharper dip than the calculated curves. These measurements in chapter 4 in particular suggest that theoretical investigation of the influence which the simplifying assumptions have on the final results might prove very interesting, although it would possibly also be very difficult.

Experiments with X-ray irradiation illustrate the importance of the broadening of the quantized energy level by the randomly distributed charged centres in the oxide–semiconductor interface. The spread in the energy levels necessary to explain the measurements is of the same order as found in previous measurements [65]) and as can be expected on theoretical grounds [12,22].

List of symbols

a	variational-calculus parameter
$a_{\min}$	value of a for the minimum of the mean energy
B	magnetic induction
C_{ox}	oxide capacitance per unit square ε_{ox}/d_{ox}
C_g	gate capacitance per unit square dQ/dV_g
$\Delta C = C_{ox} - C_g$	
C_b	depletion-layer capacitance per unit square between inversion layer and bulk of semiconductor
$C_1 = \left(\frac{1}{q^2}\frac{\partial E_F}{\partial N_0}\right)^{-1}$	
C_d	depletion-layer capacitance between source–drain contacts and bulk contact of an MOST
C_{gp}	overlap capacitance between gate and source–drain regions of an MOST
C_p	parasitic header capacitance between gate and bulk contacts
C_t	total capacitance between bulk and source–drain contacts including the parasitic capacitances
$D(E)$	density-of-states function
d_{depl}	depletion-layer thickness
E	energy
E_c, E_v	energy of conduction- and valence-band edges
E_g	band gap
E_F, E_{Fn}, E_{Fp}	Fermi level and quasi-Fermi level for electrons and holes
$E_{z\,i}^{\,j}$	ith energy level of valley j for motion perpendicular to the surface
E_0	lowest energy level for motion perpendicular to the surface
E_s	spreading parameter for energy level E_0
F	electric field at the surface due to depletion-layer charge $q\,N_a\,d_{depl}/\varepsilon_s$
F_2	electric field due to inversion-layer charge $-dV_2/dz$
$F_{B,ox}$	breakdown field strength of oxide
g_d	transconductance
$\hbar$	Planck constant devided by $2\pi = 1{\cdot}0545.10^{-34}$ J s
h	step size in numerical calculations
i_g, i_b	small-signal a.c. current flowing through gate and bulk contacts per unit square of the inversion layer of an MOST
k	Boltzmann constant $= 1{\cdot}38.10^{-23}$ J K^{-1}

k_x, k_y	wave-vector components in x and y directions
m	free-electron mass $= 9{\cdot}11.10^{-31}$ kg
m_e, m_h	effective mass of electrons and holes
m_d	density-of-states mass
$(m_{kl}{}^j)^{-1}$	reciprocal effective-mass-tensor elements for valley j
m_r	ratio of effective masses m_{zz} for the valleys with the next-highest and highest values of m_{zz}
N	number of discrete steps in numerical calculations
N_a, N_d	ionized-acceptor and -donor concentrations
$N^j{}_i$, N_0	number of inversion-layer carriers per unit square at energy level $E_z{}^j{}_i$ and E_0
N_{inv}, N_{acc}	total number of inversion-layer and accumulation-layer carriers per unit square
n	electron concentration
n_i	intrinsic concentration
$n_v{}^j$	number of degenerate valleys j
n_1, n_0'	ratio of occupation of the energy levels ε_1 and ε_0' to the occupation of the level ε_0
p	hole concentration
Q	total variable charge per unit square in the semiconductor
q	elementary charge $= 1{\cdot}602.10^{19}$ C
R, R_L	feedback resistor and load resistor in measuring circuit
Ry*	Rydberg constant for two-dimensional inversion layer
S_{inv}	surface area of inversion layer
T	absolute temperature
V, V_1, V_2	total potential, potential owing to depletion-layer charge and potential owing to inversion-layer charge in the semiconductor
V_b	applied bias between bulk and source–drain contacts
V_g	applied gate voltage
V_{th}	threshold voltage of an MOST
V_d	diffusion voltage for the depletion layer between inversion layer and bulk of the semiconductor
v_g, v_b	small-signal a.c. voltages on gate and bulk contacts of an MOST
x, y	coordinates parallel to surface
z	coordinate perpendicular to surface
$z_{\max}$	maximum distance from the surface to which numerical calculations are extended
$\langle z \rangle$	average distance of inversion-layer carriers from the surface
α	inversion-layer charge density normalized to charge density in depletion layer $= q\, N_0/\varepsilon_s F$

β	normalized absolute temperature
γ	dimensionless constant
ε_s, ε_{ox}	permittivity of semiconductor and oxide
ε_0, ε_1, ε_0'	normalized energy levels for motion perpendicular to the surface
$\Delta\varepsilon$	normalized energy difference between the two lowest energy levels for motion perpendicular to the surface
ε_F	normalized Fermi level
ε_m	normalized mean energy
η	normalized trial function for variational calculus
ϑ_0	normalized wave function for energy level ε_0
ξ, $\langle\xi\rangle$	normalized distance and average distance in the inversion layer
ϱ	space-charge density
τ	average collision time
φ	normalized potential
$\Delta\chi$	electron-affinity difference at an interface
$\Psi(x,y,z)$, $\psi(z)$	wave functions of charge carriers in the surface layer
ω_c	cyclotron frequency

REFERENCES

[1]) J. R. Schrieffer, in R. H. Kingston (ed)., Semiconductor surface physics, Univ. Pennsylvania Press, Philadelphia, 1957, p. 55.
[2]) S. R. Hofstein and F. P. Heiman, Proc. I.E.E.E. **51**, 1190-1202, 1963.
[3]) C. N. Berglund and R. J. Strain, Bell Sys. tech. J. **51**, 655-703, 1972.
[4]) A. Kobayashi, Z. Oda, S. Kawaji, H. Arata and K. Sugiyama, J. Phys. Chem. Solids **14**, 37-42, 1960.
[5]) P. Handler and E. Eisenhour, Surface Sci. **2**, 64-74, 1964.
[6]) N. St. J. Murphy, Surface Sci. **2**, 86-92, 1964.
[7]) F. F. Fang and S. Triebwasser, Appl. Phys. Letters **4**, 145-147, 1964.
[8]) D. Colman, R. T. Bate and J. P. Mize, J. appl. Phys. **39**, 1923-1931, 1968.
[9]) A. P. Gnädinger and H. E. Talley, Sol. St. Electr. **13**, 1301-1309, 1970.
[10]) K. Ohta, Jap. J. appl. Phys. **10**, 850-863, 1971.
[11]) See for instance: M. Abramowitz and I. A. Stegun, Handbook of mathematical functions, Dover Publ., New York, 1965, p. 466.
[12]) F. Stern and W. E. Howard, Phys. Rev. **163**, 816-835, 1967.
[13]) See for instance: A. Messiah, Quantum mechanics, North Holland Publ. Comp., Amsterdam, 1964, pp. 773-781.
[14]) F. F. Fang and W. E. Howard, Phys. Rev. Letters **16**, 797-799, 1966.
[15]) E. D. Siggia and P. C. Kwok, Phys. Rev. **B2**, 1024-1036, 1970.
[16]) F. Stern, J. Comput. Phys. **6**, 56-67, 1970.
[17]) F. Stern, Phys. Rev. **B5**, 4891-4899, 1972.
[18]) C. B. Duke, Phys. Rev. **159**, 632-644, 1967.
[19]) D. J. BenDaniel and C. B. Duke, Phys. Rev. **160**, 679-685, 1967.
[20]) C. B. Duke, Phys. Letters **24A**, 461-463, 1967.
[21]) C. B. Duke, Phys. Rev. **168**, 816-831, 1968.
[22]) M. E. Alferieff and C. B. Duke, Phys. Rev. **168**, 832-842, 1968.
[23]) J. A. Appelbaum and G. A. Baraff, Phys. Rev. **B4**, 1235-1245, 1971.
[24]) J. A. Appelbaum and G. A. Baraff, Phys. Rev. **B4**, 1246-1250, 1971.
[25]) G. A. Baraff and J. A. Appelbaum, Phys. Rev. **B5**, 475-496, 1972.
[26]) J. C. Hensel and G. Feher, Phys. Rev. **129**, 1041-1062, 1963.
[27]) T. Sato, Y. Takeishi and H. Hara, Jap. J. appl. Phys. **8**, 588-598, 1969.
[28]) H. Maeda, J. Jap. Soc. appl. Phys. Suppl. **39**, 185-191, 1970.
[29]) V. N. Dobrovolskii and Yu. S. Zharkikh, Phys. Stat. sol. (a) **6**, 655-663, 1971.
[30]) A. Amith, J. Phys. Chem. Solids **14**, 271-290, 1960.
[31]) F. Stern, Proc. 10th Int. Conf. Phys. Sem. (Cambridge, Massachusetts, 1970), pp. 451-458.
[32]) B. A. Tavger and M. Sh. Erukhimov, Sov. Phys. JETP **24**, 354-359, 1967.
[33]) V. B. Sandomiskii, Sov. Phys. JETP **25**, 101-106, 1967.
[34]) M. Sh. Erukhimov and B. A. Tavger, Sov. Phys. JETP **26**, 560-565, 1968.
[35]) V. J. Ryzhii, Sov. Phys. sol. St. **10**, 2286-2287, 1969.
[36]) B. A. Tavger and V. Ya. Demikhovskii, Sov. Phys. Uspekti **11**, 644-658, 1969.
[37]) A. A. Romanov, L. J. Magarill and V. S. Sardaryan, Sov. Phys. Sem. **4**, 1072-1075, 1971.
[38]) M. D. Blokh, V. A. Margulis and B. A. Tavger, Sov. Phys. Sem. **5**, 1177-1181, 1972.
[39]) J. R. Schrieffer, Phys. Rev. **97**, 641-646, 1955.
[40]) R. Seiwatz and M. Green, J. appl. Phys. **29**, 1034-1040, 1958.
[41]) D. R. Frankl, J. appl. Phys. **31**, 1752-1754, 1960.
[42]) R. F. Greene, J. Phys. Chem. Solids **14**, 291-298, 1960.
[43]) R. F. Greene, D. R. Frankl and J. Zemel, Phys. Rev. **118**, 967-975, 1960.
[44]) S. Kawaji and Y. Kawaguchi, J. phys. Soc. Jap. Suppl. **21**, 336-339, 1966.
[45]) F. Stern, Phys. Rev. Letters **18**, 546-548, 1967.
[46]) S. Kawaji, J. phys. Soc. Jap. **27**, 906-908, 1969.
[47]) H. Ezawa, S. Kawaji, T. Kuroda and K. Nakamura, Surface Sci. **24**, 654-658, 1971.
[48]) H. Ezawa, T. Kuroda and K. Nakamura, Surface Sci. **24**, 659-662, 1971.
[49]) A. V. Serebryakov, Sov. Phys. sol. St. **11**, 781-783, 1969.
[50]) E. M. Baskin and M. V. Entin, Sov. Phys. JETP **30**, 252-256, 1970.
[51]) A. B. Fowler, F. F. Fang, W. E. Howard and P. J. Stiles, Phys. Rev. Letters **16**, 901-903, 1966.
[52]) A. B. Fowler, F. F. Fang, W. E. Howard and P. J. Stiles, J. phys. Soc. Jap. Suppl. **21**, 331, 1966.
[53]) F. F. Fang and P. J. Stiles, Phys. Rev. **174**, 823-828, 1968.
[54]) F. F. Fang and A. B. Fowler, J. appl. Phys. **41**, 1825-1831, 1970.

[55]) T. Ando, Y. Matsumoto, Y. Uemura, M. Kobayashi and K. F. Komatsubara, J. phys. Soc. Jap. **32**, 859, 1972.
[56]) Y. Katayama, N. Kotera and K. F. Komatsubara, Proc. 10th Int. Conf. Phys. Sem. (Cambridge, Massachusetts, 1970), pp. 463-468.
[57]) N. Kotera, Y. Katayama and K. F. Komatsubara, Phys. Rev. **B5**, 3065-3078, 1972.
[58]) K. von Klitzing and G. Landwehr, Sol. St. Comm. **9**, 2201-2205, 1971.
[59]) M. Kaplit and J. N. Zemel, Phys. Rev. Letters **21**, 212-215, 1968.
[60]) M. Kaplit and J. N. Zemel, Surface Sci. **13**, 17-29, 1969.
[61]) A. F. Tasch, D. D. Buss, R. T. Bate and B. H. Breazeale, Proc. 10th Int. Conf. Phys. Sem. (Cambridge, Massachusetts, 1970), pp. 458-463.
[62]) D. C. Tsui, Phys. Rev. Letters **24**, 303-305, 1970.
[63]) D. C. Tsui, Proc. 10th Int. Conf. Phys. Sem. (Cambridge, Massachusetts, 1970), pp. 468-477.
[64]) D. C. Tsui, Phys. Rev. **B4**, 4438-4449, 1971.
[65]) R. G. Wheeler and R. W. Ralston, Phys. Rev. Letters **27**, 925-928, 1971.
[66]) Y. Katayama, N. Kotera and K. F. Komatsubara, J. Jap. Soc. appl. Phys. Suppl. **40**, 214-218, 1971.
[67]) F. F. Fang and A. B. Fowler, Phys. Rev. **169**, 619-631, 1968.
[68]) A. Ohwada, H. Maeda and K. Tanaka, Jap. J. appl. Phys. **8**, 629-630, 1969.
[69]) T. Sato, Y. Takeishi, H. Hara and Y. Okamoto, Phys. Rev. **B4**, 1950-1960, 1971.
[70]) V. N. Dobrovol'skii, Yu. S. Zharkikh and L. N. Abissonova, Sov. Phys. Sem. **5**, 633-638, 1971.
[71]) A. A. Guzev, G. L. Kuryshev and S. P. Sinitza, Sov. Phys. Sem. **4**, 1755-1759, 1971.
[72]) O. A. Ribeiro and J. C. Pfister, Sol. St. Comm. **10**, 63-69, 1972.
[73]) G. Dorda, Appl. Phys. Letters **17**, 406-408, 1970.
[74]) G. Dorda, J. appl. Phys. **42**, 2053-2060, 1971.
[75]) R. Poirier, J. de Physique Suppl. **29**, 65-68, 1968.
[76]) S. Tansal, A. B. Fowler and R. F. Cotellessa, Phys. Rev. **178**, 1326-1327, 1969.
[77]) H. Sakaki and T. Sugano, Jap. J. appl. Phys. **10**, 1016-1027, 1971.
[78]) F. Stern, Phys. Rev. Letters **21**, 1687-1690, 1968.
[79]) Y. Uemura and Y. Matsumoto, J. Jap. Soc. appl. Phys. Suppl. **40**, 205-213, 1971.
[80]) S. Kawaji, H. Huff and H. C. Gatos, Surface Sci. **3**, 234-242, 1965.
[81]) S. Kawaji and H. C. Gatos, Surface Sci. **6**, 362-368, 1967.
[82]) S. Kawaji and H. C. Gatos, Surface Sci. **7**, 215-228, 1967.
[83]) D. C. Tsui, Sol. St. Comm. **9**, 1789-1792, 1971.
[84]) See for instance: L. B. Valdes, The physical theory of transistors, McGraw-Hill, New York, 1961, p. 245.
[85]) See for instance: A. Many, Y. Goldstein and N. B. Grover, Semiconductor surfaces, North. Holland Publ. Comp., Amsterdam, 1965, Ch. 4.
[86]) A. O. E. Animalu and V. Heine, Phil. Mag. **12**, 1249-1270, 1965.
[87]) M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789-796, 1966.
[88]) G. E. Pikus and G. L. Bir, Sov. Phys. sol. St. **1**, 1502-1517, 1960.
[89]) E. Spenke, Elektronische Halbleiter, Springer, Berlin, 1965, p. 295.
[90]) J. M. Ziman, Electrons and phonons, Clarendon Press, Oxford, 1960, p. 119.
[91]) See for instance: P. Henrici, Discrete variable methods in ordinary differential equations, Wiley, New York, 1962, Ch. 1.
[92]) See for instance: J. H. Wilkinson and C. Reinsch, Linear algebra, Springer, Berlin, 1971, pp. 227-240.
[93]) J. A. Pals, Phys. Letters **39A**, 101-102, 1972.
[94]) See for instance: A. Ralston, A first course in numerical analysis, McGraw-Hill, New York, 1965, Ch. 9.
[95]) J. A. Pals, Phys. Rev. **B5**, 4208-4210, 1972.
[96]) See for instance: R. S. C. Cobbold, Theory and applications of field-effect transistors, Wiley, New York, 1970.
[97]) See for instance: R. M. Warner and J. N. Fordemwalt, Integrated circuits, McGraw-Hill, New York, 1965.
[98]) P. E. Gray and R. B. Adler, I.E.E.E. Trans. **ED-12**, 475-477, 1965.
[99]) A. S. Grove, O. Leistiko and C. T. Sah, J. appl. Phys. **35**, 2695-2701, 1964.
[100]) J. M. Shannon, Sol. St. Electr. **14**, 1099-1106, 1971.
[101]) H. C. de Graaff and J. van Hengel, Internal report, Philips Res. Labs.
[102]) L. Goossens and Th. J. Tulp, Internal report, Philips Res. Labs.
[103]) P. V. Gray and D. M. Brown, Appl. Phys. Letters **8**, 31-33, 1966.

[104]) G. Abowitz, E. Arnold and J. Ladell, Phys. Rev. Letters **18**, 543-546, 1967.
[105]) E. Kooi, Philips Res. Repts **20**, 595-619, 1965.
[106]) A. G. Revesz, K. H. Zaininger and R. J. Evers, J. Phys. Chem. Solids **28**, 197-204, 1967.
[107]) E. H. Nicollian and A. Goetzberger, Bell. Sys. tech. J. **46**, 1055-1133, 1967.
[108]) B. E. Deal, M. Sklar, A. S. Grove and E. H. Snow, J. Electr. Chem. Soc. **114**, 266-274, 1967.
[109]) M. V. Whelan, Philips Res. Repts Suppl. **1970**, No. 6.
[110]) H. Deuling, E. Klausmann and A. Goetzberger, Sol. St. Electr. **15**, 559-571, 1972.
[111]) N. St. J. Murphy, F. Berz and I. Flinn, Sol. St. Electr. **12**, 775-786, 1969.

Summary

This thesis is concerned with a theoretical and experimental study of the two-dimensional electron or hole gas in an inversion or accumulation layer at the surface of a semiconductor. The electrons or holes in the surface layer are bound in a potential well to the surface, which causes a quantization of the motion perpendicular to the surface.

The equations for quantized carriers in a surface layer are given in the effective-mass approximation. For the motion perpendicular to the surface a Schrödinger equation has to be solved in a potential well which is partly influenced by the charge carriers in the surface layer themselves, via the space-charge density in the Poisson equation. We give a numerical solution method for these coupled equations. We show that at the electric quantum limit, with only the lowest energy level occupied, a general solution may be given by introducing dimensionless variables for inversion layers as well as for accumulation layers. The influence of effective-mass values, bulk dopant concentration and externally applied biases is only involved via the quantities used to make the variables dimensionless.

We further show that the electric quantum limit may be violated for high surface concentrations of carriers in the inversion or accumulation layer even at zero-absolute of temperature.

The variation in the average distance of the quantized carriers from the oxide–silicon interface in a metal–oxide–semiconductor system is determined with accurate gate-capacitance measurements. Measurements are made at 77 K and 300 K for silicon *n*- and *p*-type inversion layers and at 4·2 K for silicon *n*- and *p*-type accumulation layers. These measurements are in good agreement with the calculations based on a quantized motion perpendicular to the surface, an agreement which cannot be obtained with a "classical" continuum calculation in which the quantization is neglected. These measurements show that up to room temperature the quantization exerts a considerable influence.

A second group of measurements more directly illustrates the importance of the quantization for temperatures from about 25 K to 75 K. These measurements are based on anomalous behaviour of the derivative of the Fermi level at the surface with respect to the total charge in the inversion layer. This quantity is shown to have a dip at a sufficiently low temperature when, roughly speaking, the Fermi level passes the lowest quantized energy level for the motion perpendicular to the surface. We show that this derivative of the Fermi level can be experimentally measured by measuring the open voltage which occurs on the bulk contact of a metal–oxide–semiconductor transistor if a small a.c. voltage is applied between the gate and the short-circuited source and drain contacts. A good qualitative agreement between measurements and

calculations for devices with a {100} surface orientation is found for silicon *n*- and *p*-type inversion layers.

The much flatter dip in the measured curves for devices with {111}- and {110}-oriented surfaces, and for devices with a {100}-oriented surface after X-ray irradiation is ascribed to the broadening of the quantized energy levels by the charged centres in the oxide–semiconductor interface. The broadening which is needed to explain the measured curves is shown to be of the same order as can be expected on the basis of previous theoretical and experimental results.

Samenvatting

Dit proefschrift bevat de resultaten van een theoretisch en experimenteel onderzoek van een twee-dimensionaal elektronen- of gatengas in een inversie- of accumulatielaag aan het oppervlak van een halfgeleider. De elektronen of gaten in de oppervlaktelaag zijn gebonden aan het oppervlak in een potentiaalkuil, die een kwantisatie van de beweging loodrecht op het oppervlak veroorzaakt.

De vergelijkingen voor gekwantiseerde ladingsdragers in een oppervlaktelaag zijn gegeven met de effektieve massa benadering. Voor de beweging loodrecht op het oppervlak moet een Schrödingervergelijking worden opgelost met een potentiaalverloop, dat gedeeltelijk door de ladingsdragers zelf wordt beinvloed via een ruimteladingsbijdrage in de Poissonvergelijking. We geven een numerieke oplossingsmethode voor deze twee gekoppelde differentiaalvergelijkingen. We tonen aan dat in de elektrische kwantum limiet, wanneer alleen het onderste energieniveau voor de beweging loodrecht op het oppervlak is bezet, een algemene oplossing voor inversie- of accumulatielagen gegeven kan worden door dimensieloze grootheden in te voeren. De waarden van de effektieve massa's, de koncentratie van verontreinigingen in de halfgeleider en de aangelegde spanningen beinvloeden alleen de grootheden waarmee alle variabelen dimensieloos gemaakt zijn.

We tonen ook aan dat de elektrische kwantum limiet zelfs bij het absolute nulpunt van de temperatuur niet mag worden toegepast voor hoge oppervlaktekoncentraties van ladingsdragers in de inversie- of accumulatielaag.

De verandering van de gemiddelde afstand van de gekwantiseerde ladingsdragers tot het halfgeleider oppervlak van een metaal–oxide–halfgeleider struktuur is gemeten als functie van de aangelegde spanning op het metaal door de kapaciteit tussen het metaal en de inversielaag in de halfgeleider nauwkeurig te bepalen. De metingen zijn gedaan bij 77 K en 300 K voor silicium *n*- en *p*-type inversielagen en bij 4·2 K voor silicium *n*- en *p*-type accumulatielagen. Deze metingen stemmen goed overeen met de berekeningen waarbij de kwantisatie van de beweging loodrecht op het oppervlak in aanmerking is genomen. Deze overeenstemming bestaat niet met de berekeningen waarin de kwantisatie van de beweging loodrecht op het oppervlak wordt verwaarloosd. De metingen tonen aan dat de kwantisatie een belangrijke invloed uitoefent voor temperaturen tot aan kamertemperatuur.

Een tweede groep metingen toont de kwantisatie direkter aan in een temperatuurgebied van ongeveer 25–75 K. Deze metingen zijn gebaseerd op een anomaal gedrag van de afgeleide van het Fermi-niveau aan het oppervlak naar de totale lading in de inversielaag. We tonen aan dat deze grootheid een minimum vertoont bij voldoend lage temperatuur wanneer globaal gesproken het

Fermi-niveau het laagste energieniveau voor de beweging loodrecht op het oppervlak passeert. We laten zien dat deze grootheid experimenteel bepaald kan worden door de open wisselspanning op het kontakt met het halfgeleidersubstraat van een metaal–oxide–halfgeleider transistor te meten als functie van de gelijkspanning op het metaal met een aangelegde wisselspanning tussen het metaal en de kontakten met de inversielaag. We vinden een goede kwantitatieve overeenstemming tussen metingen en berekeningen voor silicium MOS transistoren met een {100} oppervlakte orientatie en met *n*- zowel als *p*-type inversielagen.

Een veel flauwer minimum in de gemeten krommen voor transistoren met een {111} of {110} orientatie en voor transistoren met een {100} orientatie na een Röntgenbestraling kan verklaard worden met het verbreden van de gekwantiseerde energieniveaus door de geladen centra in het oxide–halfgeleider oppervlak. De verbreding nodig om het flauwere minimum te kunnen verklaren komt overeen met al bekende theoretische en experimentele resultaten.

STELLINGEN

bij het proefschrift van

J. A. Pals

15 december 1972
T.H. Eindhoven

I

Bij hoge elektrische veldsterkte loodrecht op het oppervlak is de kwantisatie van de beweging van de ladingsdragers in een inversielaag van een halfgeleider ook bij kamertemperatuur nog merkbaar.

Hoofdstuk 3 van dit proefschrift.

II

De elektrische kwantum limiet benadering, die Uemura en Matsumoto toepassen in hun verklaring van de door Tansal e.a. verrichte magnetoweerstandsmetingen aan inversielagen, is niet toegestaan.

Y. Uemura, Y. Matsumoto, J. Jap. Soc. appl. Phys. Suppl. **40**, 205-213 (1971).

S. Tansal, A. B. Fowler, R. F. Cotellessa, Phys. Rev. **178**, 1326-1327 (1969).

III

De door Wheeler en Ralston uit fotogeleidingsmetingen aan silicium inversielagen bepaalde energieverschillen komen niet overeen met de theoretisch te verwachten waarden voor de door hen aangenomen overgangen tussen de niveaus van de oppervlaktekwantisatie.

R. G. Wheeler, R. W. Ralston, Phys. Rev. Letters **27**, 925-928 (1971).

IV

Het aannemen van een lineair potentiaalverloop in een inversielaag ook bij sterke inversie door Gnädinger en Talley leidt tot onjuiste conclusies.

A. P. Gnädinger, H. E. Talley, Sol. St. Electr. **13**, 1301-1309 (1970).

V

De vertragingstijd in de basis van een verzadigde bipolaire *n–p–n* transistor is niet zoals Poon e.a. aannemen evenredig met het kwadraat van de dikte van het effektieve basisgebied, omdat het profiel van de elektronenconcentratie niet lineair is.

H. C. Poon, H. K. Gummel, D. L. Scharfetter, IEEE Trans. **ED-16**, 455-457 (1969).

VI

De in een bepaald frekwentiegebied optredende negatieve weerstand bij een z.g. baritt-diode (barrier injection transit time) is meer een gevolg van de snelheidsmodulatie van de geïnjecteerde ladingsdragers, zoals bij een ruimteladingsbegrensde vacuumdiode, dan van injectiemodulatie door het wisselveld.

D. J. Coleman, J. appl. Phys. **43**, 1812-1818 (1972).

VII

Bij de berekening van de amplitude-modulatie ruis in Read-diode oscillaties verwaarlozen Johnson en Robinson ten onrechte de frekwentie- en gelijkstroomafhankelijkheid van de admittantie van het lawinegebied.

H. Johnson, B. B. Robinson, Proc. IEEE **59**, 1272-1273 (1971).

VIII

Het biedt voordelen om in de methode beschreven door Hart e.a. ter bepaling van het minimum ruisgetal de gepulste signaalbron te vervangen door een gepulste ruisbron.

P. A. H. Hart, J. A. Pals, J. Prins, C. A. G. Stricker, Philips Res. Repts **26**, 216-228 (1971).

IX

De verklaring van Boon van het gemeten asymmetrische Mössbauer doublet aan glazen met ingebouwde Fe^{2+} ionen is aan bedenkingen onderhevig.

J. A. Boon, Proefschrift Universiteit van Utrecht (1971), Hfdst. 8.

X

De niet-lineaire absorptie van korte laserpulsen in halfgeleiders kan in zijn algemeenheid niet worden beschreven met de door Ralston en Chang gebruikte tijdsonafhankelijke differentiaalvergelijking.

J. M. Ralston, R. K. Chang, Appl. Phys. Letters **15**, 164-166 (1969).

XI

Het is verontrustend dat de huidige technische ontwikkeling is gefundeerd in de autonomie van de mens, omdat zo de grond wordt ontnomen aan het bevrijdend perspektief dat Schuurman aangeeft.

E. Schuurman, Techniek en Toekomst, Van Gorcum, Assen (1972), Proefschrift V.U. Amsterdam, Hfdst. 4.

XII

De uitspraak, dat de wetenschap alle geloof verwerpt, is innerlijk tegenstrijdig en in wezen een belijdenis van een geloof in de wetenschap.

J. D. Fast, E.P.E. Revue Energie primaire **5**, 28-35 (1969).