Modeling and Realization of an Ultra-Short Channel MOSFET

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors
der Naturwissenschaften genehmigte Dissertation

vorgelegt von
Diplom-Physiker

Joachim Knoch, geb. Goebbels

aus Aachen

Berichter: Universitätsprofessor Dr. B. Lengeler
Universitätsprofessor Dr. H. Lüth

Tag der mündlichen Prüfing: 5.9.2001

Contents

1 Introduction

In 1943 Thomas J. Watson - chairman of IBM at that time - stated the often quoted sentence 'I think there is a world market for maybe five computers'. That was in the time before the invention of the transistor based on semiconducting material. This sentence has continuously been proven wrong in the last three decades and today the semiconductor and computer industry is a prospering market with a volume of sales of 200 billion US $.

For the last 30 years CMOS has been the predominant technology of microelectronics. The metal-oxide-semiconductor field-effect transistor (MOSFET) is at its heart and can nowadays be found millionfold in every computer chip. Over the years a tremendous down-scaling has been taken place resulting in a minimum feature size of 130 nm of today's semiconductor devices. The question arises: Why is scaling important? The answer is that scaling the geometry of the device and the supply voltage by a factor $1/\kappa$ leads to a higher density of devices $\propto \kappa^2$ and an increased speed $\propto \kappa$ while leaving the power consumption constant. The ever increasing demand for higher speed and performance of microelectronic circuits has forced this enormous miniaturization. The remarkable commercial success of CMOS technology and the unique features of the Si/SiO_2 material system explain the enormous efforts of the semiconductor industry to push the limits to ever smaller structures whilst still the same principles of operation are used as at the beginning of the era of integrated circuits.

In 1965 G. Moore observed that integration of electronic circuits proceeds exponentially resulting in a doubling of components per chip every 18 months [48]. This so-called Moore's law still holds today and has led to the set up of technology roadmaps which are simple extrapolations of the development experienced so far. According to these roadmaps the limits of CMOS will be reached in the near future. Figure 1.1 shows a compilation of the International Technology Roadmaps for Semiconductors[1] for the years 1994 - 1999. As is predicted by the 1999 Technology Roadmap, the regime of 30 nm feature size is reached in 2015. To what extend CMOS can be further extrapolated and actually realized is quite unclear. A lot of theoretical and experimental research worldwide is dedicated to answer the question of 'how small can we go?'. Theoretical studies on the limits of CMOS have continuously proven wrong by experimentalists in recent years. Devices with channel lengths as short as 15 nm have already been demonstrated successfully [38] which exhibit acceptable electrical characteristics. In particular, the advent of silicon-on-insulator technology which will be the standard material for next generation devices makes a further miniaturiza-

[1] http://www.semichips.org

tion of today's devices possible [53]. Very recent simulations have shown that the limits of CMOS based on SOI technology are around channel lengths of 8 nm [50] for an ideal layout of the MOSFET structure.

For the realization of aggressively scaled devices there are two major obstacles. First, the technological difficulties, especially the nanometer-scale lithography, have to be overcome. Today, several different technologies are investigated which shall replace optical lithography in the near future. In particular, the use of extreme-UV, ion-beam and X-ray lithography are promising candidates. The second is a fundamental physical limit. There is a minimum structure size for which CMOS works. Beyond that, completely different concepts have to be employed. For ever smaller devices quantum mechanical effects will become increasingly important and will strongly influence the performance of MOSFETs.

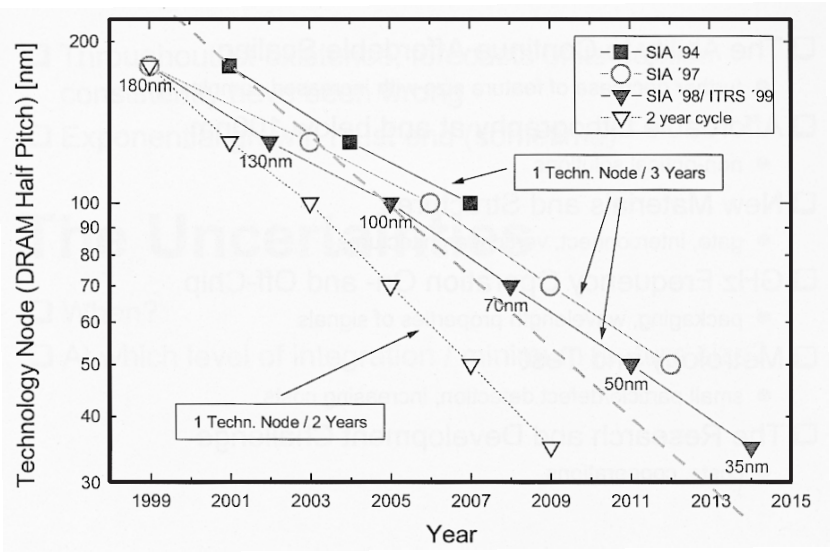

Figure 1.1: Compilation of the 1994 - 1999 International Technology Roadmaps for Semiconductors.

The present thesis studies ultra-short channel MOSFETs theoretically and experimentally. A fully quantum mechanical computer simulation of a quasi two-dimensional model for a single-gated ultra-short channel MOSFET has been developed. Simulations are performed in order to study the influence of the geometry of such devices on their electrical behavior. In addition, a technology for the fabrication of ultra-short channel MOSFETs with defined channel lengths is developed. Minimum channel lengths of 10 nm and below can be realized in principle what allows an investigation of the limits of CMOS technology. The key features of the present approach are the generation of an abrupt doping profile with the use of silicon epitaxy on SOI and the self-limited generation of source and drain with anisotropic wet chemical etching. This

results in two major advantages of the present device concept. First, ideally shallow source/drain contacts are achieved what is necessary to prevent short-channel effects. Second, a reproducible process technology is obtained due to the self-limiting behavior. In addition, because of anisotropically etching the device structure the technology is capable of fabricating source-drain separations in the nanometer regime with conventional optical lithography. Devices fabricated with this technology show state-of-the-art electrical characteristics.

The results presented here emerged out of a collaboration with the IBM T.J. Watson Research Center, USA, Massachusetts Institute of Technology, USA and the University of California at Los Angeles, USA.

2 Principles of Operation

The present chapter gives an introduction to the most relevant topics and concepts of a metal-oxide-semiconductor field-effect transistor (MOSFET). Beginning with the metal-oxide-semiconductor capacitor, the MOSFET will be introduced in the following. First, a long-channel device is examined, then the peculiarities of short-channel devices will be introduced and discussed.

2.1 The MOS-Capacitor

As its name reveals the MOS capacitor consists of three components, i.e. a metal, an insulator (oxide) and a semiconductor. In equilibrium, the edges of the conduction and valence bands are aligned according to the requirement of a common FERMI level throughout the structure. In addition, at interfaces between two components the exact position of the band edges is determined by the work-function of adjacent materials by the requirement of an equal energetic position of the vacuum level E_{vac} at the respective interface. This is illustrated in figure 2.1 (a) for the case of tungsten as metal electrode and a p-doped silicon substrate. Note, that the energetic position of the band edges is not drawn to scale.

In non-degenerately p-doped silicon the FERMI level E_f deep in the semiconductor side is determined by the doping concentration N_a and the temperature T according to [76]

$$E_i - E_f = kT \ln\left(\frac{N_a}{n_i}\right) \tag{2.1}$$

where n_i is the intrinsic carrier concentration and E_i can be considered as being the midpoint between valence and conduction band edge. The difference between the FERMI and the intrinsic level E_i is denoted ψ_b as is depicted in figure 2.1 (b) at the right end of the p-silicon. If the silicon is degenerately doped the FERMI level lies above the conduction band edge for n-type and below the valence band edge for p-type silicon.

If a voltage V_g is applied between the metal electrode and the semiconductor a potential ψ_s is present at the semiconductor-insulator interface which alters the charge distribution at the surface. In the depletion approximation the relation between V_g and ψ_s is [76]

$$V_g = V_{fb} + \psi_s + V_{ox} = V_{fb} + \psi_s + \frac{-Q_s}{C_{ox}} = V_{fb} + \psi_s + \frac{\sqrt{2\varepsilon_{si} e N_a \psi_s}}{C_{ox}} \tag{2.2}$$

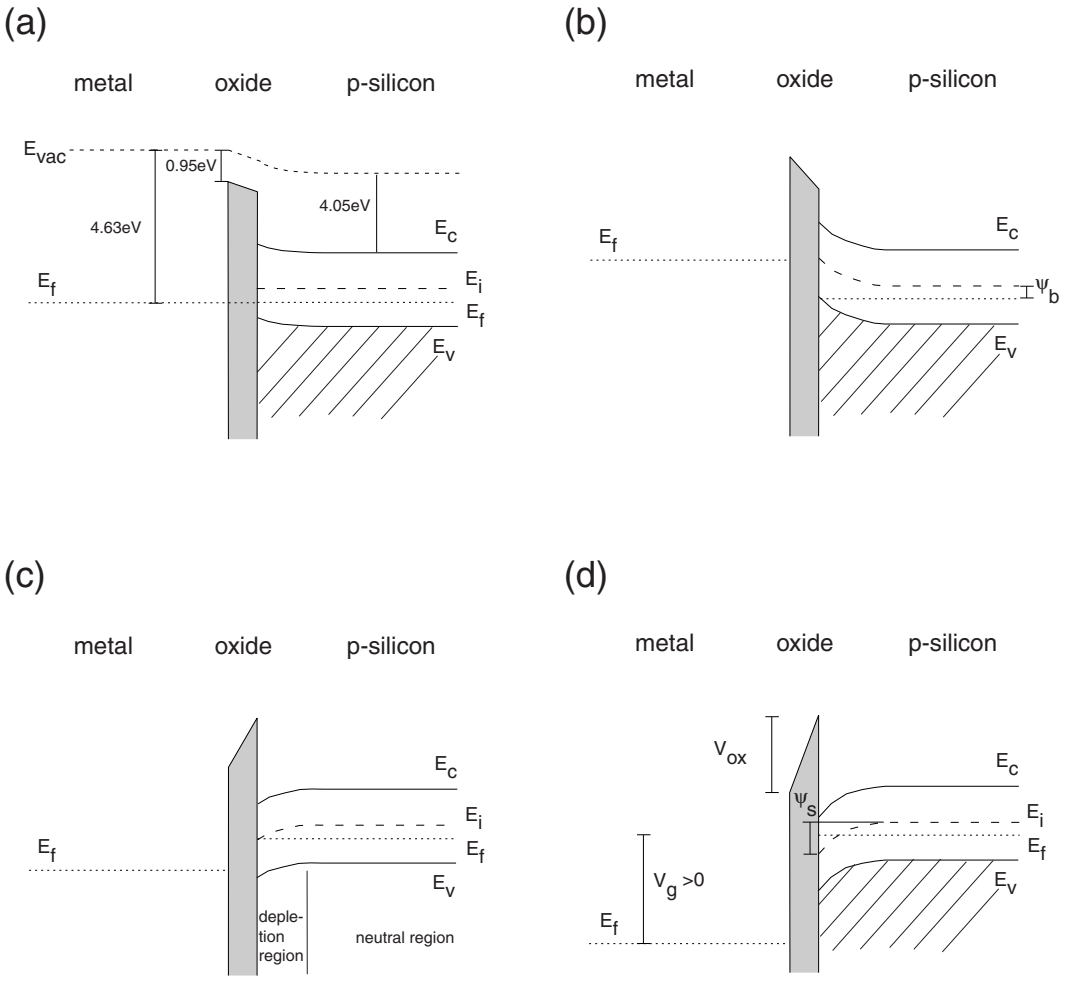

Figure 2.1: The MOS capacitor. (a) in equilibrium, (b) accumulation, (c) onset of inversion, and (d) strong inversion.

where N_a is the density of acceptors and e is the elementary charge. V_{fb} is the so-called flat-band voltage explained below. For $V_g \neq 0$ the following cases can be distinguished:

- In equilibrium the MOS system is determined by the requirement of a common FERMI level throughout the entire structure. The positions of the conduction band edges at the interfaces of the different materials are given by the differences in work function. The band-bending of the conduction and valence band of the semiconductor depends on the actual magnitude of the work-function of the metal.
- A negative voltage V_g accumulates holes at the oxide-semiconductor interface which are the majority carriers in the case of a p-silicon MOS capacitor. The MOS structure is said to be in accumulation.

- If in equilibrium the conduction and valence band edges of the p-silicon are bent a voltage applied between metal electrode and silicon can compensate this bending so that the case of flat band edges in the silicon is obtained. The corresponding voltage is called the flat-band voltage V_{fb} which can either be positive or negative depending on the work function of the gate electrode and the doping concentration in the silicon.

- If in the case considered in figure 2.1 a positive voltage larger than V_{fb} is applied the majority carriers in the p-doped region, i.e. the holes, will be repulsed from the surface. The resulting negative depletion charge of the immobile ionized acceptors will force the conduction and valence band to bend downwards. At the instant when $E_i = E_f$ at the surface an equal occupation of electrons and holes results at the insulator-semiconductor interface. Increasing V_g yields a higher occupancy of electrons, i.e. the semiconductor is inverted in a small area at the interface. The onset of weak inversion is depicted in figure 2.1 (c). This means that at the surface a small area - the inversion layer - builds up consisting of bulk minority carriers, i.e. electrons in the case considered here.

- A further increase of gate voltage leads to the case of strong inversion which is shown in figure 2.1 (d). The FERMI-level lies well above the intrinsic level. The surface potential does not change any more when altering the gate voltage because the inversion charge will screen the electric field generated by the gate voltage. In the case of strong inversion all the voltage drops across the insulator and the MOS system behaves like a capacitor.

  A criterion for the onset of strong inversion is $\psi_s = 2\psi_b = 2(E_i - E_f)$. The corresponding gate voltage V_g is called the threshold voltage V_{th} [76] which is given by equation (2.2) with ψ_s substituted by $2\psi_b$ as

$$V_{th} = V_{fb} + 2\psi_b + \frac{\sqrt{4\varepsilon_{si} e N_a \psi_b}}{C_{ox}}. \tag{2.3}$$

2.2 Long-Channel MOSFET

Attaching contacts to the right and to the left of the inversion layer one obtains a MOSFET structure a schematics of which is shown in figure 2.2. The device has three terminals, namely source, drain and gate and consists of two reversibly poled pn-junctions. All following considerations will exclusively be discussed for the case of an n-MOSFET, but can easily be transferred to the p-MOSFET. For an n-MOSFET the channel builds-up in the inversion layer of the p-doped part of the MOSFET which is in between the two n-doped contact regions. The channel is separated from the gate electrode by a thin SiO_2 film. The channel length L is measured along the x-direction, the width W of the transistor in the y-direction. Additionally, the circuitry with the accessory voltages V_g, the gate voltage, and the drain-source voltage V_{ds} are shown in figure 2.2 as well.

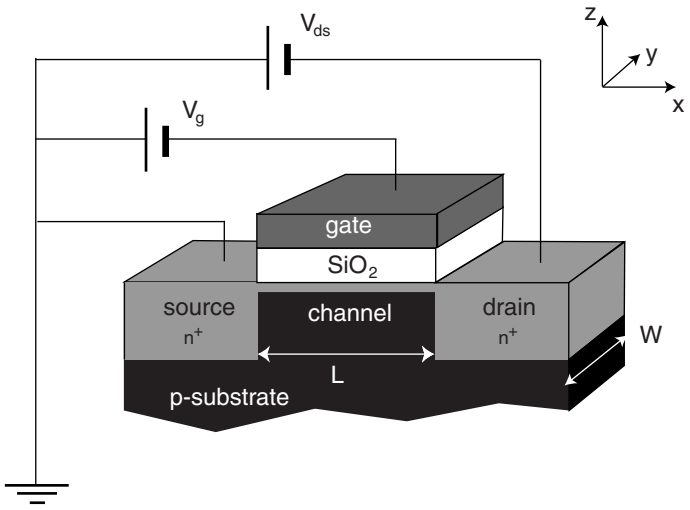

Figure 2.2: Sketch of the MOSFET with a typical circuitry and the accessory voltages.

Usually, all voltages are measured with reference to the source contact which is connected to ground.

For gate voltages $V_g \ll V_{th}$ the density of mobile carriers in the channel - i.e. electrons in the case of the n-MOSFET considered here - is exponentially small so that a drain current is prevented from flowing even when the drain-source bias V_{ds} is different from zero. Equivalently, one can consider the potential barrier generated by the two reversibly poled source-channel and channel-drain pn junctions. If this barrier is high enough no current flow is possible. In equilibrium, the barrier height is equal to the potential difference of the conduction band edges in e.g. source and the channel far away from the depletion zone of the pn junction. The resulting potential is called the built-in potential V_{bi} which is determined by the doping concentrations of the contacts and the channel. If a positive voltage is applied to the gate an inversion layer is generated connecting source and drain which is called the channel. A drain current I_d flows which can be controlled by the gate voltage.

In the case of the long-channel MOSFET the channel is much longer than the depletion zones of the pn junctions at the source and drain ends of the channel. Hence, in the case of a long-channel MOSFET these depletion zones can be neglected. The channel is so long that the potential distribution varies only slightly along the x-direction and the potential distribution can be regarded as one-dimensional. In this gradual-channel-approximation the behavior of the MOSFET is determined by the potential along the z-direction and the MOSFET can be described similar to the MOS capacitor.

2.2.1 I-V Characteristics

Figure 2.3 (a) shows a typical drain-current versus drain-source voltage plot of a long-channel MOSFET and (b) is a log-scale as well as a linear plot of the drain current I_d versus V_g for small drain-source voltages. Depending on V_g and V_{ds} one can distinguish between three different regions in the output characteristics:

- the subthreshold region: for $V_g < V_{th}$ the drain current exponentially depends on V_g since only electrons with energies higher than the conduction band edge in the channel contribute to the current. The occupancy of these energy levels is only exponentially small no matter whether MAXWELL-BOLTZMANN or FERMI-statistics is encountered.

- the linear region: for small V_{ds} the MOSFET behaves like a resistor whose resistance depends on V_g.

- the saturation region: for high V_{ds} and $V_g > V_{th}$ the drain current saturates. Its actual value solely depends on V_g.

The above listed modes of operation will now be discussed in more detail in the following paragraphs.

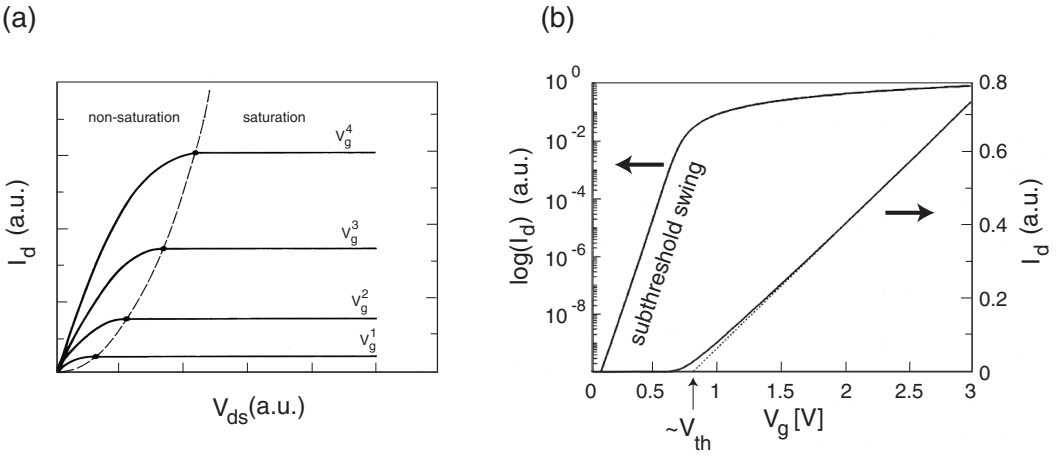

Figure 2.3: Output (a) and transfer (b) characteristics of a long-channel MOSFET.

Subthreshold Behavior

Below threshold the current is very small but not zero. In fact, it depends exponentially on the gate voltage. The left curve in figure 2.3 (b) shows a log-scale plot of the drain current versus gate voltage for small bias V_{ds}. The

straight line in the region below the threshold voltage is said to be the inverse subthreshold swing. It is defined as

$$S^{-1} = \frac{\partial \log I_d}{\partial V_g} \tag{2.4}$$

and indicates how the drain current alters when changing the gate voltage, i.e. it is a measure for the gate control of the device below threshold. Usually, the numerical value is given in mV/dec, i.e. how much gate voltage change is necessary in order to change I_d by one decade. Ideally, this change in gate voltage is given by $\Delta V_g \approx kT \ln 10$ and has a numerical value of $S = 60\,\mathrm{mV/dec}$ at room temperature. The subthreshold behavior is important for digital applications since it determines how the device can be switched on and off. An acceptable ratio between on- and off-state current I_{on}/I_{off} is about 4 - 5 orders of magnitude for V_g given by the supply voltage and $V_g = 0$, respectively. Since the supply voltage of future digital circuits must be reduced in order to diminish the power consumption S must not be larger than $\sim$100 mV/dec. Otherwise no defined off-state is obtained any more.

The Linear Region

For small drain source voltage V_{ds} the drain current I_d can be given by the following relation [76]

$$I_d = \mu_{eff} C_{ox} \frac{W}{L} (V_g - V_{th}) V_{ds} \tag{2.5}$$

where μ_{eff} is the effective mobility of the conduction electrons and V_{th} is the threshold voltage. Equation (2.5) indicates that for small V_{ds} the MOSFET behaves like a resistor where the resistance $R_{ds}(V_g) = V_{ds}/I_d$ can be modulated by the gate voltage. The linear region of $I_d - V_{ds}$ curves can be seen in figure 2.3 (a) in the left part indexed 'non-saturation'.

More important for the extraction of device parameters is the $I_d - V_g$ plot shown in figure 2.3 (b). Linear $I_d - V_g$ plots (right curve in fig. 2.3 (b)) at low drain bias are used to approximately determine the threshold voltage and to extract the effective channel length of a MOSFET. The threshold voltage V_{th} can be determined from the $I_d - V_g$ plot by linearly extrapolating the curve to $I_d = 0$. The intercept with the V_g-axis yields approximately V_{th} as is shown in figure 2.3 (b) by the dotted line.

MOSFET in Saturation

For increasing V_{ds} the drain current first raises linearly as stated above and reaches its maximum value at V_{ds}^{sat} which is shown by the dashed line in figure 2.3 (a). Current saturation occurs because at the drain end the inversion charge density vanishes if V_{ds} equals the gate overdrive $V_g - V_{th}$, i.e. the channel vanishes at the drain end which is termed pinch-off. Figure 2.4 (a) shows this

situation schematically. The saturation current depends on V_g as (dashed line in figure 2.3 (a))

$$I_d^{sat} = \mu_{eff} C_{ox} \frac{W}{L} \frac{(V_g - V_{th})^2}{2m} \tag{2.6}$$

with $m = 1+3t_{ox}/W_{dm}$ being the body-effect coefficient and W_{dm} the maximum depletion layer width [76]. A further increase of V_{ds} beyond the saturation voltage leaves I_d unchanged whilst the pinch-off point moves towards the source as shown in figure 2.4 (b).

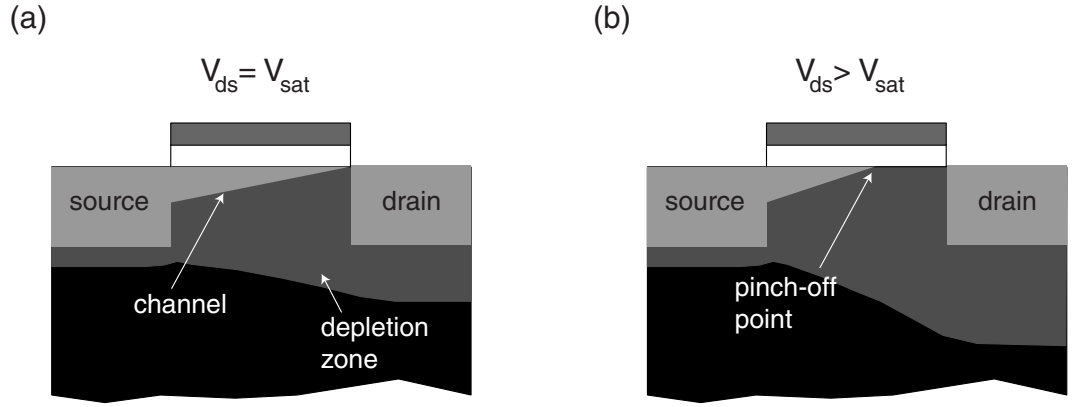

Figure 2.4: Current saturation in a MOSFET. (a) For $V_{ds} = V_{sat}$ the channel vanishes at the drain end. (b) The pinch-off point moves to the source for $V_{ds} > V_{sat}$.

A plot of I_d versus V_g in the saturation region and for $V_g \gg V_{th}$ yields the so-called transconductance g_m of the MOSFET defined according to

$$g_m = \frac{1}{W} \left. \frac{\partial I_d}{\partial V_g} \right|_{V_{ds}>V_{sat}, V_g \gg V_{th}} . \tag{2.7}$$

The transconductance is a measure of how effective the drain current in the on-state of the device is controlled by the gate voltage.

2.2.2 Channel Length

The channel length is a key-parameter in CMOS technology since it is used for circuit simulations, short-channel design etc. However, the notion 'channel length' is not well-defined. There are quite a few length scales all related to the spatial extension of the channel. One distinguishes between the mask length L_{mask}, the gate length L_g, the metallurgical channel length L_{met} and the effective channel length L_{eff}. The difference between L_{mask} and L_g is simply due to an imperfect or deliberate deviation of the lithography mask pattern and the actual gate pattern. The metallurgical channel length denotes the separation

between source and drain, i.e. denotes the geometrical channel length. The effective channel length is different from the other lengths since it is defined by the electrical characteristics of the MOSFET and is no geometrical length. L_{eff} is a measure of how much gate-controlled current a device can deliver and is therefore an important length scale for circuit models. In general L_{eff} is different from L_{met} depending on the doping profile of source and drain. In lightly-doped-drain devices, for instance, $L_{eff} > L_{met}$ whereas for abrupt doping profiles it can be significantly shorter than L_{met} [76]. There are several methods to extract the effective channel length from electrical characteristics [62]. For short channel MOSFETs the 'shift and ratio' method introduced by Taur et al. is the most effective. This method compares the linear regime of the output characteristics of a long- and a short-channel device with equal width [74].

When devices get shorter and shorter it becomes increasingly difficult to exactly define the metallurgical channel length. As long as there is no abrupt doping profile the actual source-drain separation cannot be determined unambiguously since the annealing steps in the process flow of MOSFET fabrication will smear out the borders of implanted source/drain contacts due to diffusion of the dopants. Hence, when producing ultra-short channel MOSFETs diffusion of dopants can be a severe problem for the reproducibility of a certain technology.

2.3 Short-Channel MOSFET

Although the transition from a long-channel to a short-channel MOSFET is a gradual one there exist qualitative and quantitative differences between them. For shrinking channel length the depletion zones at the source-channel and channel-drain contacts become more and more a part of the actual channel and their influence cannot be neglected any more. The resulting potential distribution is two-dimensional rather than one-dimensional. The proximity of source and drain thereby modifies the effective potential distribution and a number of so-called short-channel effects (SCE) appear which deteriorate the performance of the MOSFET. These effects are discussed in more detail in the following.

2.3.1 I-V Characteristics and Short-Channel Behavior

As already mentioned, the depletion zones of source and drain influence the electrical behavior of short-channel MOSFETs which manifests itself in the appearance of short-channel-effects. The short-channel effects come about because of an increasing loss of control of the depletion charge by the gate and an increasing influence by the drain. This can be seen in the output and transfer characteristics of such devices. The drain current of MOSFETs suffering from SCE does not saturate any more which is shown in figure 2.5 (a); I_d increases significantly even in the saturation region. Figure 2.5 (b) shows the transfer

characteristics of devices with decreasing channel length for two different V_{ds} (adjacent straight and dashed lines belong to one device). The loss of gate control manifests itself in an increasing subthreshold swing S which moreover depends on V_{ds}. Physically there is no distinction between different SCE since they are all due to the influence of the depletion zones of source and drain on the barrier height of the channel when L is reduced. However, according to the regime of operation a number of SCE are specified.

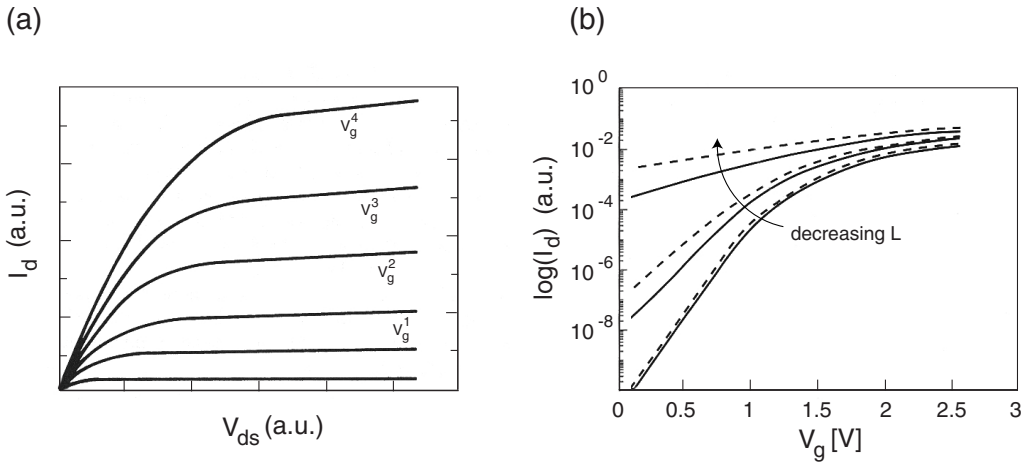

Figure 2.5: Output (a) and transfer (b) characteristics of a short-channel MOSFET. (b) shows curves with decreasing channel length for two different V_{ds}.

Drain-Induced-Barrier-Lowering

Due to the increasing extension of the drain depletion zone with increasing V_{ds} the barrier height at the edge of the source is lowered. This so-called drain-induced-barrier-lowering (DIBL) increases the carriers injected into the channel and the drain current raises in the saturation region. The situation is schematically shown in figure 2.6 contrasted to the situation in a device with long-channel behavior. The potential profile of the MOSFET is depicted for two different V_{ds}. For the long-channel device (a) the potential barrier at the source does not change under bias but depends on V_{ds} for the short-channel device (b). Therefore, the leakage current in the off-state of the device is increased with increasing bias. In the extreme case the device can be switched on under bias so that it cannot be used for digital applications because of the lack of a well defined on- and off-state. An equivalent approach to this phenomenon is that for a device suffering from SCE the depletion charge in the channel is influenced more and more by the drain electric fields as the channel length is decreased. This is referred to as charge sharing by the gate and drain electric fields.

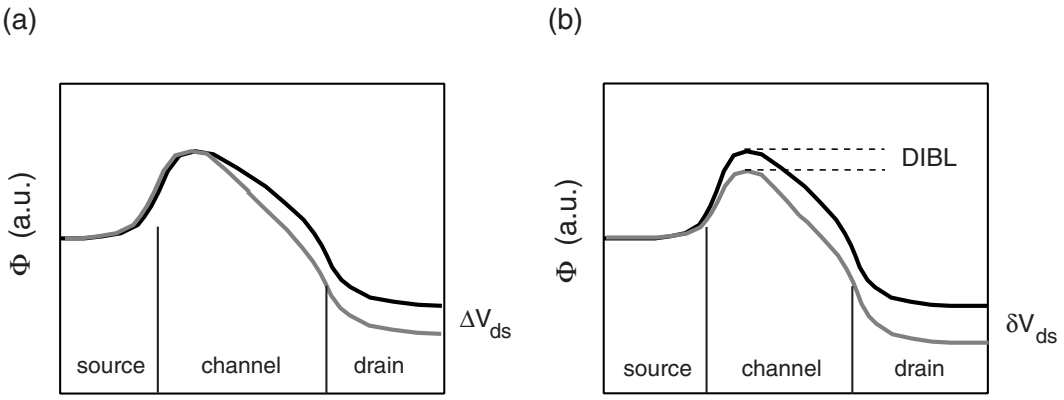

Figure 2.6: Potential profile in a long-channel (a) and a short-channel (b) MOSFET for two different V_{ds} illustrating the appearance of DIBL in (b).

Due to the barrier lowering the threshold voltage becomes bias dependent so that a measure for the DIBL is the shift of the threshold voltage V_{th} per incremental change of the bias V_{ds}, usually given in units of mV/V.

$$\mathrm{DIBL} = \left| \frac{\Delta V_{th}}{\Delta V_{ds}} \right| \tag{2.8}$$

Obviously, for an ideal long channel device the value of DIBL approaches zero.

Alternatively, looking at figure 2.6 (b) DIBL can be quantified by the ratio between the amount of barrier lowering indicated as 'DIBL' in the diagram per δV_{ds} for $V_g \approx 0\,\mathrm{V}$. This can certainly be done only for simulated curves where the potential profile is known. A practical way of extracting the DIBL is found by the observation that the barrier lowering essentially leads to a shift of the transfer characteristics to lower V_g [79]. A horizontal line is drawn in the transfer characteristics which crosses the curve for a medium V_{ds} at $V_g \approx 0$. The gate voltage interval where the horizontal line crosses transfer curves for higher and lower V_{ds} is then divided by the V_{ds}-interval of the belonging curves. This value represents the DIBL also specified as mV/V. However, this procedure results in higher numerical values (about a factor of 2) for the DIBL when compared to the direct relation possible for simulated curves.

Threshold-Voltage Roll-Off

As already mentioned above when L is of the order of the extension of the depletion zones in source and drain the potential barrier height is decreased depending on L. This results in a decrease of the threshold voltage compared to V_{th} of a long-channel device. This so-called threshold voltage roll-off is a severe problem in short-channel devices because manufacturing tolerances in the channel dimensions leads to significantly differing V_{th} from device to device

[80]. In addition, V_{th} must not be too small in order to get a defined off-state at $V_g = 0\,\mathrm{V}$. In order to overcome this problem the influence of source and drain depletion zones have to be diminished which can be done by increasing the acceptor doping level in the channel or decreasing the source/drain contact depths (ultra-shallow contacts). As will be discussed below another way is to use very thin silicon-on-insulator (SOI) films with which the fraction of the gate controlled depletion charge is increased [15]. Furthermore, quantization in ultra-thin SOI films in the direction of the film thickness can enhance the effective potential barrier height thereby increasing the threshold voltage [54].

Channel Length Modulation

The effective channel length is reduced by the extensions of the depletion zones at the source-channel and channel-drain contact. For increasing V_{ds} the depletion zone at the drain end enlarges so that the effective channel length is diminished and becomes a function of V_{ds}. The channel-length modulation ΔL is actually not a real short-channel effect since it also occurs in the long-channel MOSFET. But in the case of a long-channel device it can be neglected because $L \gg \Delta L$. However, in the case of a short channel the channel length modulation is of the order of L and influences the electrical behavior of the device. With increasing bias the channel gets shorter and shorter which leads to an increase of current in the saturation region. For high bias the depletion zones can get together which in the extreme case results in the so-called punch-through. A high drain current then flows in the MOSFET which cannot be affected by the gate any more.

2.3.2 Scaling Laws

As mentioned above, when a MOSFET is scaled down the depletion width of source and drain have to be made small which can be done by increasing the doping concentration N_a in the channel. However, this increases the threshold voltage (see equation (2.3)) which in turn reduces the drivability of a MOSFET for given supply voltage. This in turn can be overcome by reducing the oxide thickness. Hence, scaling of a MOSFET is possible if certain rules are obeyed in order to avoid SCE and corruption of device performance. Dennard et al. were the first who put forward such scaling rules [21]. All spatial dimensions and potentials are scaled by a common factor $1/\kappa$ and the doping concentration by κ. Since the electric fields in the scaled device remain the same as in the unscaled one this scaling rule is referred to as constant field scaling. The drawback of constant field scaling is that the diffusion potentials of the *pn*-junctions are not scaled so that the supply voltage cannot be scaled accordingly. Reducing the supply voltage is essential in highly integrated circuits in order to reduce the power consumption of the circuit [75] and to guarantee device reliability which is of concern due to increasing electric fields in the devices. Therefore, the scaling rules are generalized by the introduction of additional scaling factors in order to compensate the short-comings mentioned above. This approach is

called 'selective scaling' which has two spatial dimension scaling factors. One scales the gate length and the vertical dimensions of the device the other scales the width of the device as well as the wiring [20, 85]. However, the scaling rules loose their validity for channel lengths in the nanometer regime. Even for the highest doping levels there are only a few dopants in the channel which lead to fluctuating electrical behavior from device to device disabling digital applications [87]. Hence, for ultimately scaled devices doping of the channel cannot be tolerated. A special device layout must be employed in this case in order to prevent the SCE from making the device inoperative.

2.3.3 Bulk versus SOI

Instead of working with a usual silicon substrate the efforts nowadays aim at using silicon-on-insulator (SOI) technology. SOI offers advantages not only in the direct dielectric isolation of adjacent devices being part of a circuit on a chip but also for the actual device itself. This is due to the unique feature of having a silicon film of varying thickness t_{si} on SiO_2, called the buried oxide (BOX). Short-channel effects are strongly reduced in SOI-based MOSFETs. Additionally, the presence of the buried oxide reduces parasitic capacitances of the device such as the junction-to-substrate capacitance [15] which increases the response speed of such transistors.

The electrical behavior of a MOSFET fabricated on the basis of SOI technology depends strongly on the thickness t_{si} [15]. For thick films ($\geq 300\,\mathrm{nm}$) the devices are very much like bulk MOSFETs. Reducing t_{si} results in a thin film device with superior output characteristics and short-channel behavior for scaled-down devices [15, 81]. In particular, the fully-depleted (FD) SOI MOSFET is the most promising version of possible SOI devices. Fully-depleted means that t_{si} is smaller than the depletion length at threshold. When scaling MOSFETs down the depletion charge controlled by the gate decreases with decreasing device dimensions which leads to DIBL resulting in the threshold voltage roll-off and eventually in a punch-through. SOI devices on the other hand are much less vulnerable to short-channel effects. This can be understood with a simple geometrical construction [15, 81]. The charge controlled by the gate can be estimated by a trapezoidal region underneath the gate with the upper base given by the gate length and the lower reduced due to the influence of source and drain. If the channel length is scaled down the lower base reduces until it vanishes. If this happens most of the depletion charge is not controlled by the gate anymore but by source and drain and the device suffers from short-channel effects. This scenario is depicted in figure 2.7 (a) for a bulk MOSFET. The depletion region is shaded dark-gray and the part of the depletion charge controlled by the gate is the hatched area. Using SOI as shown in figure 2.7 (b), the relative amount of the gate-controlled depletion charge can be increased. This amount is maintained if the silicon film thickness is reduced accordingly when scaling the channel length down [72]. Therefore, the DIBL and the threshold-voltage roll-off in these devices is much smaller and starts at much smaller channel lengths than in bulk devices [81, 72] which is

due to the more efficient gate control of the depletion charge and the automatic ultra-shallow source/drain contacts.

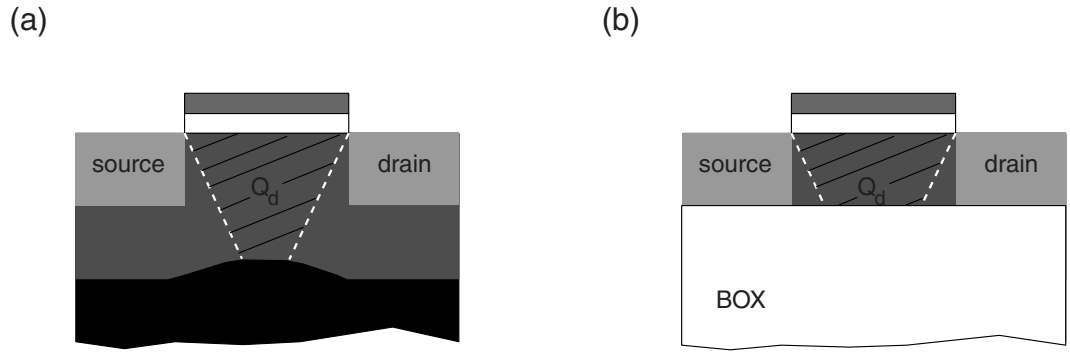

Figure 2.7: Bulk (a) versus SOI (b). The dark-gray shaded region is the total depletion charge whereas the hatched area is the depletion charge Q_d controlled by the gate [15].

If the SOI film thickness is further reduced quantum mechanical effects become important so that the quasi-continuum of two-dimensional subbands in the channel splits up into energetically separated levels [44]. Quantization plays an important role in very thin film SOI devices because the electron wave-function is confined to the SOI film by the two SiO_2-potential barriers represented by the gate oxide and the BOX [35]. In particular the first two-dimensional subband in the channel can be energetically well separated from the conduction band edge which significantly enhances the effective barrier height between source and drain [14, 23]. Since the first subband belongs to the ladder with high effective mass [5] in vertical direction quantization is of concern only for very thin t_{si} [44]. For ultra-short channel MOSFETs with undoped channel (see below) the increase in barrier height is important in order to avoid punch-through in these devices. However, quantization can degrade the performance of MOSFETs due to a reduction of the density of states in the channel and therefore a reduction of the inversion charge. Furthermore, the charge centroid is moved away from the Si/SiO_2 interface which results in a reduction of inversion capacitance. In addition, the threshold voltage increases due to the quantization [54] so that for low supply voltage the gate overdrive $V_g - V_{th}$ is reduced. This results in a performance loss concerning the drivability of the device if the SOI thickness is smaller than $\sim 3\,\text{nm}$ [12].

2.4 The Ideal Transistor Structure

2.4.1 Ballistic Transport

If the channel length is reduced to lengths in the range of 10-20 nm even for the highest doping levels there is only a small number of dopants in the channel

and every device will have its own specific spatial dopant configuration. This results in a fluctuation of threshold voltage from device to device which severely limits the applicability of MOSFETs in circuits [87, 46]. Increasing the width of the devices is no remedy for V_{th} fluctuations since in the subthreshold regime the current appears to be highly localized to the area of lowest potential barrier which can be different from device to device [12]. Therefore, one has to avoid doping in the channel and a special design is necessary to prevent the short-channel effects from deteriorating the performance of the transistor.

If the channel is undoped it was shown theoretically that the total mean free path of carriers in the channel limited by electron-electron and electron phonon-interaction [24] is such that ballistic transport, i.e. transport without scattering is possible for channel lengths smaller than 30 nm [28]. Since lowering the temperature suppresses both interaction processes ballistic transport is predicted at 77 K for $L \leq 50\,\mathrm{nm}$ [49]. Ballistic transport is favorable since the highest transconductances are anticipated for these devices.

Although electrons in ballistic MOSFETs move dissipationless the resistance of the MOSFET is not zero. There is a fundamental contact resistance which cannot be lowered no matter what contact metallurgy is employed [19, 8]. This resistance is due to the fact that the current inside the contacts is carried by infinitely many transverse modes (i.e. along the width of the device) whereas in the channel there are only a finite number of modes accessible [17]. This contact resistance is approximately $R_c = 12.9\,\mathrm{k}\Omega/M$ with M being the number of transverse modes in the channel.

The consideration of current saturation in ballistic MOSFETs is different from the notion of pinch-off in long-channel MOSFETs. The drain current I_d can be written as [49]

$$I_d \propto \int dE\,(f_s - f_d)T(E) \tag{2.9}$$

where $T(E)$ is transmission coefficient of electrons moving from source to drain. When direct tunneling between source and drain is neglected $T(E) \approx 1$ for energies above the potential barrier maximum E_{pot}^{max} represented by the source-channel *pn*-junction. In equilibrium both occupation factors are equal and cancel each other. For high bias f_d vanishes in the energy range for which $T(E) \neq 0$ and a net current flows which is given by the factor f_s for $E \geq E_{pot}^{max}$. Hence, if tunneling between source and drain can be neglected transport in ballistic MOSFETs is determined by the thermionic emission over the potential barrier generated by the two reversibly poled source-channel and channel-drain *pn* junctions.

2.4.2 Double-Gate MOSFET

There appears to be a consensus about the optimal design of an ultra-short channel device [75, 85] which is the double-gated MOSFET on SOI. Many theoretical studies have revealed its superior electrical characteristics [84, 50].

The two gates effectively terminate the drain electric field lines which results in an efficient suppression of short-channel effects [75]. There is an abrupt doping profile between contacts and the channel. The device structure of the double-gated MOSFET is depicted in figure 2.8. It consists of a very thin silicon film provided with top and back gate separated by very thin gate-oxides from the channel. Source and drain are degenerately doped ($\sim 10^{20}\,\mathrm{cm}^{-3}$) in order to keep the depletion zones and series resistances of the contacts as small as possible. The silicon thickness of the SOI film t_{si} is in the nanometer range. The reason for the superior short-channel behavior of the double-gated SOI MOSFET is the good control of the inversion charge by the gates. In addition, ideal ultra-shallow contacts are obtained automatically due to the geometry of the double-gate design [72].

As a rule of thumb for designing such a device, the ratio between gate length and silicon film thickness should not be smaller than about four in order to avoid punch-through [86]. Theoretical investigations of ultimately scaled double-

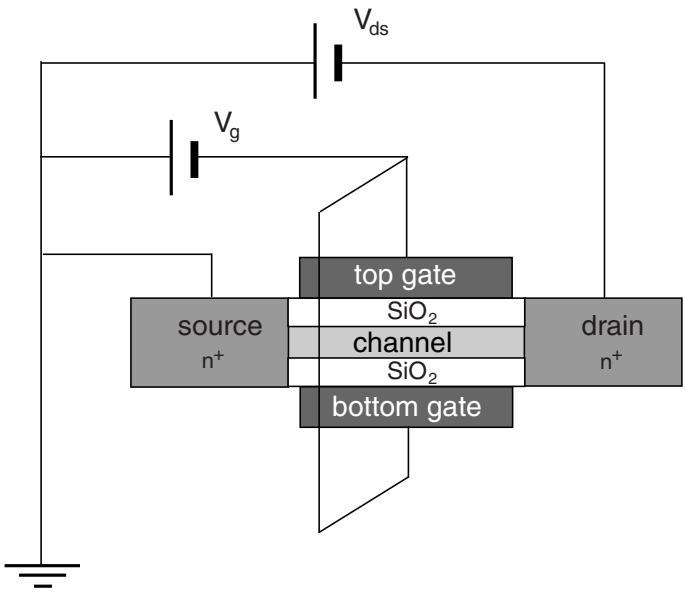

Figure 2.8: Ideal short-channel transistor structure. An ultra-thin undoped silicon layer is provided with top- and bottom-gate. Source and drain are degenerately doped with an abrupt doping profile between channel and contacts.

gated MOSFETs prognose functioning devices with channel lengths down to 8 nm for an SOI thickness of 2 nm and oxide thickness of 1 .. 1.5 nm [50]. Large transconductances up to 4000 mS/mm are predicted.

However, such double-gated MOSFETs are very difficult to realize since both gates have to be fabricated aligned to each other, surrounding a silicon layer only a few nanometer in thickness.

3 The V-Groove MOSFET

In recent years there have been proposed and realized many different concepts for the fabrication of an ultra-short-channel MOSFET [34, 36, 37, 77]. Channel lengths in the range of 30 nm down to 15 nm have been reported in the literature [13, 33, 38, 83]. With the use of electrically variable shallow source/drain junctions a MOSFET with 8 nm gate length has even been successfully fabricated [36]. However, this device needs two different gates in order to control the channel and the leads separately and suffers from very high source/drain series resistances. Most of the device concepts rely on the use of thin film SOI material with its advantages for scaled-down transistors. In general, there are two different strategies which are pursued namely the lateral MOSFET as is usually done in today's fabrication processes and the vertical MOSFET where the MOS structure is tilted at an angle of 90°. Both approaches have their specific advantages. In the case of a vertical MOSFET the active channel area can be realized, for instance, with molecular beam epitaxy (MBE) so that very short channels with atomic precision can in principle be fabricated [47]. Furthermore, a double-gated MOSFET can be realized much more easily than in the case of the lateral structure. On the other hand, vertical MOSFETs are difficult to fabricate since the technology is rather involved and requires many process steps [47, 33]. A lateral MOSFET layout on SOI offers the ability to achieve very thin channel regions which allows to reduce the short-channel effects and to exploit the vertical quantization in order to enhance the potential barrier of the channel [14, 63]. Additionally, with the use of very thin SOI films ultra-shallow source/drain contacts are automatically achieved [72].

In the present chapter a concept for the realization of an ultra-short-channel MOSFET on SOI with lateral layout is introduced and discussed. The concept relies on a combination of silicon epitaxy and a self-limited etch process which defines the channel region and separates source and drain. The device layout has some major advantages which allow the fabrication of MOSFETs with well-defined channel lengths in the 10 nm regime.

3.1 Device Principles

Figure 3.1 shows a schematic cross section of the device idea. A highly Sb doped n^{++} silicon layer (thickness t_n) is epitaxially grown on a thin, undoped p^- silicon film of thickness t_{si} being part of a silicon-on-insulator (SOI) wafer. An appropriate etch anisotropically etches a 'V' shaped groove with opening L_0 into the n^{++} layer with the tip of the 'V' just touching the p^- region. The

separated n^{++} areas serve as source and drain and the channel forms in the p^- in between source and drain, i. e. in the tip region of the V-groove. A thin gate oxide separates the channel as well as the n^{++} regions from a metal gate electrode. Except in the V-groove region the anisotropic etch cuts through the whole n^{++}/p^--stack and isolates adjacent devices from each other. Metal

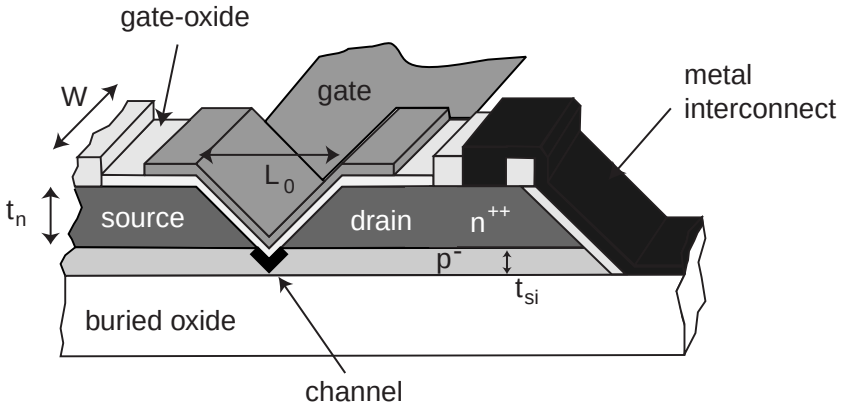

Figure 3.1: Sketch of the device geometry showing the V-groove obtained with anisotropic etching of an epitaxial silicon stack. The channel is located between the highly doped n^{++} regions.

interconnects are used in order to contact the devices. The advantages of this specific layout are the following:

- With the use of silicon epitaxy an abrupt doping profile is achieved. No implantation technique is necessary and the borders of the source and drain contacts are sharply defined by the geometry of the V-groove. Hence, a well-defined source-drain separation is obtained. The expression 'channel length' and 'source-drain separation' can therefore be used synonymically, referred to as L throughout the next chapters. Note, that raising source and drain on top of the channel layer with an abrupt doping profile is actually the optimal way of generating ultra-shallow contacts which is important in order to suppress SCE. Additionally, due to the high doping level the parasitic series resistance of source and drain is kept as small as possible.

- The channel layer is nominally undoped. This is necessary since for very small channel lengths the distinct configuration of dopant atoms influence the electrical behavior of the devices as already discussed in the previous chapter. Additionally, the p^- layer is made thin in order to reduce short-channel effects.

- Because the V-groove tapers off towards the channel a very short channel in the nanometer regime can be defined in the tip region of the groove by a much larger mask at its opening L_0. The required mask opening

L_0 is given by the geometry of the V-groove and depends only on the thickness t_n of the n^{++} layer. Hence, by choosing an appropriate t_n the V-groove process can be adjusted to any desired lithographic technique. As is shown in chapter 5, source-drain separations as short as 10 nm can be achieved with the present approach.

- The device layout inherently brings about a control of having fabricated a MOSFET with ultra-short channel. If the thickness of the SOI film is very thin only ultra-short-channel transistors can be obtained, because in case of the V-groove opening L_0 being too large the tip of the 'V' cuts through the p^- layer and no transistor action is possible. Thus, the geometry of the device gives an upper cut-off for possible channel lengths which is given by the relation $L_{max} = 2t_{Si}/\tan(\alpha)$. On the other hand, if L_0 is too small the V-groove does not cut through the entire n^{++} stack and source and drain will be shortened. Again, no transistor action is possible.

- A metal gate is chosen with the advantage that there is no depletion zone in the gate as it happens to be for poly-silicon and that a metal gate has a much lesser electrical resistance which is important for high frequency applications.

The present approach is attractive due to its conceptual simplicity. There are a lot of technological problems to be solved, though, which are set forth in chapter 5. Obviously, the generation of a proper V-groove is of central interest in the device concept. A self-limited, anisotropic etch procedure has to be developed in order to establish a reproducible and reliable technology.

4 Modeling of MOSFET Operation

Before going on to the realization of the proposed device concept, simulations of single-gated short-channel MOSFET on SOI will be performed in order to get information about the electrical behavior of such devices. Classical calculations have often failed to predict the electrical characteristics of ultimately scaled MOSFETs since quantum mechanical effects like the shift in V_{th} for ultrathin t_{si} play an important role for such devices. Therefore, a quantum mechanical simulation has been developed and will be presented here. Many publications dealing with the quantum mechanical simulation of ultra-small MOSFETs can be found in the literature. However, most of these papers incorporate only some quantum mechanical aspects in their simulations. For instance, either the enhancement of the effective barrier height in the channel due to vertical quantization [56] is considered or merely a $1D$ model perpendicular to the current flow is employed in order to describe the tunneling current from the gate to the channel [66, 30, 43]. Other works compute the $2D$ electron and current distributions relying on quasi-classical methods but neglect tunneling through the gate [50] or abandon self-consistency between the quantum and POISSON equations [49].

In this chapter a quasi two-dimensional model for an ultra-short-channel MOSFET is introduced and analyzed which fills the gap between the work mentioned above in that it combines a full quantum mechanical treatment of carrier transport with gate leakage in a self-consistent manner. The potential distribution in the direction of current flow will be reduced to a $1D$ problem making simulations using the non-equilibrium Green's function technique [18, 42, 16] feasible. Vertical quantization is taken into account through an effective barrier height in the channel. The use of the non-equilibrium Green's function formalism enables among other things the correct treatment of open boundaries for a finite computational domain which otherwise is often terminated by artificial boundary conditions [67, 71].

4.1 The Model

The device concept introduced in the previous chapter is a single-gated, fully-depleted MOSFET on SOI. Therefore, the model under investigation is chosen accordingly. In order to simplify the calculations the model is somewhat modified which is depicted in figure 4.1. A thin SOI film of thickness t_{si} serves as

channel of length L which is connected to semi-infinite source and drain contacts. The thickness t_{box} of the buried oxide is considered as being very large. The channel is assumed to be undoped whereas source and drain are degenerately doped with dopant concentration on the order of some $N_d \approx 10^{20}\,\mathrm{cm}^{-3}$. All dopants are fully ionized, i.e. they constitute a constant, positive background, called the jellium. The channel is seperated from a metal gate by a thin gate oxide (thickness t_{ox}), typically a few nanometer in thickness. Any influence of the gate on the n^{++} contacts is neglected. The width W of the device

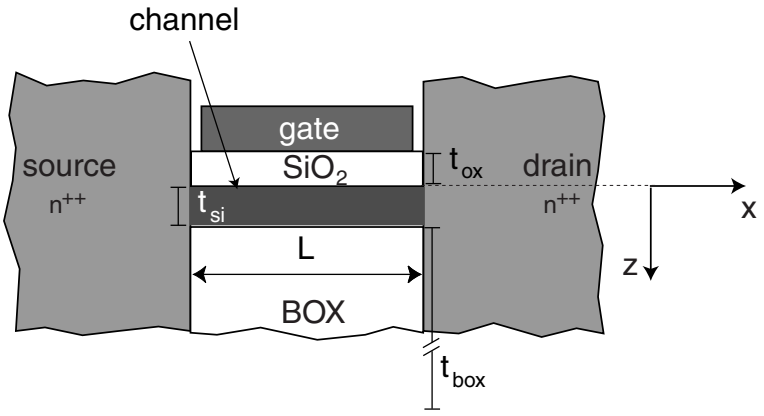

Figure 4.1: Model of the single-gated short-channel MOSFET on SOI used for the simulations. The channel length is aligned along the x-, the width along the y-direction.

is directed along the y-direction and is taken large enough so that the assumption of translational invariance along y is justified. As is shown in figure 4.1, x is directed along the channel length and z along the vertical direction, i.e. the SOI film thickness. Following Young [90, 89], the original two-dimensional potential distribution is reduced to an essentially one-dimensional one as stated in the next section.

4.1.1 Poisson's Equation

The potential distribution in the MOSFET is governed by a two-dimensional Poisson equation of the form

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)\Phi(x,z) = \frac{\rho(x,z)}{\varepsilon_{si}}. \tag{4.1}$$

In order to reduce the simulation problem to a one-dimensional calculation the so-called parabolic approximation for the potential distribution $\Phi(x,z)$ is made [88, 89, 90]

$$\Phi(x,z) \approx c_0(x) + c_1(x)z + c_2(x)z^2 \tag{4.2}$$

Equation (4.2) needs three boundary conditions in the vertical direction to be fulfilled in order to specify the 'constants' $c_i (i = 0..2)$ although there exist four

conditions namely the potential and electric field at both Si/SiO_2 and Si/BOX interfaces. According to reference [88] the exact value of the potential at the Si/BOX interface does not need to be specified since the potential essentially floats at this interface. It is assumed that the buried oxide is so thick that a finite potential difference across the BOX leads to a negligible field at the Si/BOX interface. Hence, the following boundary conditions determine the potential distribution in z-direction

- $\Phi(x,0) \equiv \Phi_f(x) = c_0(x)$, where Φ_f is called the surface potential
- the electric field at $z = 0$ is determined by the gate voltage V_g

$$\left.\frac{d\Phi(x,z)}{dz}\right|_{z=0} = \frac{\epsilon_{ox}}{\epsilon_{si}}\frac{\Phi_f(x) - V_g + V_{bi}}{t_{ox}} = c_1(x) \tag{4.3}$$

  with V_{bi} being the built-in potential. In the case of an undoped channel V_{bi} is given by $\Delta_{si}/2 + E_f$ where Δ_{si} is the energy gap of silicon and E_f is the FERMI energy relative to the conduction band edge deep inside the source/drain contact. Hence, V_{bi} is the maximum possible potential barrier height between source and drain. Doping the channel with acceptors would raise V_{bi} but is not considered here due to reasons mentioned in chapter 2. Furthermore, we have assumed a midgap workfunction gate metal such as tungsten so that $V_{fb} = 0$ for an undoped channel. However, using a metal gate with different work function yields an $V_{fb} \neq 0$ which would have to be added to V_{bi} in equation (4.3). The effect would simply be a shift in V_{th}.

- at the interface to the buried oxide the electric field is approximately zero for large t_{box}

$$\left.\frac{\Phi(x,z)}{dz}\right|_{z=t_{si}} = \frac{\epsilon_{ox}}{\epsilon_{si}}\frac{\Phi_{bs} - \Phi_b(x)}{t_{box}} = c_1(x) + 2t_{si}c_2(x) \approx 0 \tag{4.4}$$

Inserting this into the two-dimensional POISSON equation (4.1) and setting $z = 0$ one ends up with an expression for the surface potential $\Phi_f(x,z)$ in the channel. Thus, the following reduced equation for Φ_f is obtained [88, 56]

$$\frac{d^2\Phi_f(x)}{dx^2} - \frac{\epsilon_{ox}}{\epsilon_{si}}\frac{\Phi_f(x) - V_g + V_{bi}}{t_{si}t_{ox}} = \frac{\rho_{tot}(x)}{\epsilon_{si}} \tag{4.5}$$

where $\rho_{tot}(x)$ is the total charge density which in the contacts includes the uniform jellium of positive ionized donors. To be explicit, ρ_{tot} is given as follows

$$\rho_{tot}(x) = \begin{cases} \dfrac{|e|n(x)}{\epsilon_{si}} & \text{in the channel} \\ \dfrac{-|e|(N_d - n(x))}{\epsilon_{si}} & \text{in source/drain} \end{cases} \tag{4.6}$$

and $n(x)$ has to be calculated quantum mechanically as explained in the next section. Note, that in source and drain V_g is taken to be zero and V_{bi} is replaced by the terminal voltages, i.e. zero in source and V_{ds} in drain. Having calculated Φ_f the full $2D$ potential distribution can be obtained with equation (4.2) by computing $c_{1,2}$ from equations (4.3) and (4.4); the potential profile in z-direction at a particular position x is illustrated in figure 4.3.

A natural length scale λ appears in the reduced Poisson equation which plays the role of an effective screening length [2, 88]

$$\frac{d^2\Phi_f(x)}{dx^2} - \frac{\Phi_f(x) - V_g + V_{bi}}{\lambda^2} = \frac{\rho_{tot}(x)}{\epsilon_{si}} \quad \text{with} \quad \lambda = \sqrt{\frac{\epsilon_{si}}{\epsilon_{ox}} t_{si} t_{ox}}. \tag{4.7}$$

This screening length is due to the particular geometry of the considered device and has a somewhat different form for double-gated SOI MOSFET [88, 56]. It has the obvious implication that in order to achieve a small screening length the thickness of the SOI film as well as the gate-oxide thickness have to be reduced appropriately when scaling-down the channel length L. Since a small λ is effective in screening the potential variation at source and drain it is inevitable for scaled-down MOSFETs in order to suppress the short-channel effects and prevent punch-through. The channel length has to be significantly longer than the screening length. Simulations show that the inequality $2\lambda \lesssim L/3$ should be satisfied in order to get proper MOSFET operation[1].

4.1.2 Quantum Calculation

As already mentioned, for structures on a geometrical scale considered here a quantum mechanical calculation of the carrier dynamics is inevitable. Since the charge of the electrons significantly influences the potential distribution in the MOSFET structure the induced potential due to electron-electron interaction has to be taken into account. In the present approach the Hartree-potential is included by calculating self-consistently the electrostatic potential from the Poisson- and the quantum-equations; exchange correlations are neglected in the analysis.

The non-equilibrium Green's function formalism is employed in order to calculate the quantum mechanical charge- and current-distributions in the simulated structure since it allows to easily incorporate scattering in the structure and is capable of simulating an open quantum system by *exactly* taking the coupling to contacts into account [17].

The MOSFET structure is partitioned into three parts as shown in figure 4.2. Source and drain consist of contact 1/2 and the adjacent intermediate region which are regions where the electric field in x-direction is non-zero. The spatial extend of the intermediate regions depends on the screening of the electric field at the transition to the channel and therefore on the electron density in source and drain. The computational domain is truncated to the gray-shaded regions

[1] This is a rule of thumb which empirically emerged out of several different simulations.

which will be called the device or the conductor in the following. The actual contacts with zero electric field extending to $\pm\infty$ are exactly accounted for via a particular self-energy function [17, 18]. The computation is numerically performed on a tight-binding grid with lattice spacing a and nearest neighbor hopping parameter t. Details of the theory involved and the numerical procedure are given in appendix A.

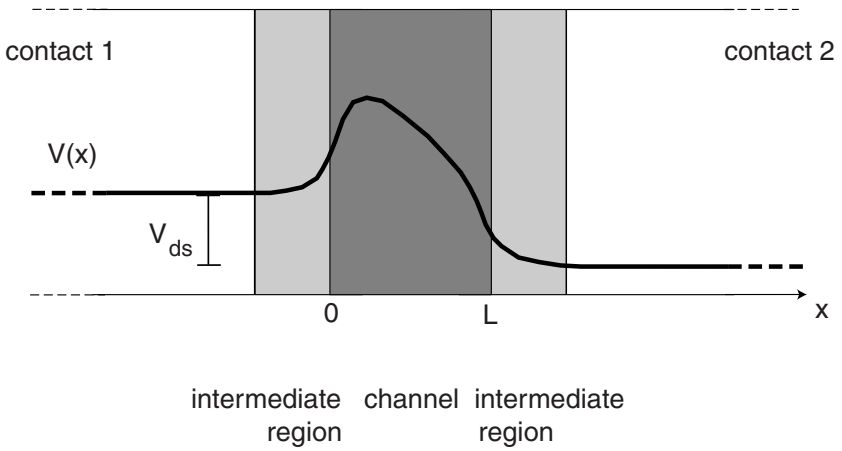

Figure 4.2: Potential distribution of the device. The light and dark-gray shaded regions represent the computational domain for which the numerical calculation is performed.

In order to describe the dynamics of the electrons in the device the retarded and advanced Green's functions have to be calculated. These Green's functions are determined by the following equations

$$\left[E - H - \Sigma^R\right] G^R(\vec{r}, \vec{r}', E) = \delta(\vec{r} - \vec{r}') \quad \text{and} \quad G^A = \left(G^R\right)^\dagger \tag{4.8}$$

where the retarded self-energy function Σ^R accounts for the coupling to the contacts, as well as for possible scattering events of electrons in the respective parts of the structure. Since a non-equilibrium situation is considered the functions G^n, G^p and Σ^{in}, Σ^{out} are necessary as well (note, that throughout the text DATTA's notation is used [17, 18]) which are the electron/hole correlation function and the in- and out-scattering functions which mediate the injection and absorption of carriers. The electron correlation function G^n is given by the following relation

$$G^n(\vec{r}, \vec{r}', E) = \int d\vec{r}_1 d\vec{r}_2 \, G^R(\vec{r}, \vec{r}_1, E) \Sigma^{in}(\vec{r}_1, \vec{r}_2) G^A(\vec{r}_2, \vec{r}', E). \tag{4.9}$$

Having calculated G^n, the local density of carriers follows as

$$n(\vec{r}) = 2 \times \frac{1}{2\pi} \int dE \, G^n(\vec{r}, \vec{r}, E) \tag{4.10}$$

where the factor of 2 accounts for spin degeneracy. With this density the electrostatic potential Φ_f can be calculated using the particular surface POISSON equation (4.5) derived in the previous section. The resulting potential is inserted into the HAMILTONIAN H in the equation for the retarded Green's function (4.8) of the device. Therefore, equations (4.8), (4.10) and (4.5) have to be solved iteratively until self-consistency is achieved.

It is well known that the iteration between the SCHRÖDINGER equation and the POISSON equation is numerically unstable and does not lead to a converged solution unless a relaxation factor is included [1]. In particular when dealing with an open quantum system self-consistency is not trivial to achieve (see the discussion in [29]) because the number of carriers is not a conserved quantity. In the simulations of open systems the reservoirs must be able to emit and absorb particles according to the requirement of overall charge neutrality in the device and zero electric field at the boundaries to the contacts. Different strategies are pursued in order to obtain self-consistency. For instance, PÖTZ employed so-called 'flexible' boundary conditions where the FERMI-energies in the reservoirs are slightly varied to ensure charge neutrality [57]. However, in the present work a NEWTON-RAPHSON iteration scheme is used (see appendix C of [42]) which is numerically stable and leads to self-consistency within a few iterations. Details of the procedure are given in appendix A.

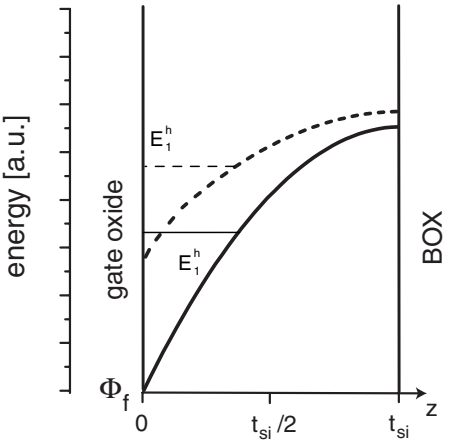

Figure 4.3: Potential profile in z-direction using the quadratic approximation. E_1^h is the first 2D-subband due to z-quantization.

Since the thickness t_{si} of the SOI film - bounded by the gate-oxide and the BOX - is only a few nanometer, the first two-dimensional subband due to the quantization in the z-direction is energetically well separated from the conduction band edge given by Φ_f, thereby enhancing the effective potential barrier height. Furthermore, the electric quantum limit [68] is considered here where due to the quantization only the lowest 2D subband is occupied which is justified for small t_{si}. This subband has the heavy effective electron mass m_h directed along the z-coordinate and a degeneracy factor of two [5]. Therefore, the transport mass is the light electron effective mass m_l which is also used in the HAMILTONIAN for the $1D$ calculations. Contributions of higher subbands are taken into account by a numerical factor according to reference [49] which is justifiable if the fraction of higher subbands is not too large. This factor is given by $2\times$*(total population)/(population of 1st subband)* where the 2 is due to the valley degeneracy of the first subband [49]. For 300 K, $t_{si} = 10$ nm and $E_f \approx 150$ meV the factor is approximately 2×2.1 and for $t_{si} = 15$ nm about 2×3.2, where E_f is the FERMI energy relative to the conduction band edge in the source contact.

The potential distribution for two different gate voltages is depicted in figure 4.3 for one point x in the channel. The potential profile is given by equation (4.2) for a particular surface potential Φ_f. The position of the first $2D$ subband is indicated for the two profiles as well. However, for simplicity the position of the first two-dimensional subband relative to the conduction band is approximated by the energy E_1^h of a rectangular potential well[2] which is constant throughout the channel and given by

$$E_1^h = \frac{\hbar^2 \pi^2}{2 m_h t_{si}^2} \tag{4.11}$$

One could compute an $\tilde{E}_1^h(x)$ depending on the position in the channel using explicitly the parabolic approximation of the potential under the assumption that it is possible to decouple the z-part of the HAMILTONIAN at each position x. This approximation holds until the potential variation along the x-direction is comparable to the variation in z-direction [67, 5] so that for ultra-short channel MOSFETs the validity can be questioned. A fully two-dimensional calculation would be necessary in this case.

Since all quantities will be averaged over z by integrating over z and dividing by t_{si} the exact form of the wave function of the first subband is of no importance as long as it is normalized so that the integration over the z-coordinate yields a factor of one. Translational invariance in the y, z-directions is assumed in source and drain, i.e. in the contacts and the intermediate regions for the computation of the carrier density. Hence, three-dimensional densities can easily be obtained by replacing all FERMI distribution factors by the so-called supply functions as stated in appendix A. The supply function is essentially the FERMI-integral of the order 1/2 multiplied by the energy-independent $2D$ density of states. In the channel, translational invariance can only be assumed in the y-direction whereas in z-direction the quantization has to be taken into account. Therefore, a special averaging procedure has to be performed which is described in the appendix. This procedure yields the three-dimensional carrier density of the first $2D$ subband in the channel and is also employed to get the $3D$ current density. As was mentioned in section 2.4.1, there is a fundamental contact resistance R_c due to a finite number of current carrying modes in the channel. Hence, if the width of the device is rather small the assumption of translation invariance is not valid any more. Here, we estimate the fundamental contact resistance according to reference [19] to be $70\,\Omega$ for a device with a width of 700 nm and $E_f = 150$ meV. In order to keep the simulation simple translational invariance in y-direction is assumed and the current I_d is corrected afterwards by adding R_c and $R_{ds} = V_{ds}/I_d$ which gives for unchanged V_{ds} the corrected

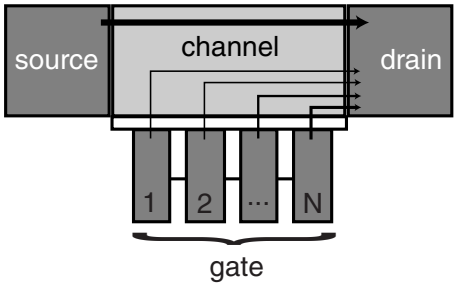

Figure 4.4: I_d contributions from source and the 'gate contacts'.

[2] E_1^h is the difference between the band bottom and the first subband.

current $\tilde{I}_d$ what is true for all series resistances. Parasitic series resistances do not only influence the drain current. The gate voltage is reduced as well according to $\tilde{V}_g = V_g - R_s I_d$ where R_s is the parasitic source resistance. This effect is neglected in the present analysis since $R_s = R_c/2 = 35\,\Omega$ results only in a slight change of V_g.

Care has to be taken with the effective masses. In source and drain the electron density is computed with the $3D$ density of states effective mass $m_{dos} = \sqrt[3]{m_h m_l^2}$ [5]. In the channel the $2D$ $m_{dos} = \sqrt{m_h m_l}$ and the conductivity or transport effective mass is simply the light electron mass m_l.

After self-consistency is achieved the source-drain current through the structure is calculated according to (see appendix A)

$$I_{sd} = W \times \frac{2e}{h} \int dE \, (f_s(E) - f_d(E, V_{ds})) \times T(E) \tag{4.12}$$

where $T(E)$ is the transmission function for carriers moving from source to drain and $f_{s,d}$ are the equilibrium FERMI distribution functions of source and drain.

For very thin gate oxides direct tunneling of electrons from the gate to the channel becomes more and more important and increases the current I_d in the subthreshold regime. Hence, the off-current of devices with ultra-thin gate oxides 'saturates' for $V_g \approx 0\,\mathrm{V}$ at the magnitude of the tunneling current. This means that for $V_g \lesssim 0$ the transport is dominated by the tunneling current from gate to channel, in particular by that part of the gate adjacent to the drain. For this part of the gate there is only the thin gate oxide as barrier so that the tunneling current raises with increasing difference $V_{ds} - V_g$. The subthreshold saturation of I_d can corrupt logic applications due to the lack of an off-state. Tunneling can be incorporated in the computations by attaching N 'gate contacts' with length L/N to the gate oxide along the channel as depicted in figure 4.4. All these contacts have the same E_f, i.e. they are shortened as is the case for the real gate. In the limit $N \to \infty$ a continuous gate follows. For simplicity, the gate electrode is assumed to consist of the same highly doped n^{++} silicon as source and drain. The total drain current is the sum of the current from source to drain and the part from gate to drain. This is illustrated in figure 4.4 with arrows going from source to drain and from each 'gate contact' to drain. The total drain current is explicitly given by

$$I_d = I_{sd} + \sum_{i=1}^{N} I_{gd}^i. \tag{4.13}$$

Again, details can be found in the appendix.

4.2 Simulations

In the following the computations presented above are used to investigate the electrical behavior of single-gated short-channel MOSFETs. In particular, the

influence of the SOI and oxide thickness on the output characteristics is studied. For all simulations room temperature and a width of 700 nm is assumed. In addition, the doping concentration in source and drain is always taken to be $3\times10^{20}\,\mathrm{cm}^{-3}$ which yields an $E_f = 150\,\mathrm{meV}$. According to DATTA et al. [19] the fundamental contact resistance is estimated to be $70\,\Omega$ for the chosen width. As mentioned above, the calculated drain current has to be corrected appropriately which is done for all simulations displayed below.

4.2.1 A 30 nm Device

Figure 4.5 (a) and (b) show typical gray-scale images of the local density of states versus energy for two different V_g and equal bias. Additionally, the potential profile of the conduction band is shown in the figure as well. The

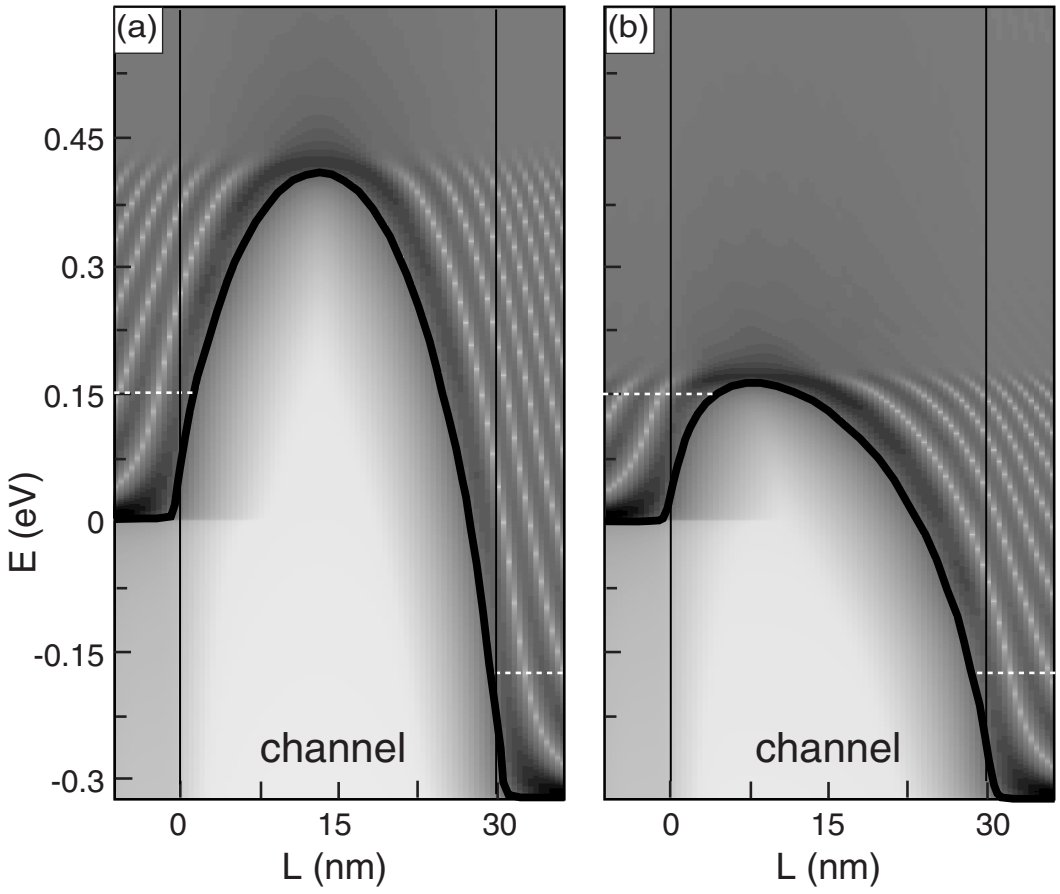

Figure 4.5: Local density of states for a 30 nm transistor at room temperature. The darker the image the higher the density of states. The position of the FERMI energies in source and drain are indicated by the white dotted lines. The vertical black lines illustrate the transition between channel and source/drain. Parameters for the simulation: $V_{ds} = 0.325\,\mathrm{V}$ and (a) $V_g = 0.05\,\mathrm{V}$, (b) $V_g = 0.55\,\mathrm{V}$.

parameters for the calculation are $L = 30\,\text{nm}$, $t_{ox} = 2\,\text{nm}$ and $t_{si} = 10\,\text{nm}$. The width and the doping concentration have the values stated above; V_{ds} and V_g are as indicated in the figure.

The local density of states exhibits the typical structure for energies below the maximum potential energy due to back-reflection of carriers [40]. Above this energy a continuum of states is available. The light gray within the potential barrier indicates a non-zero density of states which is due to the leakage of the electron wave-function into the potential barrier. A very high density of states can be seen right above the conduction band edge. Hence, vertical quantization is not important for $t_{si} \geq 10\,\text{nm}$.

Figure 4.6 (a) shows the corresponding simulated output characteristics of the device. The range of gate voltages is indicated in the diagram in steps of 0.05 V. In the saturation region the drain current increases with a significant slope which is a clear signature of SCE in this device. A closer inspection shows a DIBL of 110 mV/V, extracted from a direct inspection of the corresponding potential profile. The relatively high DIBL stems from the fact that the 10 nm thick SOI and $t_{ox} = 2\,\text{nm}$ results in a large amount of charge in the channel shared by the contacts and the channel, or equivalently the effective screening length $\lambda = \sqrt{\frac{\varepsilon_{si}}{\varepsilon_{ox}} t_{si} t_{ox}} = 7.8\,\text{nm}$ is too large compared to L. The transfer characteristics are shown in figure 4.6 (b) for V_{ds} =0.3..0.5 V in steps of 0.05 V. A threshold voltage of $V_{th} = 0.45\,\text{V}$ and a subthreshold swing of 87 mv/dec can be extracted.

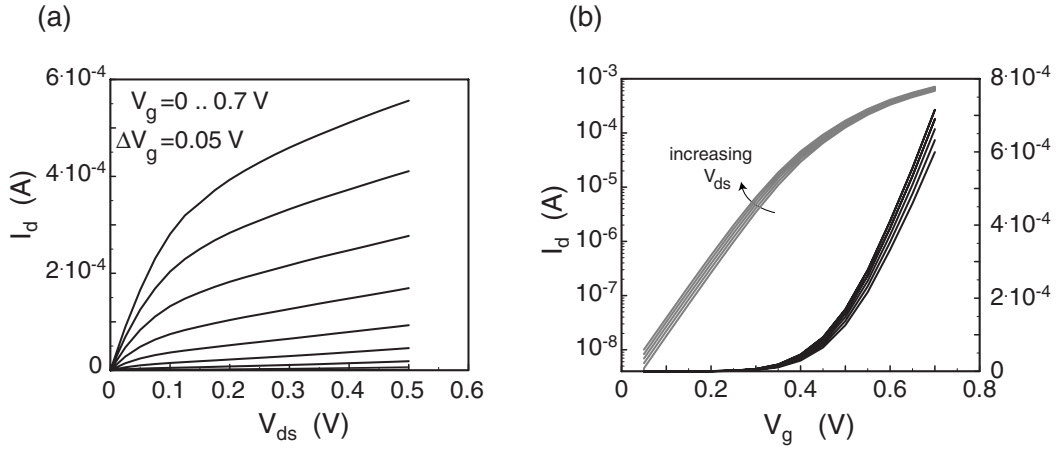

Figure 4.6: Simulated Output (a) and transfer (b) characteristics of a device with 30 nm channel length. $t_{si} = 10\,\text{nm}$, $t_{ox} = 2\,\text{nm}$, W =700 nm.

4.2.2 Channel-Length Dependence

The evolution of short-channel to long-channel behavior can be seen in figure 4.7. Transfer characteristics of devices with channel lengths $L = 30$, 40, 50 nm are shown in the diagrams. The parameters are equal to the ones used so far except that the SOI thickness is 15 nm instead of 10 nm.

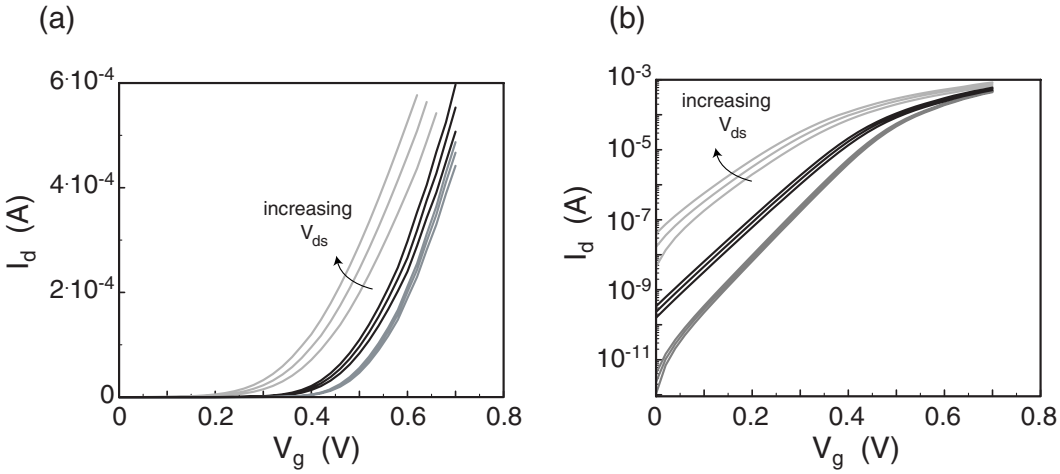

Figure 4.7: Simulated transfer characteristics of devices $L = 30, 40, 50$ nm for $V_{ds} = 0.3, 0.4, 0.5$ V.

The simulations are performed for three drain-source voltages $V_{ds} = 0.3$, 0.4, and 0.5 V. The transconductance g_m is extracted from these curves which is listed with the other figures of merit in table 4.1. One might wonder about the increase of transconductance for the longer devices which is contrary to experimental observations. The decrease of g_m with increasing channel length usually observed in real MOSFETs is due to the diffusive carrier transport in the channel. However, for ballistic MOSFETs where carriers move without scattering the transconductance should be constant for varying channel length if SCE are not important. If the channel length is reduced such that short-channel effects appear, the reversed dependence of g_m on channel length is expected since the gate control of the charge in the channel in the short devices is much worse than in the longer ones. Nevertheless, the variation of g_m-values is rather small. Ratios of 1.16 and 1.04 of the transconductances between the 30/40 nm and 40/50 nm devices, respectively, are found which reflect the discussed behavior.

The curves in figure 4.7 (b) show the subthreshold swing for the three devices. The MOSFET with $L = 50$ nm has a subthreshold slope of $S = 65$ mV/dec which is close to the ideal value at room temperature of 60 mV/dec. S increases with decreasing channel length. For the device with $L = 40$ nm $S = 80$ mV/dec and for $L = 30$ nm $S = 95$ mV/dec. Note, that the bending of the curves near $V_g \approx 0$ belonging to the 50 nm device is due to the finite energy interval

	30 nm	40 nm	50 nm
g_m	3700 mS/mm	4300 mS/mm	4500 mS/mm
S	95 mV/dec	80 mV/dec	65 mV/dec
V_{th}	0.36 V	0.48 V	0.52 V

Table 4.1: Comparison of devices with different channel lengths.

chosen for the simulation. The subthreshold swing increases with decreasing channel length which is a clear signature of the onset of short-channel effects in the $L = 40$, 30 nm devices. Additionally, the subthreshold swing shows with decreasing channel length an increasing dependence on V_{ds}. This is due to the DIBL as discussed in chapter 2. The potential barrier decreases with increasing bias and a higher drain current follows. Values found for the DIBL range from 58 mV/V for the 50 nm to 105 mV/V for the 30 nm device.

The threshold voltages are extracted from (a) by a linear extrapolation of the curve belonging to $V_{ds} = 0.5$ V and are also given in table 4.1. As is expected for devices with increasing SCE the threshold voltage rolls off for decreasing channel length.

4.2.3 Going to the Limits

In the reduced Poisson equation derived at the beginning of the chapter the length scale λ appears. This effective screening length allows a rough estimate of the design parameters for a short-channel MOSFET. As already mentioned the relation $2\lambda \lesssim L/3$ has to be fulfilled in order to suppress short-channel effects. In this subsection the dependence of the electrical behavior on t_{si} and t_{ox} is studied in detail. A channel length of 10 nm is chosen which is close to the theoretically predicted limit for MOSFETs with double-gate layout [50]. In contrast to the work of Wong [84] we will see that in principle MOSFET operation is possible for such short single-gated MOSFETs.

First, the evolution for decreasing oxide thickness is simulated followed by the lowering of the SOI film thickness. Tunneling from gate to the channel is taken into account as described above.

Reducing t_{ox}

Figure 4.8 shows output and transfer characteristics for MOSFETs with $t_{si} =$ 5 nm throughout. The gate oxide thickness t_{ox} is 2 nm, 1.5 nm and finally 1 nm from top to bottom; again, $W = 700$ nm.

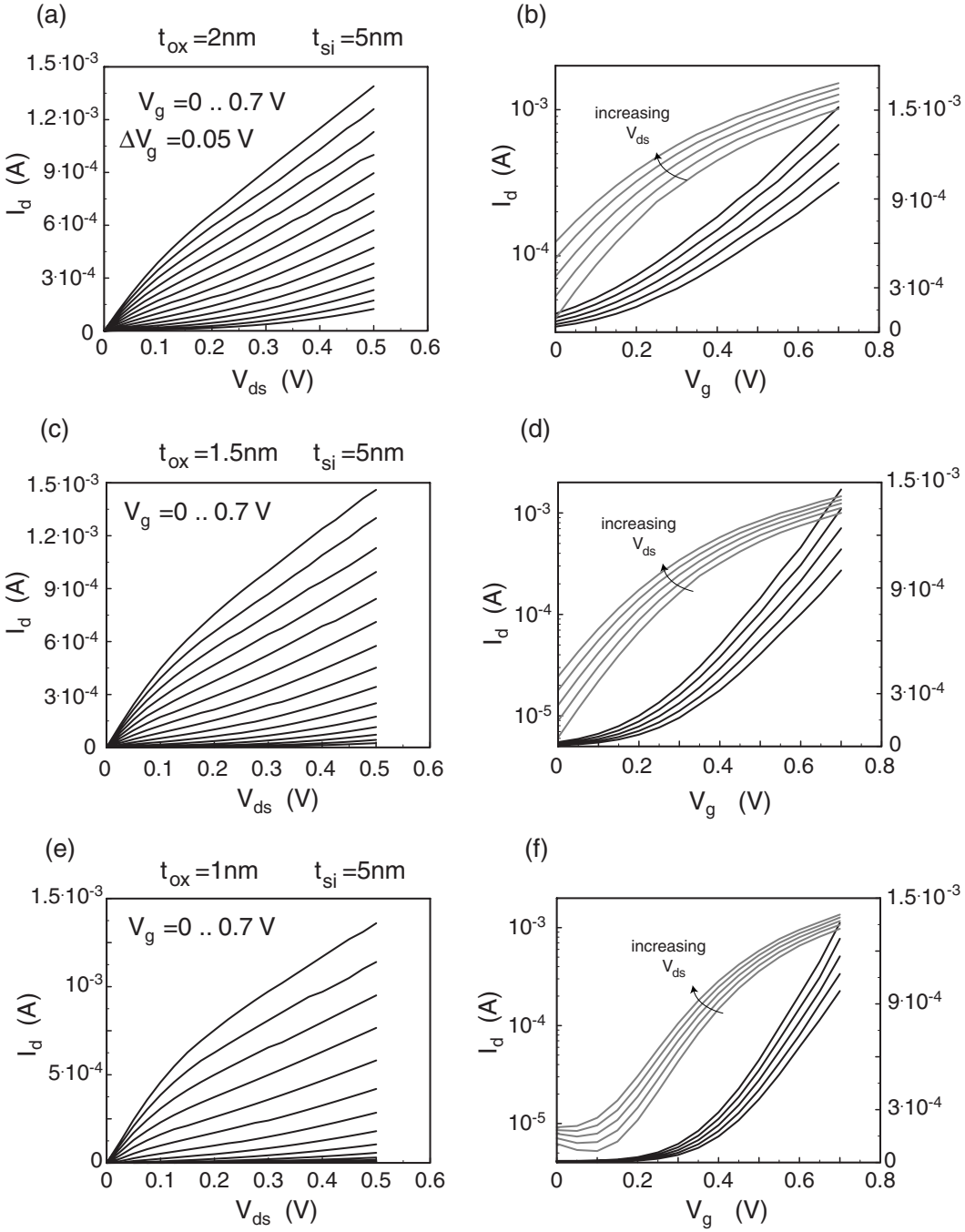

Figure 4.8: Simulated output and transfer characteristics for a 10 nm MOSFET. (a) and (b) $t_{ox} = 2\,\text{nm}$, $t_{si} = 5\,\text{nm}$, (c) and (d) $t_{ox} = 1.5\,\text{nm}$, $t_{si} = 5\,\text{nm}$, (e) and (f) $t_{ox} = 1\,\text{nm}$, $t_{si} = 5\,\text{nm}$.

All devices suffer from severe short-channel effects which can immediately be seen when looking at the increasing current in the saturation region of the output characteristics and the strong dependence of the subthreshold swing on V_{ds}. The first device shows the strongest increase of I_d in the saturation region. Even for $V_g = 0\,\mathrm{V}$ the drain current significantly increases with higher bias. This means that the device has no off-state since it can be switched on by raising V_{ds} alone. This is confirmed by the transfer characteristics. A threshold voltage cannot be determined since the device is in its on-state even for $V_g = 0\,\mathrm{V}$. An off-current of $\sim 10^{-4}\,\mathrm{A}$ flows, only one order of magnitude smaller than the drive current at $V_g = 0.7\,\mathrm{V}$. Hence, this device is useless for digital applications. The second device shows somewhat improved characteristics but still there is a lack of current saturation and no defined off-state. $S > 200\,\mathrm{mV/dec}$ shows the bad control of the channel in the subthreshold regime and the off-current is only 2..3 orders of magnitude smaller than the on-current. A further reduction of t_{ox} to 1 nm improves the characteristics again but in the subthreshold region I_d saturates at a value similar to the device with $t_{ox} = 1.5\,\mathrm{nm}$ which is due to direct tunneling from the gate to the channel. The improvement concerning SCE can be seen at the decrease of DIBL which lowers as $230\,\mathrm{mV/V}$ for the first device, $217\,\mathrm{mV/V}$ for the second and $185\,\mathrm{mV/V}$ for the device with minimal t_{ox}. Nevertheless, the DIBL for all three devices is very high and the devices are inoperative for digital applications.

Reducing t_{si}

Since tunneling would disable the device applicability for even thinner gate oxides one has to reduce the SOI film thickness in order to further improve the electrical behavior of the device. An alternative way would be the use of high-k gate dielectrics. But this has not yet been proven to work as good as SiO_2 and is still a matter of intense research.

The first two diagrams in figure 4.9 show the output and transfer characteristics for equal $t_{ox} = 1.5\,\mathrm{nm}$ and $t_{si} = 3\,\mathrm{nm}$ and $t_{si} = 2\,\mathrm{nm}$, respectively. Comparing the four diagrams one observes that decreasing t_{si} results in reduced SCE. The value for DIBL is lowered from $175\,\mathrm{mV/V}$ to $128\,\mathrm{mV/V}$ which can be seen in the lesser dependence of the transfer characteristics on the bias. The subthreshold swing is also reduced from $135\,\mathrm{mV/dec}$ in the first device to $105\,\mathrm{mV/dec}$ in the second one which reinforces the suppression of SCE for smaller t_{si}. Furthermore, the threshold voltage is increased from 0.46 V to about 0.65 V. This is a quantum mechanical effect since the energetic separation of the first two-dimensional subband and the conduction band bottom increases quadratically with decreasing t_{si} thereby enhancing the effective barrier height [54]. Therefore, the drive current in the second device is strongly reduced for $V_g = 0.7\,\mathrm{V}$. The simulations are extended to higher gate voltages in the case of (c) to (f) to compensate the quantum mechanically induced loss of gate overdrive.

The last two diagrams ((e) and (f)) belong to a device with $t_{si} = 2\,\mathrm{nm}$ and $t_{ox} = 1\,\mathrm{nm}$. V_{th} is further increased to 0.67 V but due to the thin gate oxide the drive current is almost twice as big as in the second device. The subthreshold

swing $S = 90\,\mathrm{mV/dec}$ shows the improved control of the channel by the gate when compared to the other devices. A DIBL equal to $80\,\mathrm{mV/V}$ shows the efficient suppression of short-channel effects for this device. Tunneling through the oxide truncates the off-current to $\sim 10^{-7}\,\mathrm{A}$. The device exhibits a very large transconductance of $6090\,\mathrm{mS/mm}$.

Discussion

In order to exhibit acceptable electrical characteristics concerning the short-channel effects the single-gated SOI MOSFET must have an extremely thin SOI layer of about 2 nm. With a quasi-classical estimation OMURA predicted heavier SCE for ultra-thin SOI due to an enhanced charge-sharing factor [55]. The results found here are contrary to this since a continuous suppression of SCE for decreasing t_{si} could be observed. In contrast to WONG et al. [84] who predict a threshold voltage roll-off for the single-gated MOSFET thereby disabling MOSFET operation, we find an enhanced V_{th} due to the vertical quantization. On the other hand the threshold voltage will thereby be increased to values of around 0.67 V which is rather high for low supply voltages as desired for today's semiconductor devices. Hence, for 10 nm MOSFETs the double-gated layout is favorable which exhibits a better control of the channel charge, thus reducing the SCE even for larger t_{si}. If a reduction to a $1D$ POISSON equation is performed as stated in section 4.1.1, the effective screening length is given by $\sqrt{\frac{\varepsilon_{si}}{2\varepsilon_{ox}} t_{si} t_{ox}}$ so that for equal λ the SOI thickness in a double-gated device can be twice as big as for the single-gated MOSFET of the same length. Since the energetic separation between the conduction band bottom and the first subband in the channel is inversely proportional to the square of t_{si} the deterioration due to vertical quantization is much less limiting for the double-gated device layout.

Nevertheless, MOSFET operation of a single-gated device in the 10 nm regime is, in principle, possible. The subthreshold swing $S = 90\,\mathrm{mV/dec}$ and a ratio of $\sim 10^4$ between on- and off-current meets the requirements for logic applications. The major challenge, however, lies in the fabrication of a homogeneous and extremely smooth 2 nm thin SOI film. Variations in t_{si} due to roughness would play an important role since they would appear as a varying effective barrier with specific height for different samples on a chip. This in turn would result in threshold voltage fluctuations since a variation of $\pm 0.2\,\mathrm{nm}$ yields a difference in barrier height of about 42 meV. This is almost the energetic difference between the position of the first $2D$ subband in the $t_{si} = 3\,\mathrm{nm}$ and the $t_{si} = 2\,\mathrm{nm}$ device which resulted in a shift of V_{th} of 0.19 V. On the other hand, MOSFETs based on a 2 nm thin SOI layer have already successfully been fabricated which exhibit a homogeneous silicon film thickness [55]. Even devices based on a 1 nm thin SOI film were experimentally realized [23]. Although the SOI was inhomogeneous in that case the devices exhibited well-behaved electrical characteristics. Additionally, gate oxides as thin as 0.8 nm have recently been demonstrated to work in a MOSFET with 30 nm channel length [13].

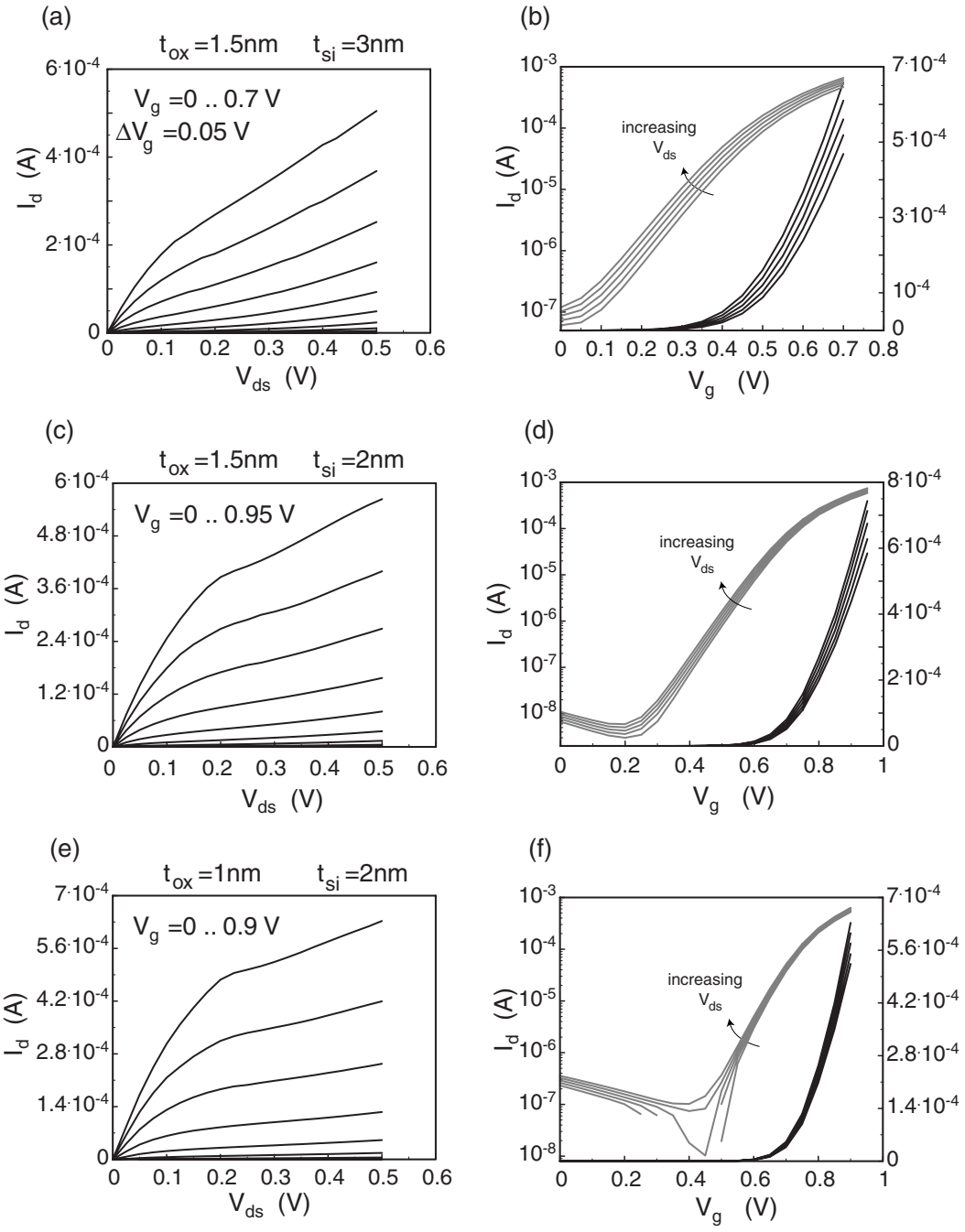

Figure 4.9: Simulated output and transfer characteristics for a 10 nm MOSFET. (a) and (b) $t_{ox} = 1.5\,\text{nm}$, $t_{si} = 3\,\text{nm}$, (c) and (d) $t_{ox} = 1.5\,\text{nm}$, $t_{si} = 2\,\text{nm}$, (e) and (f) $t_{ox} = 1\,\text{nm}$, $t_{si} = 2\,\text{nm}$.

Going to smaller channel lengths reduces the gate leakage current so that such thin gate oxides might work very well in ultra-short channel MOSFETs. However, producing the required ultrathin SOI remains the most challenging part in order to get functioning single-gated ultra-short channel devices on SOI.

4.3 Results

In this chapter simulations of a ballistic MOSFET were presented. Using a quasi one-dimensional model a fully quantum mechanical treatment became feasible. In particular, the influence of the SOI film and gate oxide thickness on the electrical behavior of ultra-short channel MOSFETs has been studied. Additionally, the presence of potential barriers at source and drain was investigated. The following results can be summarized:

- The transconductance for ballistic MOSFETs of various lengths remains constant or decreases with decreasing channel length if SCE become important.
- For $t_{ox} < 1.5\,\mathrm{nm}$ direct tunneling from gate to the channel becomes significant and truncates the off-state current of the devices.
- MOSFET operation of a single-gated device with 10 nm channel is possible for $t_{si} = 2\,\mathrm{nm}$ and $t_{ox} = 1\,\mathrm{nm}$ which show satisfying electrical characteristics. Extremely smooth ultra-thin SOI layers are required for such a device in order to avoid threshold voltage fluctuations due to vertical quantization.

For devices with channel lengths exceeding $\approx 20 - 30\,\mathrm{nm}$ the assumption of ballistic transport is not justified any more and inelastic scattering of carriers in the channel must be taken into account. This can be done in the present formalism by including self-energy functions for the respective scattering mechanisms [42]. Future work on the present simulations has to address the proper incorporation of scattering in order to enable realistic simulations even for longer devices.

5 Realization of an Ultra-Short Channel MOSFET

5.1 The Technology

In the present chapter the technology for the fabrication of the V-groove MOSFET introduced in chapter 3 is described in detail. Particular emphasis is put on the choice of etch-mask materials and the wet chemical patterning of the source and drain regions.

5.1.1 How to prepare a V-Groove?

It has long been known that silicon can be etched anisotropically with an aequous KOH solution [65, 22]. Although there exist other possible etch solutions (such as EDP or TMAH [45, 10, 78]), here, we will restrict the considerations to KOH because it is non-toxic and easy to handle. Therefore, it is used for all anisotropic etch steps.

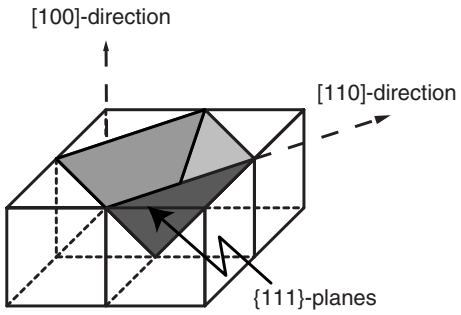

Figure 5.1: Schematics of {111} planes in a Si(100)wafer. The intersection of the planes forms a V-groove.

The most striking feature of KOH is that it has very different etch rates for different crystal planes of silicon. In particular, the etch rate of the {111} planes is very small compared to the {100} planes. On an (100) silicon wafer the {111} planes enclose an angle of 54.7° with the surface plane and are aligned along the ⟨110⟩ directions. Hence, in order to obtain a V-groove an etch mask has to be prepared consisting of rectangular structures aligned along the ⟨110⟩ directions of a (100) wafer. This is schematically shown in figure 5.1. The V-groove then consists of two intersecting {111} planes.

In general, the etch rate of KOH depends on the concentration and temperature of the solution [65, 22]. For the transistor to have a homogeneous channel length along its width smooth etch flanks and (100) surface are necessary. It is well known that a small KOH concentration leads to wavy {111} planes and to

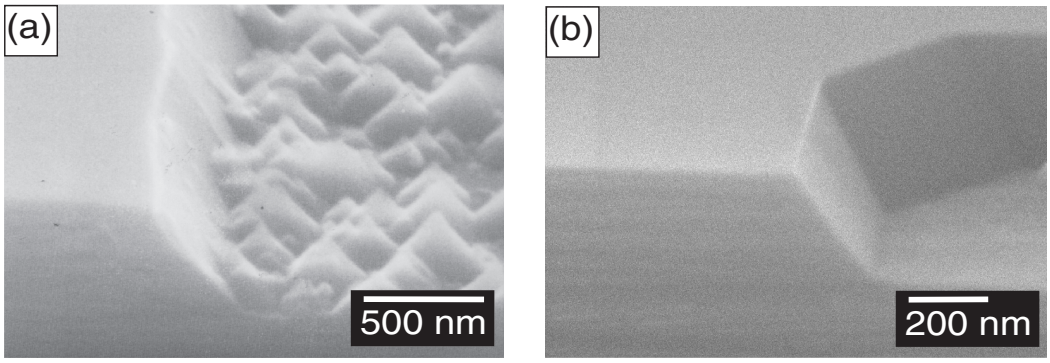

Figure 5.2: Electron micrographs of two etch flanks generated with (a) 20% KOH and (b) 40% KOH concentration.

a rough surface of the (100) plane [65, 22]. Figure 5.2 (a) shows an etch flank generated using a 20 % KOH solution at 30° C without stiring during the etching. One clearly sees that at the bottom (100) plane pyramidal structures are formed which would inhibit a proper V-groove generation and shorten source and drain. In contrast, figure 5.2 (b) shows an etch flank generated with a 40 % solution and stiring[1]. The etch flanks as well as the surface are smooth. Therefore, a 40 % KOH solution is chosen at a temperature of 30° C; the relatively low temperature ensures a controllable etch velocity.

KOH Etch Solution (40%)
• 112.5 g H_2O, 100 g KOH pellets • add 12.5 g isopropyl alcohol at 40° C • stir during the etching

When etching a silicon (100) wafer with appropriate mask the etching in the V-groove region essentially stops as soon as the two {111} planes meet. Only areas not bounded by {111} planes are further etched. For the V-groove MOSFET this means that everywhere except in the channel region the whole n^{++}/p^{-} stack is cut through resulting in an isolation of the devices from each other. This behavior is shown in figure 5.3 (a) for the case of a silicon (100) wafer. A mask consisting of two parallel long lines was generated on the wafer which was subsequently etched in KOH. The (unbounded) etch flank to the right of the structure is etched much further than the depth of the V-groove. Hence, a self-limited definition of source and drain and the device isolation can be performed

[1] Stiring is necessary in order to avoid concentration gradients.

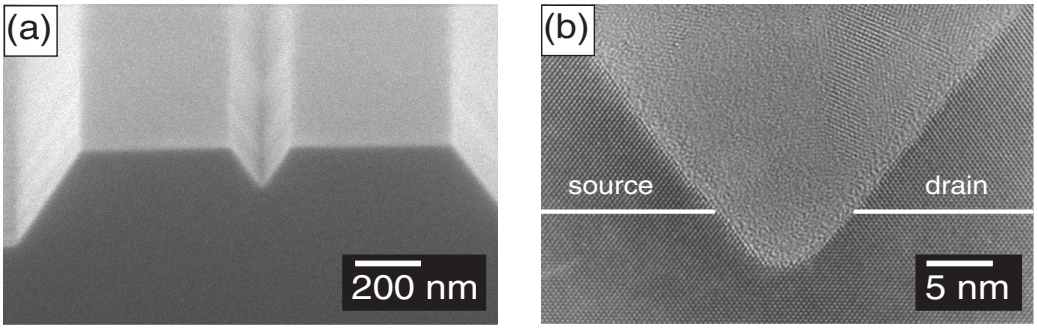

Figure 5.3: (a) self-limited behavior of the KOH etching. (b) TEM image of the V-groove tip. Lines are added to illustrate the p^-/n^{++} interface for a 10 nm device. The TEM image was taken by Ch. Dieker, Zentrum für Mikrostrukturanalyse, Universität Kiel.

in a single etch step. The self-limiting behavior is of major importance for a reliable and reproducible fabrication of the V-groove devices. Figure 5.3 (b) shows a transmission electron microscope image of the tip of the V-groove. Smooth etch flanks on a nanometer scale are obtained which is inevitable for the fabrication of homogeneous channels on this length scale. For illustration purposes, the position of the n^{++}/p^- interface for a transistor with 10 nm channel length is shown. The radius of curvature is approximately 3 nm which represents the lower cut-off of possible source-drain separations.

5.1.2 Mask Materials

The preparation of functioning devices critically depends on the formation of a proper V-groove as the essential part of the device concept which itself depends on the choice of an appropriate mask material for the KOH etch. In this section we will therefore discuss suitable materials which can, in principle, serve as etch masks.

Resist is not useful since it is readily dissolved in alkaline solutions like KOH. In principle, one can choose between SiO_2, Si_3N_4 or a metal which is not attacked by KOH. Such metals are for example gold, tantalum or chromium [22, 11]. Metals are advantageous because the samples can be kept at room-temperature during the evaporation which is important whenever one deals with a restricted thermal budget, to preserve a distinct doping profile, for instance. The most straightforward material certainly is SiO_2. In the specific case of the present transistor concept we are limited to a maximum temperature of 600° C to prevent diffusion of Sb from the n^{++} into the p^-, as will be discussed in section 5.1.5 [7]. Therefore, in order to use SiO_2 as mask material a low temperature oxidation process has to be employed. For this purpose the same process is utilized as is used for the gate oxidation (see section 5.1.5). This process yields a high quality oxide with excellent stability in KOH. Thus, rather

thin oxides of about 50-80 Å are sufficient. The SiO_2 is patterned with wet chemical etching in buffered oxide etch (BOE) which therefore makes thin oxides necessary to ensure a good pattern transfer. The definition of a SiO_2 mask pattern is depicted in figure 5.4 (a). A resist mask is generated with electron beam lithography which is used as etch mask for the BOE. After the removal of the resist mask the V-groove can be prepared with KOH.

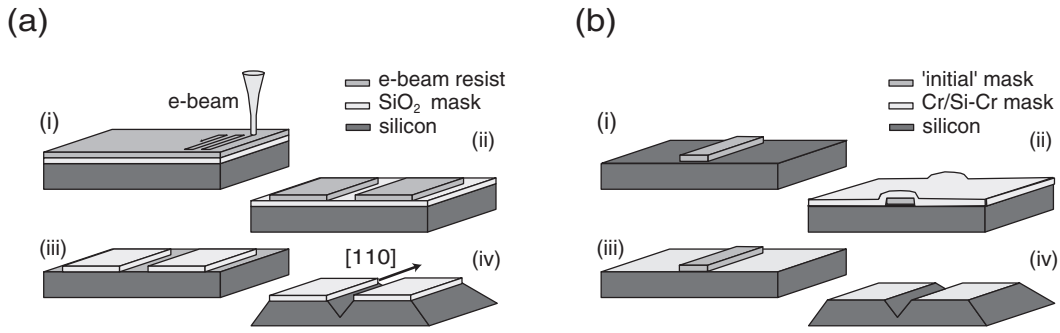

Figure 5.4: Fabrication of V-grooves using (a) a SiO_2 and (b) a Si-Cr etch mask prepared as described in the main text.

Alternatively, chromium can be used as mask material making use of a particular process schematically shown in figure 5.4 (b). First, the sample surface is cleaned with an RCA1/2 procedure [39] leaving an approximately 10 Å chemically generated SiO_2 on the surface. Afterwards, an 'initial' mask is generated on the sample surface with the only restriction that it must withstand an HF-dip. The chemically generated oxide is removed using $HF : H_2O = 1 : 100$ which yields a clean and hydrogen passivated silicon surface [31]. The samples are mounted in a UHV chamber immediately after the oxide removal and are subsequently coated with a 20 nm thick chromium layer evaporated with an electron beam gun. The evaporation of chromium on a clean silicon surface under UHV conditions is known to generate a Si-Cr interface layer about 10-13 Å in thickness [26, 27]. After the evaporation the sample is baked on a hot plate for one hour at 120° C and subsequently the Cr layer is removed using commercially available Cr-etch. The Si-Cr interface layer which is not removed by the Cr-etch step is thus generated only in areas where the chromium was in direct contact with the clean silicon substrate. Hence, having removed the 'initial' mask as well a KOH etch step can be performed since the Si-Cr layer is not attacked by KOH for usual etching durations. The advantages of this technique are that no lift-off technique is needed to generate the actual etch mask since the 'initial' mask serves only as shadow mask. Therefore, a great variety of mask materials and very small mask structures can be used to pattern the Si-Cr layer. Additionally, the Si-Cr mask is the exact invers of the initial mask, i.e. a non-deformable mask generation is obtained.

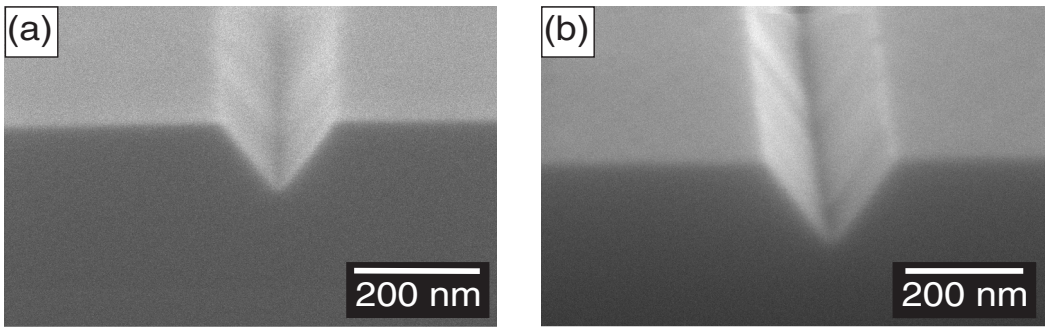

Figure 5.5: Electron micrographs of V-grooves fabricated with (a) SiO_2 and (b) SiCr as etch mask. Both images were taken under a tilt angle of 70°. The flanks of the V-groove are aligned along a [110] direction.

Figure 5.5 shows V-grooves etched in silicon using SiO_2 (a) and Si-Cr (b) as mask material, respectively. One clearly sees that both processes lead to the formation of a V-groove with smooth etch flanks. Thus, both techniques are in principle useful for the V-groove formation. The drawback of the Cr-technology is that it is difficult to generate a homogeneous Si-Cr layer extending over the entire sample which is attributed to the lack of a perfect hydrogen passivation and contaminations present on the sample surface. This difficulty could be overcome by removing the chemically generated oxide in nitrogen atmosphere and by using an ethanol-HF mixture which is known to yield a perfect hydrogen passivation of silicon surfaces [31, 32].

Due to the reasons mentioned above SiO_2 is exclusively used as mask material in the following since better results concerning the homogeneity are achieved. In addition, it is intended to establish a process which relies on standard procedures of semiconductor technology.

5.1.3 Etch Mask Definition

The patterning of the SiO_2 mask for the actual transistor structure was performed using a negative e-beam resist[2] since only very small parts of the whole sample have to be covered by SiO_2. The advantage of the resist is that it is e-beam as well as optically sensitive. Whereas the e-beam process is negative the optical one is positive. In order to transfer the resist pattern into the oxide mask we desist from taking reactive ion etching (RIE) because it is very difficult to exactly control the RIE process to stop right at the Si/SiO_2 interface. This is important since etching the silicon is equivalent to a widening of the mask opening. Another important issue is the roughness of the silicon surface introduced by a RIE step which leads to a non-uniform V-groove formation. Therefore, we preferred to pattern the SiO_2 mask with wet chemical etching using BOE which etches SiO_2 readily but attacks the highly doped n^{++} sili-

[2]Allresist GmbH; Resist: AR 7500/107, Developer: AR 300-47

con only slightly. In addition, rather thin oxides have to be used so that the transferred SiO_2 mask has the same shape as the resist pattern. As already mentioned, the actual thickness was in the range of 50-80 Å.

Source/Drain Definition
• resist: 6000 rpm, 5 min, 90°
• pre-patterning: 7 s UV exposure, 11 mW/cm^2, develop 5 s
• e-beam: ∼400 μC/cm^2, water 3 min, UV flood-exposure 3 s
• develop 30 s, post-bake 120° C, 1 h
• BOE ∼12 s, remove resist in NMP at 80° C for 2 h
• immerse in acetone, IPA and water
• dip in BOE before KOH, 30° C, 40%
• water for at least 5 min

The detailed process is as follows. After spinning the resist with 6000 rpm it is prebaked at 90° C for five minutes. It is important to rather exactly stop the baking after five minutes because the sensitivity of the resist on electrons depends on the time as well as on the temperature of the prebake step. The resist is then prepatterned with optical lithography so that there is resist only in the relevant areas of the sample. This is essential because the SiO_2 marks for e-beam adjustment are not visible when coated with resist. The optically positive resist is exposed with UV light for 7 s and subsequently developed 5 s in the appropriate developer. Afterwards, the sample is mounted into a scanning electron microscope for the e-beam exposure; a relatively high dose of about 400 μC/cm^2 is used. Prior to the development of the resist a UV flood-exposure for 3 s is employed. This drastically diminishes the influence of the proximity effect [60] and sharpens the resist mask pattern since areas unintentionally exposed with electrons via the proximity effect are washed away in the developer due to the UV sensitive component of the resist; only the virtually exposed areas remain where the electron dose is very high. The actual time of development is 30 s followed by a rinse in DI water. A subsequent post bake at 120° C for 1 h ensures that the resist pattern sticks well to the substrate and enhances its durability when exposed to BOE. The transfer of the resist pattern into the mask is done by etching the oxide in BOE up to the bare silicon surface. This step is easily controlled since clean silicon surfaces are hydrophobic whereas SiO_2 is hydrophilic. The resist is then dissolved in 1-methyl-dipyrolidon (NMP) at 80° C for about 2 h. Subsequently, the samples are immersed in acetone, isopropyl alcohol and DI water. Immediately before KOH etching the samples are dipped in BOE in order to remove the native oxide

($\approx 1\,$nm thick) and hence any contamination present on the sample surface. The KOH etch step is done with the 40% solution mentioned earlier followed by a thorough rinse in DI water for at least 5 min which removes all K-ions.

5.1.4 Mask Geometry

As is well known from micro-electro-mechanical-systems (MEMS) technology, (for a recent review see e.g. [22]) when etching silicon anisotropically in KOH convex corners are etched very rapidly since they are dominated by fast etching planes [65, 11, 64]. This is an unwanted effect because with increasing etching time the shape of the etched structure differs more and more from the originally intended one. The result of this effect is shown in figure 5.6 (a) which is an electron micrograph of a convex corner covered with a SiO_2 mask. The area enclosed in the broken line is the silicon dioxide mask which has been undercut by the KOH etch. The undercut is dominated by the fast etching {221} planes [64]. In the present transistor concept, for the isolation of source and drain, convex corners naturally appear so that one has to take care of the actual geometry of the etching mask. If the width W_0 of the V-groove and thus the transistor is too small the preceding undercut at the corners leads with increasing etching time to a consumption of the active transistor area. We therefore chose a mask geometry as is shown in figure 5.6 (b) where one long, straight structure ($\approx 2\,\mu$m) is opposite and parallel to a structure of various width W_0. As indicated, the two opposite structures represent source and drain each having two leads to enable four terminal measurements. The advantage of

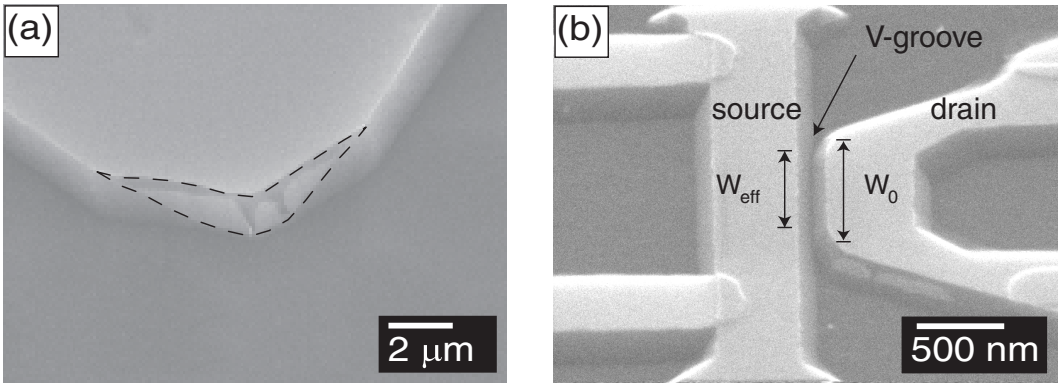

Figure 5.6: (a) Electron micrograph of the undercut of convex corners during KOH etching. The etching is dominated by fast etching {221} planes. (b) Geometry of the transistor structure generated in a Si(100) wafer. The structure enables four terminal measurements.

this particular geometry is that the long part does not change its shape since there are no convex corners in the relevant area whereas the opposite part can be adjusted so that V-grooves and thus transistors of various widths can be fabricated.

In the case of figure 5.6 (b), the undercut of the SiO_2 mask at the corners of the drain pattern reduces the effective width (indicated as W_{eff}) of the transistor compared to the original intended width W_0. Thus, a substantial difference between W_0 and W_{eff} has to be taken into account. Since a corner compensation (cf. [59, 22]) is not possible in the present case because it would connect the right and left part of the structure the only way to get a functioning device is to make the width W_0 of the SiO_2 mask long enough to prevent the undercut from consuming the active transistor area. Knowing the required etch depth, the width of the mask W_0 can be adjusted accordingly in order to reproducibly fabricate V-grooves of a certain width. In general, we find a ratio of $\sim 1 : 0.85$ between etch depth and mask undercut at convex corners. For example, a n^{++}/p^{-} stack of 100 nm thickness requires a mask with a width of at least 2×85 nm.

5.1.5 Gate Oxidation and Metal Gate

After the V-groove formation a thin gate oxide has to be grown in order to isolate the gate electrode from the channel as well as from the V-groove flanks belonging to source and drain. In order to avoid the diffusion of dopants from the highly doped epitaxial n^{++} layer into the p^{-}, i.e. to preserve the abrupt doping profile, the thermal budget is restricted to a temperature of 600° C [6]. Therefore, a special gate oxidation process had to be developed which is explained in detail now.

The temperature for dry thermal oxidation with reasonable growth rates exceeds by far 600° C. On the other hand when performing a wet thermal oxidation the pyrogenic reaction between hydrogen and oxygen must for safety reasons take place at approximately 750° C. Therefore, a possibility for a low temperature oxidation is to use wet thermal oxidation and establish a temperature gradient in the oxidation furnace keeping the sample temperature at 600° C. This scenario is schematically shown in figure 5.7.

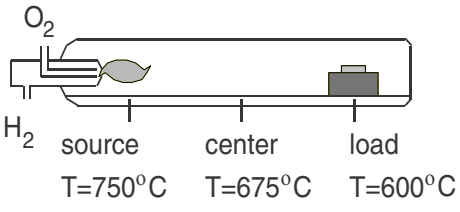

Figure 5.7: Schematics of the temperature gradient in a standard oxidation furnace.

The oxidation rate of the wet thermal oxidation process is approximately 5 Å/h and therefore yields a controllable growth rate for very thin gate oxides. The actual procedure makes use of a dry-wet-dry cycle which enhances the oxide quality [6]. Prior to the oxidation the sample is dried in nitrogen for 1 h at 600° C. The best results are obtained when the samples are cleaned with RCA1 and RCA2 right before the oxidation, thus starting the oxidation procedure with the RCA2 generated 10 Å thin SiO_2. Samples which are dipped in HF before the oxidation in order to remove the native oxide show worse results since the incomplete hydrogen passivation of the sample surface does not yield a homogeneous oxide formation during the

dry oxidation part of the whole cycle. Therefore, all oxides are grown with the RCA2 generated oxide as starting point which is at the same time the lower cut-off of the available oxide thicknesses.

Due to the importance of the gate oxide for the performance of the devices the oxides are characterized by current versus voltage, capacitance versus voltage, and parallel conductance versus voltage measurements [6]. Tungsten is

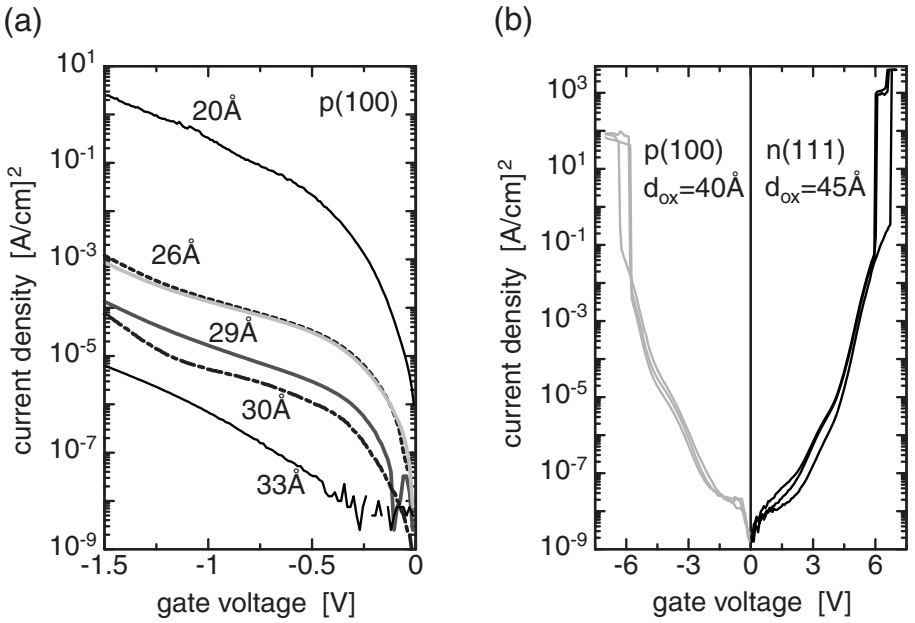

Figure 5.8: (a) Leakage current densities for low temperature wet thermal and high temperature oxide. (b) Leakage current for n/p-doped wafers with different orientation.

chosen as gate electrode material. It has the advantage of a midgap workfunction and is therefore a candidate for advanced CMOS technology [52]. A 50 nm tungsten layer is evaporated immediately after the gate oxidation which is patterned with optical lithography and wet chemical etching in H_2O_2 at room temperature.

The leakage current densities for oxides prepared with this oxidation procedure on $p(100)$ silicon substrates are shown in figure 5.8 (a) for various thicknesses ranging from 20 Å to 33 Å. In the range shown in figure 5.8 (a) the leakage current is dominated by direct tunneling. For comparison, the leakage current densities of two oxides with thickness equal to 26 Å and 30 Å grown at elevated temperatures in dry oxygen are plotted as well (dotted lines). As can be seen the wet thermal oxidation procedure yields high quality oxides comparable to the high temperature oxides grown in dry oxygen.

Figure 5.8 (b) shows the leakage current densities over a wider voltage range for a thickness of 40 Å and 45 Å grown on a $p(100)$ and an $n(111)$ substrate, respectively. Note, that the orientations of the test wafers are chosen according

to the realities of the device which the oxidation procedure is developed for. The onset of breakdown is around 12 MV/cm independent of the orientation of the wafer. Furthermore, an interface trap density of $8 \times 10^{10}\,\mathrm{cm}^{-2}\mathrm{eV}^{-1}$ is found [6] confirming the high quality of the low temperature oxide.

An important point has to be addressed concerning the dependence of the oxide thickness on the orientation as well as on the doping level. It is known [51] that the oxidation of [111] oriented silicon yields an increased oxidation rate of about 20% compared to [100] oriented silicon. Additionally, the oxide thickness on highly ($10^{20}\,\mathrm{cm}^{-3}$) doped silicon is approximately 30% higher than on lightly doped silicon. Thus, for the actual device the gate oxide on the V-groove flanks is expected to be ∼1.5 times the thickness of the gate oxide in the channel region. This is favorable since a much better isolation between source/drain and the gate is obtained due to the exponential dependence of the direct tunneling current on the thickness of the gate oxide. Additionally, the parasitic capacitance between gate and source/drain is reduced.

5.2 A Real Device

Having developed a suitable technology for processing the V-groove transistors the question now is whether the technology is transferable to the actual epitaxial material system. All steps until now were exclusively done on simple silicon (100) wafers. The next sections are devoted to answer this question and show the outcome of the fabrication process.

5.2.1 The Material System

As mentioned above the material system consists of a degenerately doped silicon n^{++} layer grown with solid phase epitaxy (SPE) on a thin nominally undoped p^- ($\sim 5\times10^{14}\,\mathrm{cm}^{-3}$) silicon on insulator (SOI) wafer[3]. The SOI layer is thinned down to 10 - 15 nm using cycles of dry thermal oxidation and etching with HF. The deposited amorphous silicon film recrystallizes when annealed at a temperature of 600° C. The advantages over usual molecular beam epitaxy are the high doping level accessible only with SPE [70] and the low process temperatures not exceeding 600° C which prevents the Sb dopants from diffusing. Hence, abrupt doping profiles can be realized with this technique [69].

The thickness of the n^{++} layer is intended to be around 100 nm in order to get mask openings in the range of 140 nm, easily accessible with electron beam lithography. A transmission electron microscope image of the material package with a 15 nm p^- and a 30 nm n^{++} layer is shown in figure 5.9; the Cr on top of the sample enhances the contrast of the image. The n^{++} shows the formation of twin defects starting at the interface to the p^- substrate. These twins always appear but do not limit the use of the SPE grown films for the V-groove formation since they do not affect the self-limiting etch behavior.

[3] Unibond®, S.O.I.TEC, France

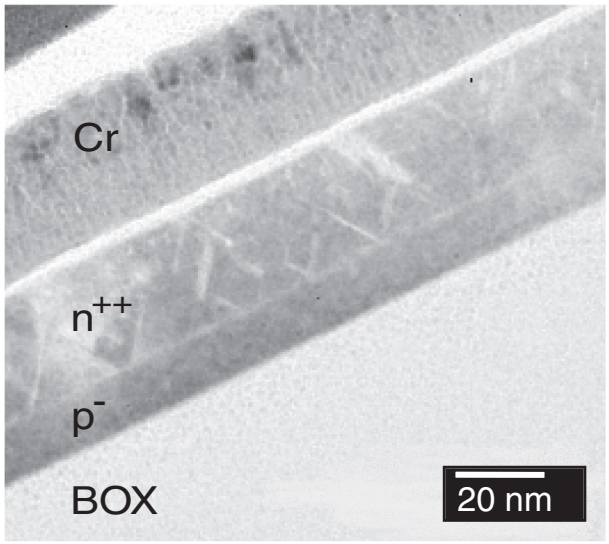

Figure 5.9: Transmission electron microscope image of the epitaxial silicon stack. A 30 nm n^{++} layer is grown on a 15 nm nominally undoped SOI wafer. The TEM image was taken by J. Benedict, IBM Semiconductor Research and Development Center, Hopewell Junction, New York.

The most delicate part of the growth of the epitaxial silicon stack is the cleaning of the SOI wafer right before the epitaxy. There are essentially two ways to do that. First, the wafer is cleaned with an RCA procedure [39]. As already mentioned, the RCA2 leaves the sample surface with a thin oxide layer. This layer is heated off under UHV conditions in the MBE chamber at elevated temperatures immediately before the actual epitaxy. The process has the advantage that a clean surface is achieved but has the drawback that it induces a relatively high roughness on the sample surface [32]. The second way is to do the RCA1/2 cleaning procedure as before but to remove the chemically generated SiO_2 immediately before mounting the sample in the MBE chamber relying on the hydrogen passivation of the silicon surface [32]; the hydrogen passivation is subsequently heated off at a temperature of 600° C for 5 min under UHV conditions. No roughness is induced using this procedure but an interface layer one or two monolayers in thickness remains between n^{++} and p^{-} which is only slightly attacked by KOH. This interface consists of SiO_x and carbon-hydroxides due to an incomplete hydrogen passivation and contaminations of the silicon surface prior to the MBE [31, 32].

A vertical roughness of the sample surface is essentially equivalent to a horizontal roughness of the etch mask along its width which would result in an inhomogeneous V-groove formation that cannot be tolerated. Therefore, all samples are grown with the HF-dip before the epitaxy although it is difficult to control the composition of the resulting interface layer.

Cleaning Procedure
• SC1, 65° C, 5 min • $HF : H_2O = 1 : 100$ • SC2, 65° C, 5 min • HF-dip immediately before mounting the sample in UHV chamber • heating off the hydrogen passivation: 600° C, 5 min

Particular care has to be taken during the cleaning of the SOI wafer in order to minimize the effect of the interface layer since contaminations are known to strongly deteriorate the crystalline quality of SPE silicon films [3]. Furthermore, if the interface layer is too strong that it is not attacked by KOH the devices will not be isolated from each other and no well-defined channel area can be obtained. On the other hand, a moderate interface is advantageous as is revealed by the transmission electron micrographs shown in figure 5.10. This figure shows V-grooves etched in a p^-/n^{++} material system generated with the second cleaning method described above. All images were taken from the same sample but with different V-groove openings L_0 increasing from (a) to (f). Images (a) and (b) show V-grooves with openings so small that the whole n^{++} layer is not cut through. Thus, left and right side, i.e. source and drain, are shortened. The V-grooves in images (c) and (d) have an appropriate opening whereas (e) and (f) are too large so that the entire n^{++}/p^- stack is cut through. The images reveal the effect of a moderate interface layer: for (c) and (d) the tip of the V-groove is flattened since the KOH attacks the interface only slightly. As long as the flat area is not too large the interface can withstand the etch. If this area gets enlarged ((e) and (f)) the interface layer and thus the p^- is cut through and no device action is possible anymore. Thus, a moderate interface allows for the device isolation but yields desirable flat channel regions. This is particularly important when the p^--thickness is reduced for a better control of short-channel effects. For example, the maximum source-drain separation L_{max} for a 5 nm p^- SOI film without consuming the whole p^- by the KOH is only about 7 nm when the usual V-groove geometry is encountered. Thus, the interface layer makes it possible to utilize very thin SOI films in order to suppress short-channel effects and to exploit the vertical quantization without carrying the channel to unreasonably short lengths.

A further inspection of image (c) shows that a separation of 10 nm between the two triangularly shaped n^{++} regions (i.e. source and drain) can be achieved. This demonstrates the ability of the present approach to provide nanometer-scale control for the fabrication of 10 nm MOSFET devices.

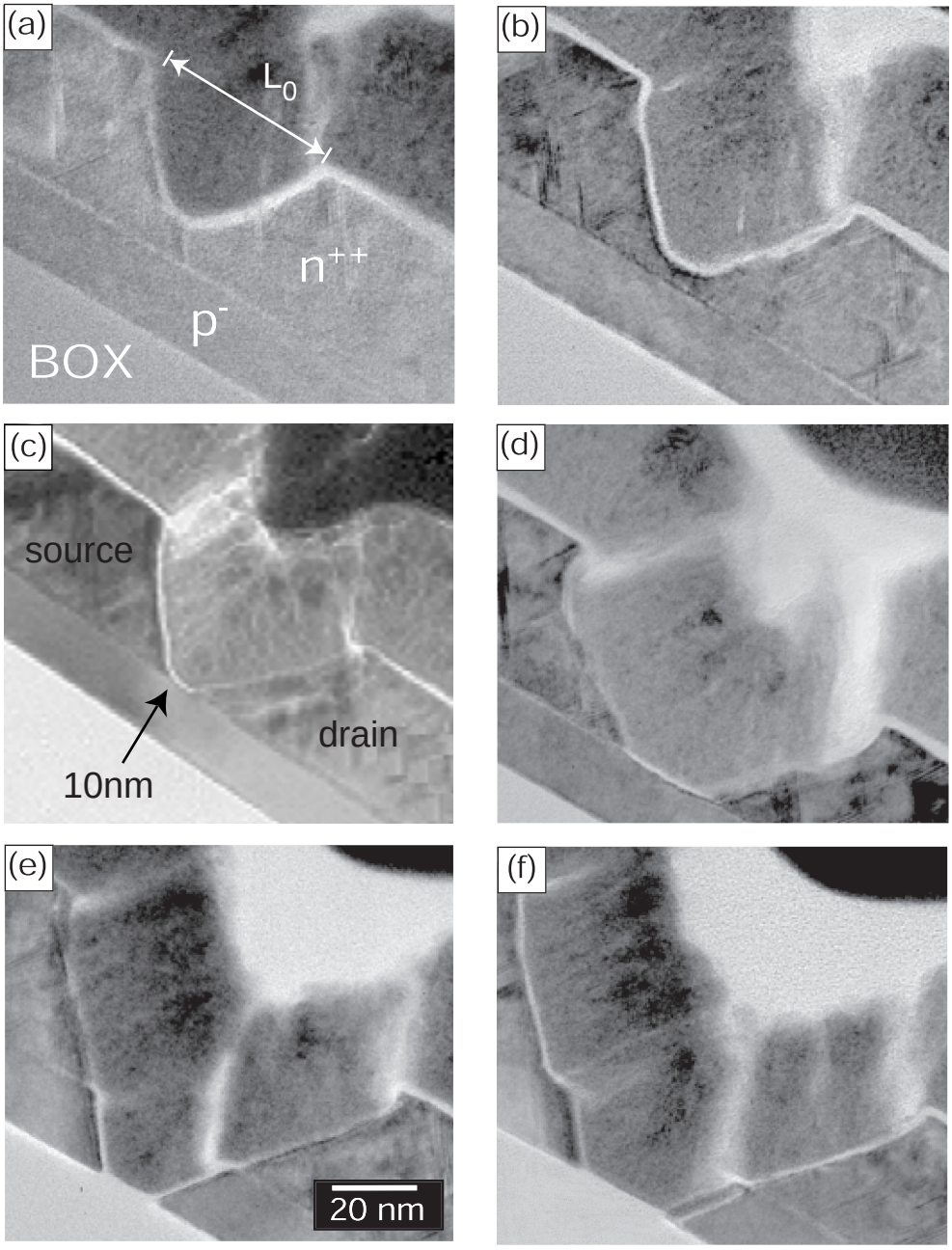

Figure 5.10: V-grooves generated in an n^{++}/p^{-} silicon stack. The etch mask opening L_0 increases from (a) to (f). (c) and (d) have appropriate openings for a functioning device. Due to the interface layer the channel area is flat. Image (c) has a source-drain separation of only 10 nm which proves the feasibility of the V-groove approach to generate ultimately short source-drain separations. The TEM image was taken by J. Benedict, IBM Semiconductor Research and Development Center, Hopewell Junction, New York.

5.2.2 Optical and Electron-Beam Lithography

The first step of device fabrication is the generation of a hard mask consisting of a low temperature SiO_2 (LTO, typically 100 nm in thickness) which is raised on the sample surface. This LTO is patterned with optical lithography and wet chemical etching in BOE. Care must be taken with the arrangement of the optical pattern since it must allow the active structure to be aligned along a $\langle 110 \rangle$ direction.

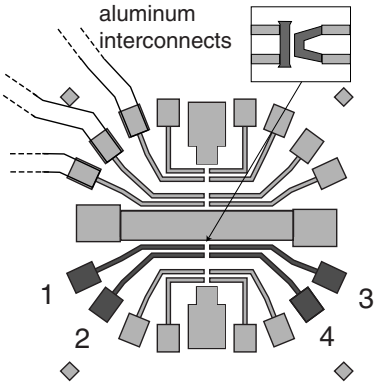

Figure 5.11: Hard-mask pattern. 1/2 are leads to source, 3/4 to drain.

The hard-mask pattern can be seen in figure 5.11 which allows to contact four devices each having two leads for the source contact and two for the drain contact. As an example, four such contacts belonging to one device are indicated with numbers 1..4 (dark gray). The four rhombs around the optical mask pattern serve as marks for the adjustment of the electron beam lithography. The actual device is patterned with electron beam lithography within the optical mask pattern as stated in the next section. The inset shows how the e-beam pattern (dark gray in the inset) for the V-groove fits into the LTO mask. The contact pads at the end of the n^{++} leads will be contacted with aluminum interconnects as the last step of the whole fabrication process (after the removal of the LTO at the pads).

Having transferred the optical resist mask pattern into the LTO the resist mask is ashed in an oxygen plasma and the sample is cleaned with RCA1/2. The next step is the generation of the thin SiO_2 mask for the KOH etch step using the low temperature wet thermal process. This oxide is patterned with the negative electron-beam procedure explained in detail in section 5.1.3 and the geometry introduced in section 5.1.4. The progression of the subsequent KOH etching can actually be seen due to interference of light reflected at the top surface and the Si/SiO_2. The etching is stopped as soon as the BOX is reached followed by a thorough rinse in DI water in order to remove all K-ions from the sample surface. Afterwards, the samples are cleaned with RCA1/2.

Figure 5.12 shows the device without gate-metallization etched in the n^{++}/p^{-} material system. The dashed line is directed along the channel length. This image has to be compared to figure 5.6 (b). It can be seen that although the flanks of the V-groove in the epitaxial silicon stack are not as smooth as in bare silicon the process nevertheless yields the desired V-groove between source and drain. A width of 700 nm is taken from a magnification of this SEM image showing the already mentioned deviation between the originally intended width $W_0 = 1000\,\text{nm}$ and the effective width of the processed device.

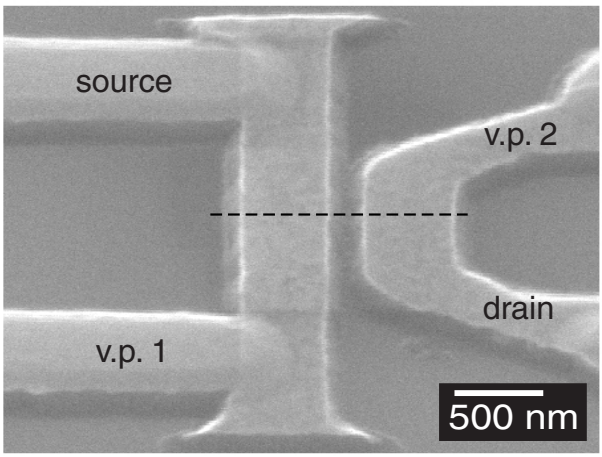

Figure 5.12: Device without gate-metallization fabricated in the n^{++}/p^{-} silicon stack. Measured devices are cut through along the dashed line to facilitate TEM images. The device has four terminal, source/drain and two additional voltage probes denoted 'v.p. 1/2'.

5.2.3 A Complete Device

The final steps of the device preparation are the gate oxidation, the preparation of the tungsten gate and the metal interconnects which allow measurements in a probe station. The gate oxidation proceeds as described above. Immediately after the oxidation a 50 nm tungsten layer is deposited on the sample surface which is patterned using optical lithography and wet chemical etching in H_2O_2. The interconnects are patterned with optical lithography and standard lift-off technique. A 100 nm thick layer of aluminum is chosen as interconnect material. Right before the evaporation of the interconnects the LTO is removed at the pads of the optical mask (see figure 5.11) in order to ensure a good electrical contact between the aluminum and the n^{++}. Finally, an additional aluminum layer is patterned with optical lithography and lift-off in order to enhance the contact between the tungsten gate electrode and the appropriate interconnect (see appendix D). Afterwards a forming gas step at 400° C is performed to anneal the interconnects-n^{++} silicon contact and to reduce the interface trap density of the gate oxide. Figure 5.13 shows a complete sample. A total of four devices is present which are connected by the n^{++} leads to the aluminum interconnects. A significant overlap of the gate metallization and the active device area is tolerated because in order to show the feasibility of the approach presented in this thesis only *DC* measurements are made.

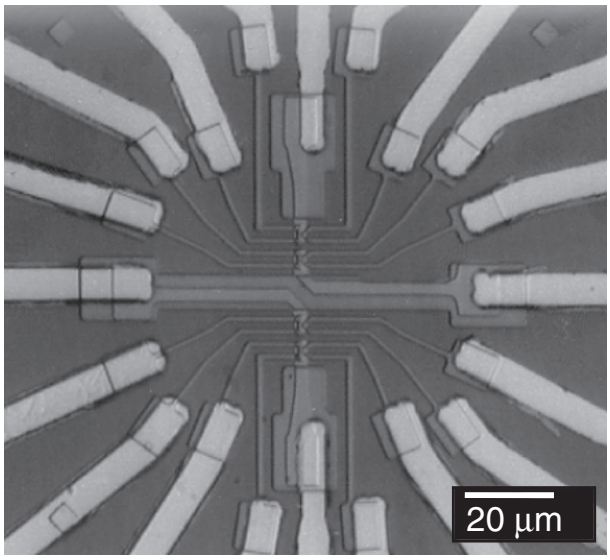

Figure 5.13: A complete device. The LTO pattern with four devices is connected to aluminum interconnects via the n^{++} leads.

5.2.4 Deviations from the Ideal Behavior

After having taken all measurements the samples are cut through in order to take TEM images[4]. These images are necessary to unambiguously determine the source-drain separation, i.e. the channel length. Moreover, they give detailed information about any deviation of the fabrication process from the ideal behavior presented at the beginning of this chapter. Figure 5.14 shows a cross-sectional TEM image of the device (compare with figure 3.1). The V-groove in between the n^{++} regions is clearly visible. To the right and to the left of source and drain the entire n^{++}/p^{-} stack is cut through as is required for device isolation. The whole structure including the tungsten gate is covered with aluminum serving to enhance the contact between gate and its belonging interconnect. Of particular interest is the behavior of the tungsten gate layer. Due to the wet chemical preparation the tungsten is cut through at convex bends (see appendix D) which results in a quasi self-aligned gate process in that the electrically relevant part of the gate in the V-groove is isolated from the rest of the tungsten gate. Obviously, in order to make use of this self-alignment, when thinking of high frequency applications, the aluminum has to be removed from the sample surface.

A closer look at the actual V-groove in figure 5.15 (a) reveals several effects concerning the interplay of the KOH etch with the epitaxial material. First, the original opening L_0 was 100 nm whereas the TEM image shows an opening of 170 nm. This cannot be explained by the mere fact that the oxide mask

[4]The samples are cut through along the dashed line in figure 5.12

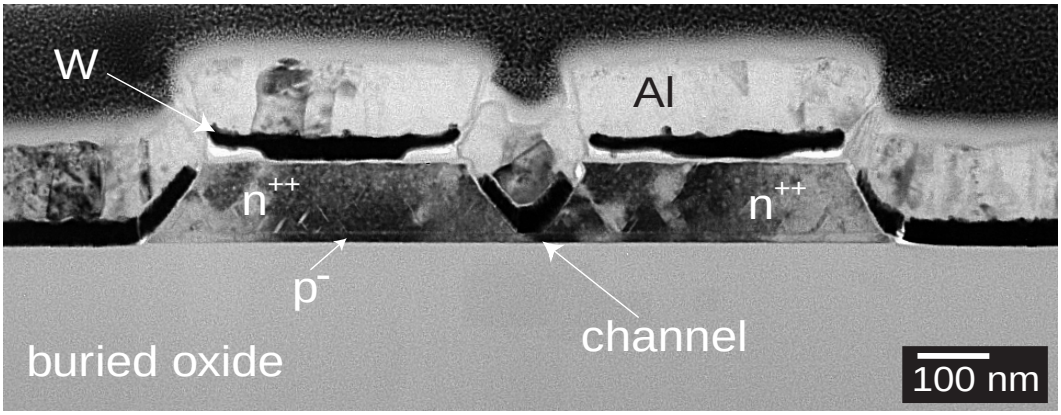

Figure 5.14: TEM image of a functioning device. The image was taken by Ph. Rice, IBM Research Division, Almaden Research Center, San Jose.

opening is wider than the negative e-beam resist mask due to the wet chemical preparation with BOE. One has to assume an increased etch rate of the {111} planes of the highly Sb doped n^{++} when compared to usual silicon. This assumption is supported by the fact that right at the V-groove opening the n^{++} material is partly etch 'inversely', i.e. there is a shallow V-groove rotated at 90°, also bounded by {111} planes. This effect always exists even in 'normal' silicon but is usually very small and appears only when etching very long. From the geometry of the V-groove in figure 5.15 compared to a V-groove in a Si (100) wafer with $L_0 = 100\,\text{nm}$ an etch rate of the highly doped material of about 2.5 nm/min is deduced when a time of ∼10 min to cut through the n^{++} is taken into account. Thus, an etch ratio of {111} planes between normal silicon and the highly doped epitaxial material of 2.5 follows. The increased etch rate of the highly doped silicon also explains the behavior of the tungsten gate at convex bends. The part of the SiO_2 mask which is undercut by the KOH plays the role of a shadow mask for the tungsten so that the tungsten is disconnected at every etch-flank.

Due to the increased etch rate of the n^{++} a source-drain separation of about 36 nm is achieved (see figure 5.15 (b)) although a V-groove mask opening of $L_0 = 100\,\text{nm}$ would in fact lead to a shortened device. The exact channel length is difficult to determine since the {111} planes are in the case of the n^{++} material not as smooth as they are with bare silicon as already discussed.

Figure 5.15 (b) shows that the p^-/n^{++} interface layer is already slightly attacked. This implies that samples with larger source-drain separations may have a relatively inhomogeneous channel area since only areas where the interface layer between n^{++} and p^- was strong enough the p^- is not cut through. Hence, only p^--islands connect source and drain in devices with larger L_0. It is therefore expected that such devices exhibit a reduced effective width.

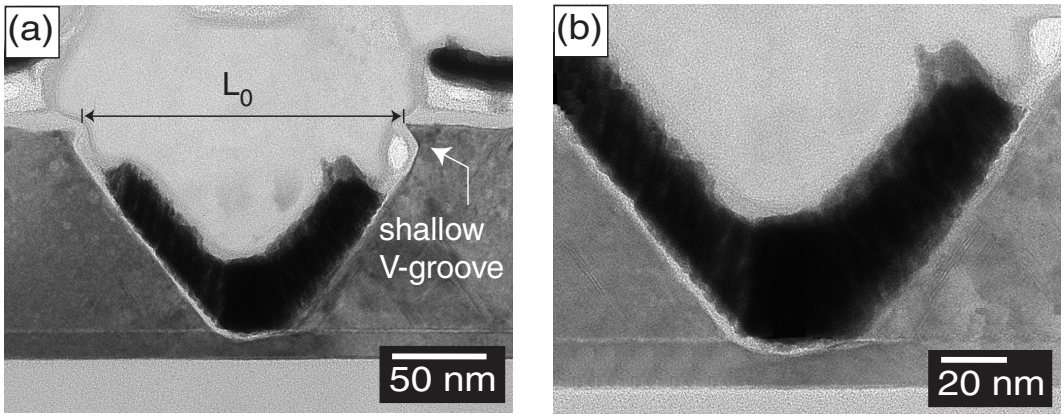

Figure 5.15: TEM image of the V-groove device. (a) shows the etch behavior of the n^{++}, (b) is a magnification. The image was taken by PH. RICE, IBM Research Division, Almaden Research Center, San Jose.

The thickness of the gate oxide is determined from a further magnification of the TEM images to be ∼3.5 nm on the V-groove flanks and ∼2.6 nm in the channel region. This results in an 1.35 higher oxidation rate of the V-groove flanks consistent with the expectation of an enhanced growth rate for highly doped (111) silicon surfaces as discussed in section 5.1.5.

5.3 How to Align the Gate?

In the actual device a significant overlap of the gate and the device is tolerated since only *DC* measurements are made to show the feasibility of the present approach. In order to facilitate high frequency applications the gate overlap has to be reduced which requires a self-aligning process for the geometrical scales considered here. As was shown above due to the wet chemical patterning of the tungsten gate such a quasi self-aligned process is automatically achieved. However, a deliberate process which ensures a small gate is desirable, in particular if one thinks of patterning the tungsten gate with e.g. reactive ion etching. A self-aligned process can be established when the etch mask for the generation of the V-groove serves at the same time as mask to pattern the gate. For the tungsten gate desired here a lift-off process is the means of choice. Figure 5.16 schematically shows how the process works in principle. Starting point is again the epitaxial silicon stack covered with SiO_2 (light-gray) and an additional layer (gray). First, this layer is patterned and afterwards the SiO_2 using the negative e-beam process described above (a). Since the oxide is etched isotropically the mask is undercut a certain amount. Next, the KOH etch step is performed followed by the gate oxidation and the evaporation of the tungsten gate layer ((b)and (c)), typically about 50 nm in thickness. The last step is a lift-off process using the additional layer as lift-off mask and the tungsten solely remains in the tip of the V-groove (d).

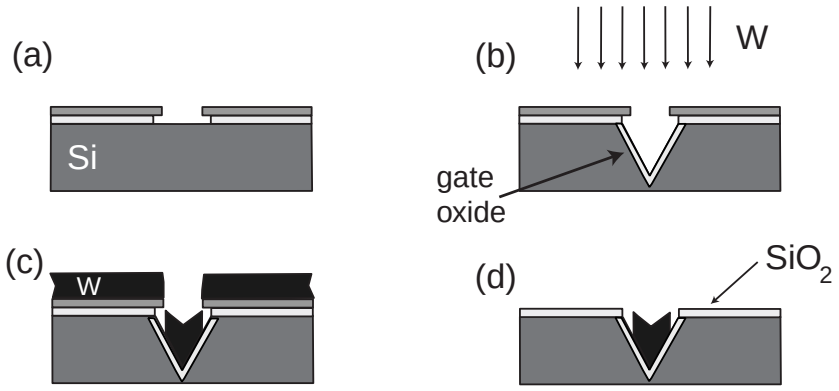

Figure 5.16: Sketch of the self-aligning gate process. An additional mask layer (gray) serves as lift-off mask after the tungsten has been evaporated on the sample surface.

The process flow of the V-groove MOSFET severely limits the materials which can be used as KOH etch- and lift-off mask at the same time. The following constraints have to be met:

1. The material must not be etched by KOH.
2. It has to withstand the wet thermal oxidation.
3. A selective etch relative to tungsten and silicon must exist.
4. It must withstand a short dip in BOE.

The most straightforward way would be to combine etch- and lift-off mask in a thick SiO_2. On the other hand the lift-off in HF could attack the gate oxide. Therefore, a two-layer etch/lift-off material system has to be used. Si_3N_4 can be used as additional layer since it can be etched by hot H_3PO_4 which attacks neither silicon nor SiO_2 (see appendix B). Furthermore, Si_3N_4 is not attacked by KOH and withstands the thermal oxidation.

However, in order to show that this process works in principle, 50 nm chromium is used as the additional layer which can be etched selectively by commercially available Cr-etch relative to SiO_2 and tungsten (see appendix B). Whereas the patterning of the chromium layer is done with Cr-etch at room temperature (etch rate approximately $\sim$120 nm/min) the lift-off process is done with Cr-etch heated to 50° C which very rapidly etches the chromium thereby washing away the tungsten on top of it as well. The results of the process are shown in figure 5.17 (a) and (b). Both images show a V-groove with the SiO_2 mask to the left and right. In the V-groove there is tungsten which was left there after the Cr-lift-off. The borders of the tungsten are rather rough which is attributed to two facts. First, the chromium mask which serves as shadow mask for the tungsten in the V-groove region has itself rough borders due to

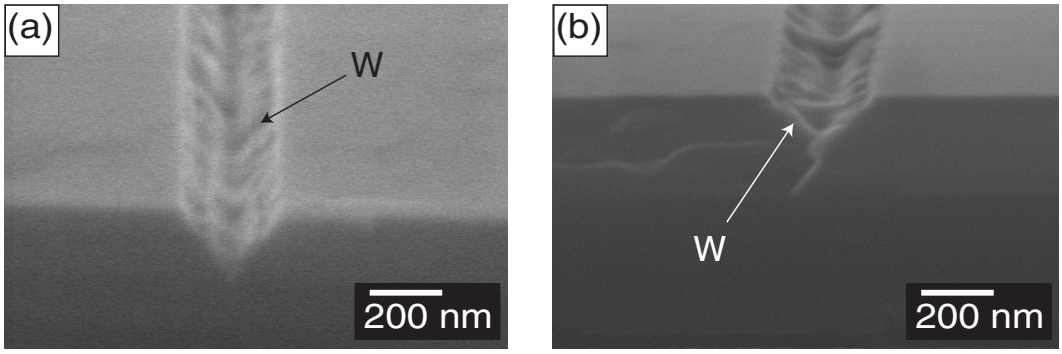

Figure 5.17: The self-aligning gate process. Both images show a tungsten gate in the V-groove patterned with a lift-off technique using chromium as mask. The images were taken at a tilt angle of 70°.

the wet chemical patterning; a process using reactive ion etching would be advantageous here. Second, the tungsten is, although only slightly, attacked by the hot chromium etch (see appendix B).

Of course the process in this form cannot be used for the actual transistor because the gate oxidation has to be carried out with the chromium layer on top of the sample. This would deteriorate the Cr during the gate oxidation so that it cannot be used for the lift-off process afterwards. On the other hand the self-aligning process realized above shows a way of how to fabricate an appropriate gate for the V-groove MOSFET. This approach has to be transferred to Si_3N_4 for the actual V-groove device in order to facilitate high-frequency applications for future samples. Technological problems, in particular the patterning of the Si_3N_4 mask, remain to be solved in order to use the self-aligned process for device application.

5.4 Results

In conclusion, the findings of the present chapter can be summarized as follows

- A technology for the fabrication of ultra-short channel MOSFETs has been developed and successfully accomplished.
- The material system can be fabricated with sufficient crystalline quality with the desired thickness. V-grooves with satisfying shape can be etch into this material system.
- A moderate interface layer is favorable for the fabrication of flat channels and inevitable when preparing MOSFETs on ultra-thin SOI films. Device isolation is possible with such an interface layer.
- The n^{++} material shows different etch rates in KOH than usual silicon. In

particular, the {111} planes are etched faster with a rate of approximately 2.5 nm/min.

- Devices with 36 nm source-drain separation have been fabricated.
- A way to fabricate a self-aligned tungsten gate has successfully been demonstrated.

6 Measurements and Results

After the fabrication, the samples are characterized by $I_d - V_{ds}$ and $I_d - V_g$ measurements. All measurements are performed with a HEWLETT PACKARD semiconductor parameter analyser and are presented in the following. Functioning devices were obtained for samples with original V-groove openings of 100, 110, and 120 nm. For larger openings the p^- was cut through and no transistor action could be measured. The effective width of the devices is extracted from the 100 nm device and is 700 nm. For convenience, the geometrical data of the samples is displayed in table 6.1. Measurements presented in this section

t_{si}:	15 nm	t_n:	105 nm
t_{ox}:	2.6 nm	t_{ox} on V-groove flank:	3.5 nm
L_0:	100 .. 120 nm	$W =$	700 nm

Table 6.1: Geometrical Data

are representative for devices having the same V-groove opening. Although the actual channel length was determined with TEM images only for the transistor with $L_0 = 100$ nm to be 36 nm we anticipate channel lengths of 46 nm and 56 nm for the devices with V-groove opening $L_0 = 110$ and 120 nm. This is a reasonable assumption because the TEM images (see figure 5.15) showed a regular V-groove formation so that the channel length variation is according to the variation of L_0. The devices are therefore referred to according to their channel length L. Four terminal measurements were performed in order to obtain the intrinsic terminal voltages which are otherwise masked by the parasitic source and drain resistances caused by the respective n^{++} leads. Output and transfer characteristics were measured at room temperature as well as low temperatures. One of the major benefits of the V-groove approach is that MOSFETs with different channel lengths can be fabricated reliably and reproducibly. Because of the unique combination of a self-limiting generation of the device with an abrupt doping profile the V-groove concept allows the fabrication of samples with a well-defined channel length and channel length variation in ~ 10 nm steps. This is a major advantage when compared to other ultra-short-channel MOSFETs relying on ion implantation or annealing techniques.

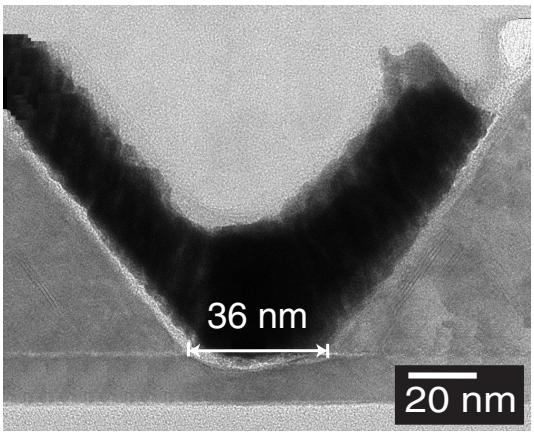

Figure 6.1: TEM image of a device with $L_0 = 100\,\text{nm}$. A channel length of ∼36 nm can be taken from the image. This is the shortest channel length of devices fabricated with the V-groove approach to date. The image was taken by PH. RICE, IBM Research Division, Almaden Research Center, San Jose.

6.1 Room-Temperature

For the room-temperature measurements a special probe station equipped with six probes and the capability of flushing the sample space with nitrogen was used. After flushing the sample 24 h in dry nitrogen four terminal measurements were done in a crossed configuration, as indicated in figure 5.12. This configuration ensured the most homogeneous current flow and was therefore used for all measurements. During the measurements the leakage current through the gate oxide and the BOX was monitored. It never exceeded values of 10 - 100 pA. The $I - V$ characteristics shown below are representative for different devices having the same L_0.

6.1.1 I-V Characteristics

As already mentioned, the present approach allows the fabrication of devices with defined channel length and channel length variation. This enables the investigation of the geometrical influence of the samples on the electrical behavior. Figure 6.2 shows the $I_d - V_{ds}$ as well as the $I_d - V_g$ curves of samples with L =56, 46 and 36 nm from top to bottom; the range of gate voltages is as indicated in the figure with steps of 0.1 V. The actual figures of merit are listed in table 6.2 for all three devices. Two different values for the threshold voltage are given in this table depending on the way this entity is extracted from the measurements. From a linear extrapolation of the $I_d - V_g$ curves the threshold voltage is extracted which is referred to as V_{th}^1 in the table. The alternative method is to extract the threshold voltage from the linear regime of the output characteristics (see below). Values found with this method are higher and are

referred to as V_{th}^2. The DIBL can be quantified for drain-source voltages in the saturation region at $V_g \approx 0\,\mathrm{V}$ as described at the end of section 2.3.1.

The output characteristics show a consistent evolution of devices with decreasing channel length. The drive current and the transconductance increase with decreasing channel length whereas the threshold voltage (V_{th}^1 in table 6.2), the off-current I_{off} and the DIBL increase at the same time. Furthermore, the subthreshold swing gets larger for smaller channel length. The output and transfer characteristics (a) to (f) show the continuous evolution from a device with rather long-channel behavior to a transistor exhibiting short-channel effects. Current saturation can only be observed for the longest device (a). The shortest device has a significant DIBL almost twice as much as the other two samples.

	56 nm	46 nm	36 nm
drive current	$93\,\mu\mathrm{A}/\mu\mathrm{m}$	$214\,\mu\mathrm{A}/\mu\mathrm{m}$	$490\,\mu\mathrm{A}/\mu\mathrm{m}$
S	80 mV/dec	90 mV/dec	160 mV/dec
g_m at $V_{ds} = 1\,\mathrm{V}$	$220\,\frac{\mathrm{mS}}{\mathrm{mm}}$	$570\,\frac{\mathrm{mS}}{\mathrm{mm}}$	$900\,\frac{\mathrm{mS}}{\mathrm{mm}}$
V_{th}^1	0.32 V	0.26 V	0.22 V
V_{th}^2	0.53 V	0.51 V	0.5 V
DIBL at $V_g \approx 0$	128 mV/V	171 mV/V	314 mV/V
I_{off} at $V_g = 0, V_{ds} = 0.4\,\mathrm{V}$	1.1×10^{-11} A	4×10^{-10} A	2×10^{-8} A

Table 6.2: Figure of merits of three V-groove MOSFETs.

Consider the curves in figure 6.2 and (b). The $I_d - V_g$ curve is plotted for several drain-source voltages in steps of 0.1 V. In the subthreshold region the curves for different V_{ds} are almost the same. This means that the switching behavior of such a device is only slightly dependent on the drain-source voltage. Hence a well-defined off-state exists as is expected for a transistor with long-channel behavior. This is reflected in the value for the subthreshold swing $S = 80\,\mathrm{mV/dec}$ and DIBL of 128 mV/V. The device has a transconductance of $g_m = 220\,\mu\mathrm{S}/\mu\mathrm{m}$ at $V_{ds} = V_g = 1\,\mathrm{V}$. The increase of I_d for $V_g < -0.2\,\mathrm{V}$ is typical of fully depleted SOI devices. For large negative V_g holes in the valence band can travel through the channel region from drain to source, resulting in an increased I_d [61]. Figure 6.2 (c) and (d) show output and transfer characteristics for a device with a nominal V-groove opening of 110 nm, i.e. anticipated channel length of 46 nm. The following observations can be made.

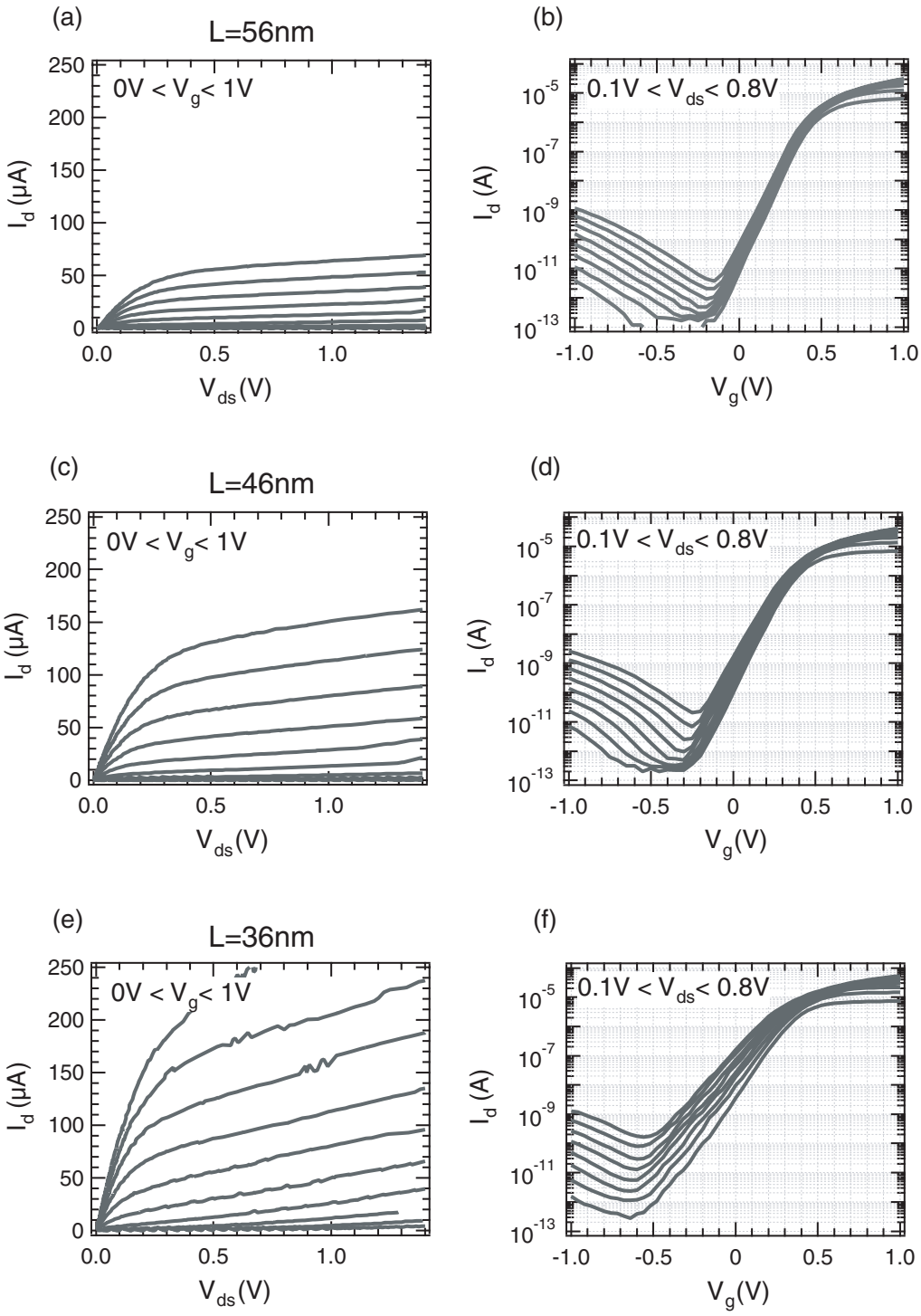

Figure 6.2: Output and transfer characteristics of devices with $L = 36$, 46 and 56 nm. The range of applied voltages is as indicated in the figures with $\Delta V_g = \Delta V_{ds} = 0.1\,\mathrm{V}$.

The drive current is increased to 214 μA/μm, the transconductance at $V_{ds} = V_g = 1\,\mathrm{V}$ is now $g_m = 570\,\mu\mathrm{S}/\mu\mathrm{m}$. The smaller channel length results in a stronger appearance of short-channel effects which is reflected by the strong V_{ds} dependence of the subthreshold transfer characteristics; the DIBL is raised to 171 mV/V. From the linear $I_d - V_g$ plot a slightly lower threshold voltage $V_{th}^2 = 0.51\,\mathrm{V}$ than for the $L = 56\,\mathrm{nm}$ device is found. The transfer characteristics figure 6.2 (d) show a subthreshold swing of $S = 90\,\mathrm{mV/dec}$.

The shortest device fabricated with the V-groove approach to date is the 36 nm device. Output and transfer characteristics are shown in figure 6.2 (e) and (f). The device exhibits SCE which is reflected in the loss of current saturation in the output characteristics. In addition, the transfer characteristics also show a device with short-channel behavior. The subthreshold curves exhibit a significant dependence on V_{ds}, a result of the strong DIBL in this device, reflected in the numerical value of 314 mV/V. The SCE lead to a worse gate-control in the subthreshold region resulting in a relatively large subthreshold swing $S = 160\,\mathrm{mV/dec}$. V_{th}^2 is lowered only slightly to 0.5 V.

Geometrical Issues

Figure 6.3 shows the transfer characteristics of the three devices for high bias. The slope of the curves for the highest V_g (see figure 6.3) is used to extract the transconductance of the devices which is divided by the width $W = 700\,\mathrm{nm}$. Looking at the numerical g_m values listed in table 6.2 a strong decrease of transconductance is observed. These values were obtained with the assumption of an equal width W for all three devices. However, the strong decrease cannot be explained by the channel length variation alone. As mentioned above the regular V-groove formation during the fabrication of the devices justifies the assumption of a channel length variation in 10 nm steps. Additionally, we exclude a statistical variation of device geometry from device to device being responsible for the decrease of g_m since samples with the same V-groove opening show the same electrical characteristics. Therefore, it is believed that the strong decrease in transconductance is due to a reduction of the effective width of the devices with larger V-groove opening L_0. The behavior of the KOH etch and the interplay with the interface layer present between n^{++} and p^- was already discussed in chapter 5. It was found that with increasing L_0 it is more likely that the KOH cuts through the interface layer. Since the interface layer is inhomogeneous, source and drain in devices with large L_0 will be connected only by p^--islands thereby reducing the effective width of the device.

The idea of p^--islands is supported by an analysis where the channel length is extracted from the linear regime of the output characteristics. For low V_{ds} and $V_g > V_{th}$ the MOSFET behaves like a resistor whose resistance should decrease with V_g or conversely, its conductance G_{ds} should raise linearly with V_g. If G_{ds} versus V_g is plotted for a V_{ds} in the linear regime, one observes that instead of the expected linear increase the curves tend to saturate for larger V_g. This happens if an additional series resistance R_p between source and channel or channel and drain exists which is not or only slightly influenced

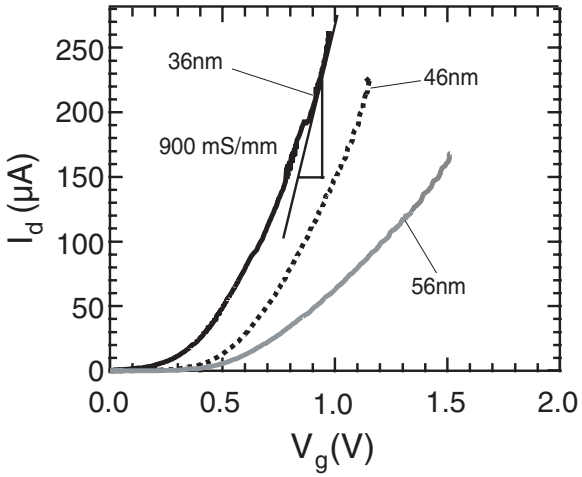

Figure 6.3: Transfer characteristics of the devices with $L = 36$, $L = 46$ and $L = 56\,\mathrm{nm}$ with $V_{ds} > V_{sat}$.

by the gate voltage. Then, as the intrinsic conductance of the channel raises with increasing V_g the overall conductance is more and more dominated by the additional resistance. Such a parasitic resistance can consist of the usual spreading resistance [9], the fundamental contact resistance already mentioned in section 2.4.1 [19] and/or a resistance due to the particular device geometry or fabrication. If the resistance R_p is subtracted from the $G_{ds} - V_g$ curves a straight line is obtained when the correct value of the parasitic resistance R_p is encountered. This is shown in figure 6.4 (a) which shows a $G_{ds} - V_g$ plot. The gray curves are the measured curves whereas the black points are the belonging corrected values which lie on a straight line. The black points are obtained by subtracting a varying test resistance from the $G_{ds} - V_g$ curves until they yield a straight $G_{ds} - V_g$-line. Thus, the numerical value of the parasitic resistance as well as the intrinsic conductance of the channel is obtained from the linear dependence of G_{ds} on V_g.

Having found the intrinsic values of G_{ds} the channel lengths of the devices can be estimated as follows in order to check the assumption of a channel length variation in 10 nm steps. If for the moment an equal width for the three devices is assumed, then R_p should be equal for all three devices. The numerical value of $R_p \approx 300\,\Omega$ is determined from the curve belonging to the 36 nm device and is subtracted from the $G_{ds} - V_g$ curves of all three devices. If furthermore diffusive transport in the channel is assumed so that the intrinsic conductance of the channel should scale as $1/L$ and knowing the channel length of the shortest device the channel lengths of the other two samples can be obtained. The ratio of the intrinsic conductances of two devices (taken with $V_g = 1\,\mathrm{V}$ from the corrected $G_{ds} - V_g$ plots) scales reciprocal to their channel lengths, e.g. $G_{36}/G_{46} = L^{46}/L^{36}$. With $L^{46} = L^{36} + \Delta L^1$ and $L^{56} = L^{36} + \Delta L^1 + \Delta L^2$ and

the values of G_i (i = 36, 46, 56) the equations can be solved to give $\Delta L^{1,2}$. Channel lengths of $L^{36} = 36\,\text{nm}$, $L^{46} = 63\,\text{nm}$ and $L^{56} = 188\,\text{nm}$ follow what is contrary to the assumption of channel lengths stated above. Additionally, the latter channel length of the device with $L_0 = 120\,\text{nm}$ is far too large to yield a functioning device. The KOH would cut through the whole silicon stack for such source-drain separations. Hence, a variation in the width of the devices must be responsible for the reduction in conductance with increasing V-groove opening L_0.

Therefore, a closer inspection of the $G_{ds} - V_g$ curves is necessary. Since we have evidence for a decreasing width W with increasing channel length the parasitic resistance R_p should increase as well. The following is done in order to get an estimation for the variation of W.

1. Due to the reasons mentioned at the beginning of this chapter we take the channel lengths to be 36, 46 and 56 nm.

2. The width of the 36 nm device is 700 nm taken from the SEM image 5.12.

3. The resistance R_p must scale as $1/W$. For example, the spreading resistance R_{sp} and the fundamental contact resistance R_c are given by the following expressions [9, 19]

$$R_{sp} = \frac{2\rho_j}{\pi W} \ln\left(0.75\frac{x_j}{x_c}\right) \qquad R_c = \frac{\pi\hbar}{e^2 W}\sqrt{\frac{\pi}{2n_s}}. \tag{6.1}$$

   where ρ_j and x_j are the resistivity and depth of the contacts and x_c is the depth of the channel; n_s is the $2D$ surface carrier density. Hence, plotted versus $1/W$ all values found for R_p should lie on a straight line which in addition has to cross the origin

4. The intrinsic conductances G_{ds} scale as W/L.

The analysis proceeds as follows: For all three devices a maximum and minimum value of R_p is subtracted from the respective $G_{ds} - V_g$ curve such that the resulting curve is still linear. In addition, a value in between is subtracted, too. Hence, we get three R_p values and the corresponding intrinsic conductances G_{ds}. Since we already know the width of the 36 nm device the values of R_p for this device can be drawn in a plot $R_p - 1/W$ which are the black circles in figure 6.4 (b). Because all R_p values must lie on a straight line which crosses the origin in this plot the gray-shaded cone represents the area where possible pairs of $R_p - W$ values can lie. Hence, with the ratio $G_{36}/G_{46} = W_{36}/W_{46} \times L_{46}/L_{36}$ and equivalently for the 56 nm device the widths of the 46 nm and 56 nm samples, $W_{46,56}$, can be calculated for all nine combinations of G_{36}/G_{46} and is then plotted in the $R_p - 1/W$-plot. The dark-gray circles are obtained for the 46 nm device and the light-gray circles for the 56 nm device. All circle fall into the area of the cone so that on principle all these values are possible ones. Table 6.3 displays the values obtained for the devices with this analysis. Additionally, the recalculated values of the transconductances g_m with the new widths are given as well.

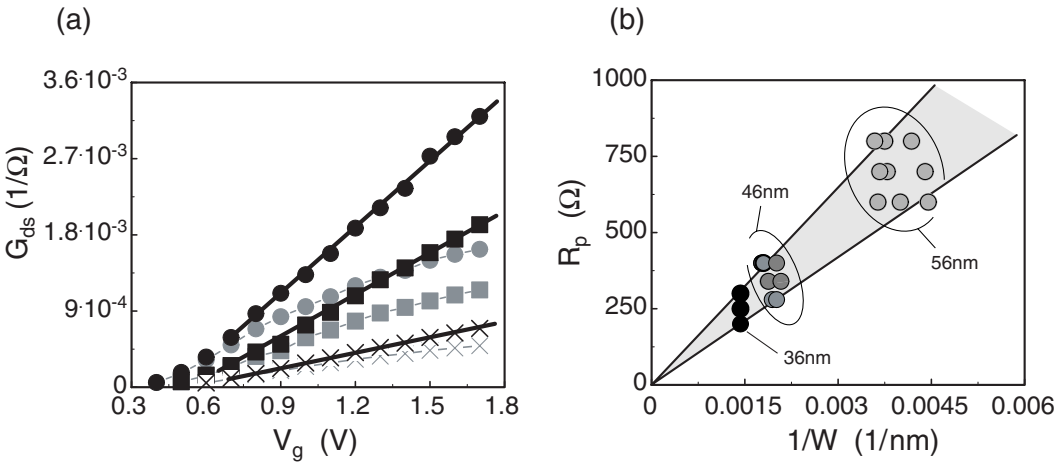

Figure 6.4: (a) Measured conductance (gray curves) of the three devices in the linear regime. The resistance R_p as well as the intrinsic conductances are obtained from the corrected curves (black). (b) Possible pairs of $R_p - W$ values for the three devices.

	56 nm	46 nm	36 nm
R_p	600..800 Ω	280..340 Ω	200..300 Ω
W	230..280 nm	500..560 nm	700 nm
g_m at $V_{ds} = 1$ V	550..670 $\frac{\text{mS}}{\text{mm}}$	710..800 $\frac{\text{mS}}{\text{mm}}$	900 $\frac{\text{mS}}{\text{mm}}$

Table 6.3: Parasitic resistances and recalculated widths and transconductances of the devices.

The recalculated values for g_m show a much more reasonable decrease with increasing channel length, consistent with the anticipated channel length variation.

The corrected curves in figure 6.4 (a) also allow the extraction of the threshold voltage. The intercept of the linear fit with the V_g-axis determines V_{th}. The numerical values are given in table 6.2 referred to as V_{th}^2.

Comparison with Theory

When comparing the experimental results with the simulations presented in chapter 4 the first thing to mention is the transconductance. The analysis of the previous paragraph has revealed that g_m decreases with increasing channel length. This is the typical behavior of MOSFETs with diffusive carrier transport through the channel. The simulations on the other hand have shown that

for ballistic transport g_m remains rather constant or even shows the reverse dependence on channel length if the SCE become important. Therefore, at least the 46 nm and the 56 nm device exhibit diffusive transport and deviations between simulated and experimental curves are expected.

Figure 6.5 shows a comparison between simulated output characteristics (a) and the characteristics of the 36 nm device (b). The parameters for the simulation as described in chapter 4 are extracted from the electrical measurement and the structural information obtained by the TEM images. $N_d = 6\times10^{19}\,\mathrm{cm}^{-3}$ $\rightarrow E_f = 53\,\mathrm{meV}$ (see section 6.2), $W = 700\,\mathrm{nm}$, $t_{si} = 9\,\mathrm{nm}$ and $t_{ox} = 2.6\,\mathrm{nm}$. t_{si} was taken as the minimum thickness of the p^- silicon layer as can be seen in figure 6.1. The calculated I_d of the output characteristics is corrected to

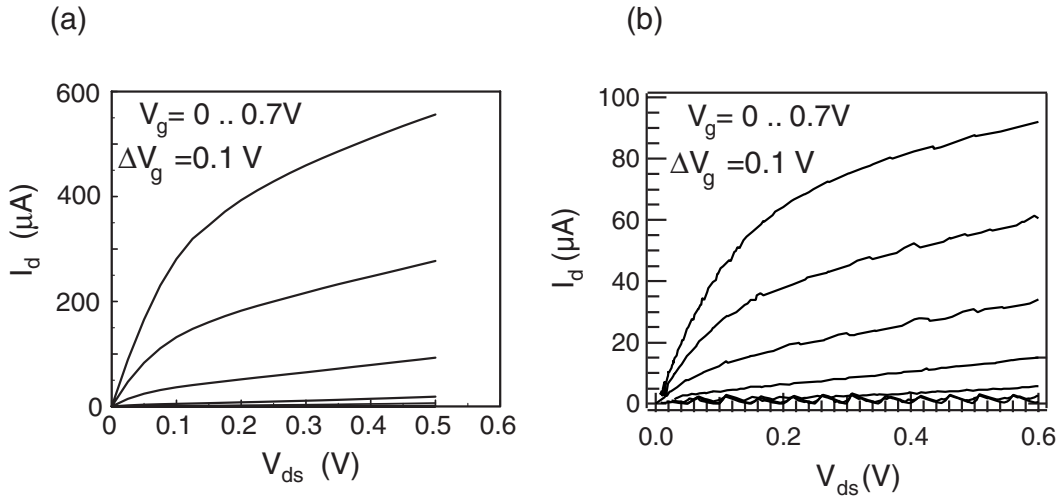

Figure 6.5: Comparison between simulated (a) and experimental (b) output characteristics.

account for a total parasitic resistance of 300 Ω. The simulated curves show a very similar behavior as the experimental ones although the drive current and transconductance is higher. The following fact has to be considered. Although the drain current has been corrected according to a parasitic resistance of 300 Ω the gate voltage has not. The displayed V_g is the intrinsic gate voltage which is smaller than the external one if there is a significant parasitic source resistance. Such a resistance degrades the gate drive according to $V_g^{intrin} = V_g^{ext} - R_{source} I_d$. In the simulated curve in figure 6.5 (a) this means that for e.g. the curve belonging to $V_g = 0.6\,\mathrm{V}$ the external gate voltage is ∼0.64 V when the source resistance is taken to be $R_p/2 = 150\,\Omega$. From the plot in (b) we estimate a current of 70 μA at $V_g = 0.64\,\mathrm{V}$ so that the simulated device exhibits ∼4 times the drive current of the experimental device. The transconductance, $g_m = 3690\,\mathrm{mS/mm}$, is a factor of four larger, too.

Inelastic scattering of carriers in the channel due to e.g. electron-electron and electron-phonon interaction [25] neglected so far must be included in the simulations in order to get a conformity with the experimental curves. However,

ballistic carrier transport in MOSFETs is predicted to occur for L below 50 nm at 77 K [49] in undoped channels so that we believe that the simulations will describe the transport in our shortest devices properly at lower temperatures.

6.1.2 Where are we?

In order to appreciate the device characteristics obtained with the V-groove approach the parameters of the $L = 36$ nm device are displayed in table 6.4 together with values published by other groups most recently [83, 38, 13]. The

	Intel	NEC	UCB	V-groove
gate length	30 nm	24 nm	15 nm	36 nm
g_m	1200 $\frac{\text{mS}}{\text{mm}}$	1000 $\frac{\text{mS}}{\text{mm}}$	400 $\frac{\text{mS}}{\text{mm}}$	900 $\frac{\text{mS}}{\text{mm}}$
t_{ox}	8 Å	25 Å	40 Å	26 Å
drive current	514 $\frac{\mu\text{A}}{\mu\text{m}}$ at 0.85 V	796 $\frac{\mu\text{A}}{\mu\text{m}}$ at 1.2 V	190 $\frac{\mu\text{A}}{\mu\text{m}}$ at 1.2 V	490 $\frac{\mu\text{A}}{\mu\text{m}}$ at 1.1 V
V_{th}	0.3 V	0 V	-0.1 V	0.5 V
S	100 $\frac{\text{mV}}{\text{dec}}$	180 $\frac{\text{mV}}{\text{dec}}$	150 $\frac{\text{mV}}{\text{dec}}$	160 $\frac{\text{mV}}{\text{dec}}$
I_{off}	100 $\frac{\text{nA}}{\mu\text{m}}$	300 $\frac{\text{nA}}{\mu\text{m}}$	80 $\frac{\text{nA}}{\mu\text{m}}$	280 $\frac{\text{nA}}{\mu\text{m}}$

Table 6.4: Comparison with other state-of-the-art devices.

electrical characteristics of the V-groove MOSFETs are comparable to state-of-the-art devices. Since the geometrical parameters like t_{ox} and gate length directly influence the behavior of the devices a closer inspection is needed in order to classify the V-groove MOSFET. For instance, the thinner the gate oxide the better the control of the charge in the channel. Therefore, the transconductance as well as the drive current should be increased for decreasing gate oxide thickness. Hence, our values compared with the ones from INTEL are excellent because t_{ox} of the V-groove MOSFET is three times larger than that of the INTEL device. Moreover, in order to compare our device with the NEC sample the gate overdrive $V_g - V_{th}$ has to be considered instead of V_g. Thus, an overdrive of 0.6 V yields a drive current of 490 μA/μm in comparison to 796 μA/μm at an overdrive of 1.2 V for the NEC device. Furthermore, the channel length of the V-groove MOSFET is the longest of the four devices. Hence, although some of the data found by the other groups are better the V-groove MOSFET has an enormous potential to outperform the data presented above. The present V-groove MOSFET has an minimum SOI thickness of ∼9 nm and a $t_{ox} = 2.6$ nm. As was discussed in section 2.3.3 decreasing t_{si} yields improved electrical behavior concerning SCE due to an increased portion of gate-controlled charge in the channel. Additionally, a reduction of t_{ox} also results in a suppression of

SCE and increases the drive current. In order to estimate the improvements that can be obtained by simply reducing t_{si} to 5 nm and t_{ox} to 1.5 nm keeping the channel length at 36 nm, simulations are performed. Figure 6.6 (a) and (b) show the output and transfer characteristics for a device with $t_{si} = 9$ nm and t_{ox} =2.6 nm; (c) and (d) refer to the devices with $t_{si} = 5$ nm and t_{ox} =1.5 nm.

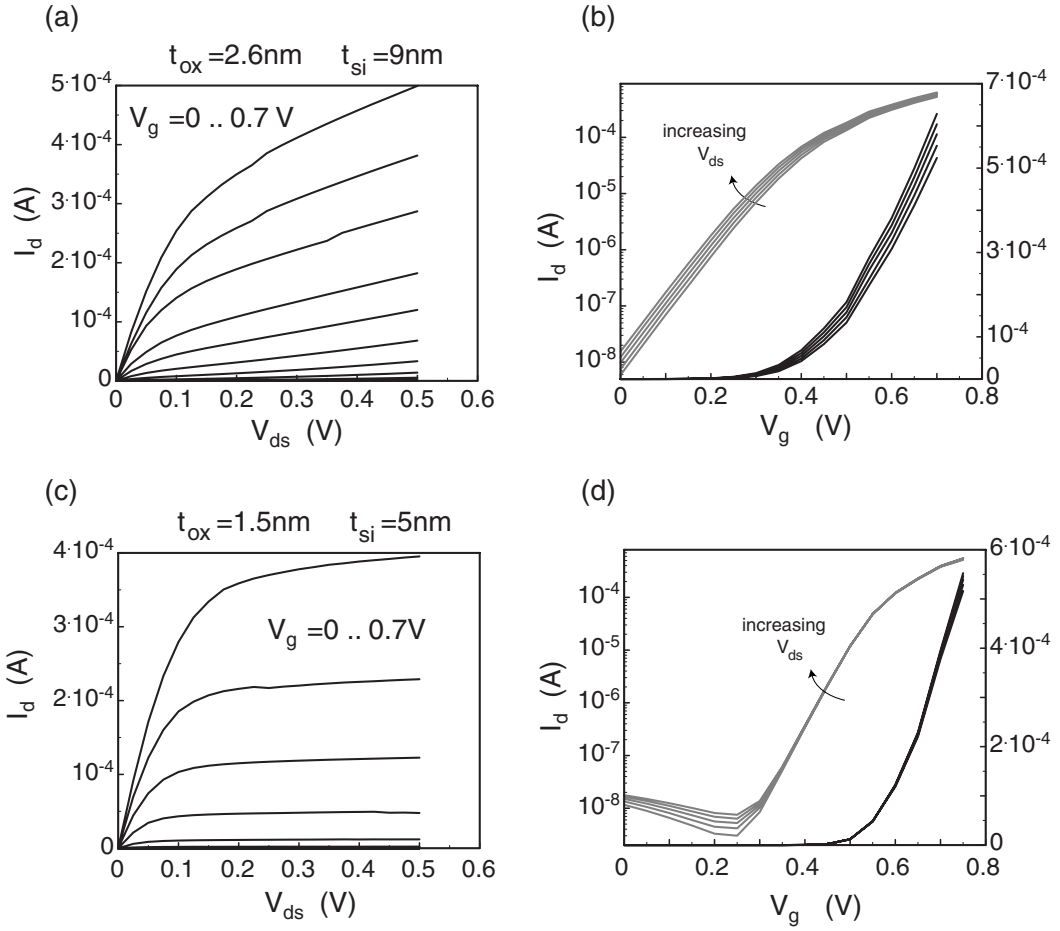

Figure 6.6: Simulated output and transfer characteristics for the 36 nm device with $t_{ox} = 2.6$ nm, $t_{si} = 9$ nm ((a) and (b)) and with $t_{ox} = 1.5$ nm, $t_{si} = 5$ nm ((c) and (d)). The gate voltage changes in steps of 0.05 V.

The obvious improvement when looking at the characteristics of the two devices is confirmed by the numerical values of g_m, S and V_{th}. These values change from the thick-channel to the thin-channel device as follows. g_m : 3690 $\frac{\mu S}{\mu m} \rightarrow 4630 \frac{\mu S}{\mu m}$, $S : 100 \frac{mV}{dec} \rightarrow 67 \frac{mV}{dec}$ and $V_{th} : 0.36\,V \rightarrow 0.55\,V$ and DIBL: 130 mV/V $\rightarrow$ 14 mV/V. The subthreshold slope for the improved device has almost the ideal room temperature value which shows the excellent gate control in the off-state of the device. DIBL is suppressed very efficiently what is reflected in the numerical value and can be seen in the transfer characteristics.

In order to compare the drive currents only the gate overdrive $V_g - V_{th}$ is considered which yields values of $6.3{\times}10^{-4}\,\mathrm{A}$ at $0.7\,\mathrm{V}{-}V_{th} = 0.34\,\mathrm{V}$ and $4{\times}10^{-4}\,\mathrm{A}$ at $0.7\,\mathrm{V}{-}V_{th} = 0.15\,\mathrm{V}$, respectively. This means, that the device with smaller $t_{si,ox}$ has a drive current almost equal to the device with $t_{si} = 9\,\mathrm{nm}$ although the gate overdrive is less than half as large. The transfer characteristics in (d) show a saturation of I_d for small V_g due to tunneling from the gate to the channel. Therefore, the off-currents of both devices are approximately equal.

The electrical behavior is significantly improved for the thin-channel, thin-oxide device in all important figures of merit. Hence, decreasing $t_{si,ox}$ allows to reduce the channel length without deteriorating the performance of the device by severe SCE. The roll-off of V_{th} accompanying the reduction of the channel length is desirable here in order to get a gate overdrive large enough for high drive currents using low voltage power supplies[1]. A reduction of channel length will further improve the characteristics of the real devices since scattering of carriers in the channel is more and more suppressed.

All in all, a significant improvement of the already excellent device characteristics obtained so far is expected when the silicon and oxide thickness will be reduced. Both, smaller t_{si} as well as t_{ox} have already been successfully demonstrated.

6.2 Low-Temperatures

Temperature dependent measurements of the $I - V$ characteristics are taken in order to study the behavior of the MOSFET in more detail. In particular, potential barriers at source and drain - if present - should have an impact at low temperatures, although they are obviously small enough that they do not disturb electron transport at room temperature. For this purpose devices are used which were fabricated according to the technology presented in the previous chapter from the same wafer. The difference between the two types of samples is that for the latter we used a lower concentrated KOH. This KOH solution does not attack the interface layer that much as the higher concentrated does. Thus, devices with significantly longer channel length have been produced. As mentioned above, for devices with longer channels there are only p^--islands connecting source and drain which is also true for the devices presented here. Hence, the effective width is significantly reduced and the devices used for the low temperature measurements show a much lower drive current than the devices presented so far. Since the samples were fabricated from the same wafer we believe that the findings presented in this section can be transferred to the former devices.

First of all, the electron density and mobility of the n^{++} silicon layer is characterized with longitudinal transport and HALL-effect measurements. Hallbar structures made of the n^{++} material are patterned together with the transistor structures. Measurements were done in a ^{3}He evaporation cryostat. It is found

[1] Alternatively, a gate metal with smaller work function can be used to tune V_{th}.

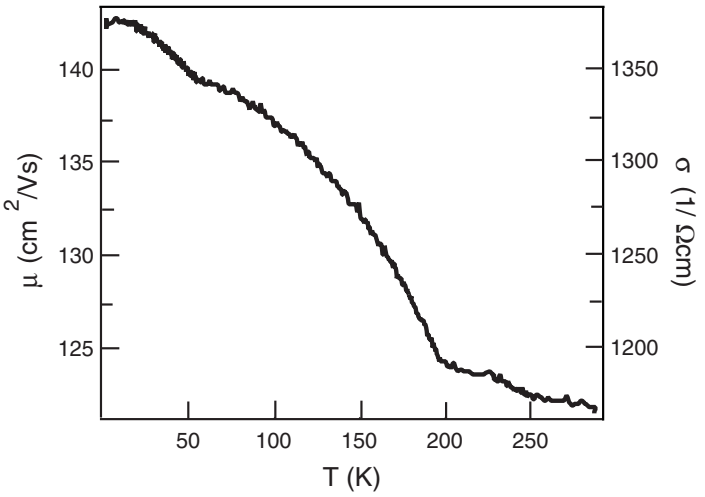

Figure 6.7: Temperature-dependent mobility and conductance of the n^{++} epitaxial silicon.

that in the temperature range between 0.3 and 80 K the density of carriers is actually constant at a value of $6 \times 10^{19}\,\mathrm{cm}^{-3}$. We therefore assume a constant density, i.e. fully ionized donors, over the entire temperature range from 0.3 - 300 K. The actual doping density therefore is about $6 \times 10^{19}\,\mathrm{cm}^{-3}$, a bit smaller than originally intended.

Figure 6.7 shows the temperature dependent mobility of the n^{++} and the conductance which are related by $\sigma = en\mu$, $\propto \mu$ for constant n. The values for the carrier density and conductance are in agreement with values found in literature [73]. Furthermore, the temperature dependent source-source and drain-drain resistance is measured. This is necessary because the low temperature measurements could only be done in two-terminal configuration since our low-temperature equipment did not allow four terminal measurements. The measured values of the source-source/drain-drain resistance are divided by 2 due to the symmetry of the layout in order to get the parasitic source and drain resistance. It was found that both resistances are almost temperature independent with a value of 20 kΩ.

6.2.1 I-V Characteristics

The low temperature measurements are performed in a liquid nitrogen cooled probe station capable of cooling the samples down to 80 K. Output characteristics are taken between 80 K and 200 K. In the range of the output characteristics considered here, the two terminal resistance is so large that the 20 kΩ parasitic resistances of the leads can actually be neglected. The data is shown in figure 6.8 for $T = 200, 170, 130, 100$ and 80 K (from (a) to (e)). The range of V_g is indicated in the diagrams. At 170 K the output characteristics show the onset

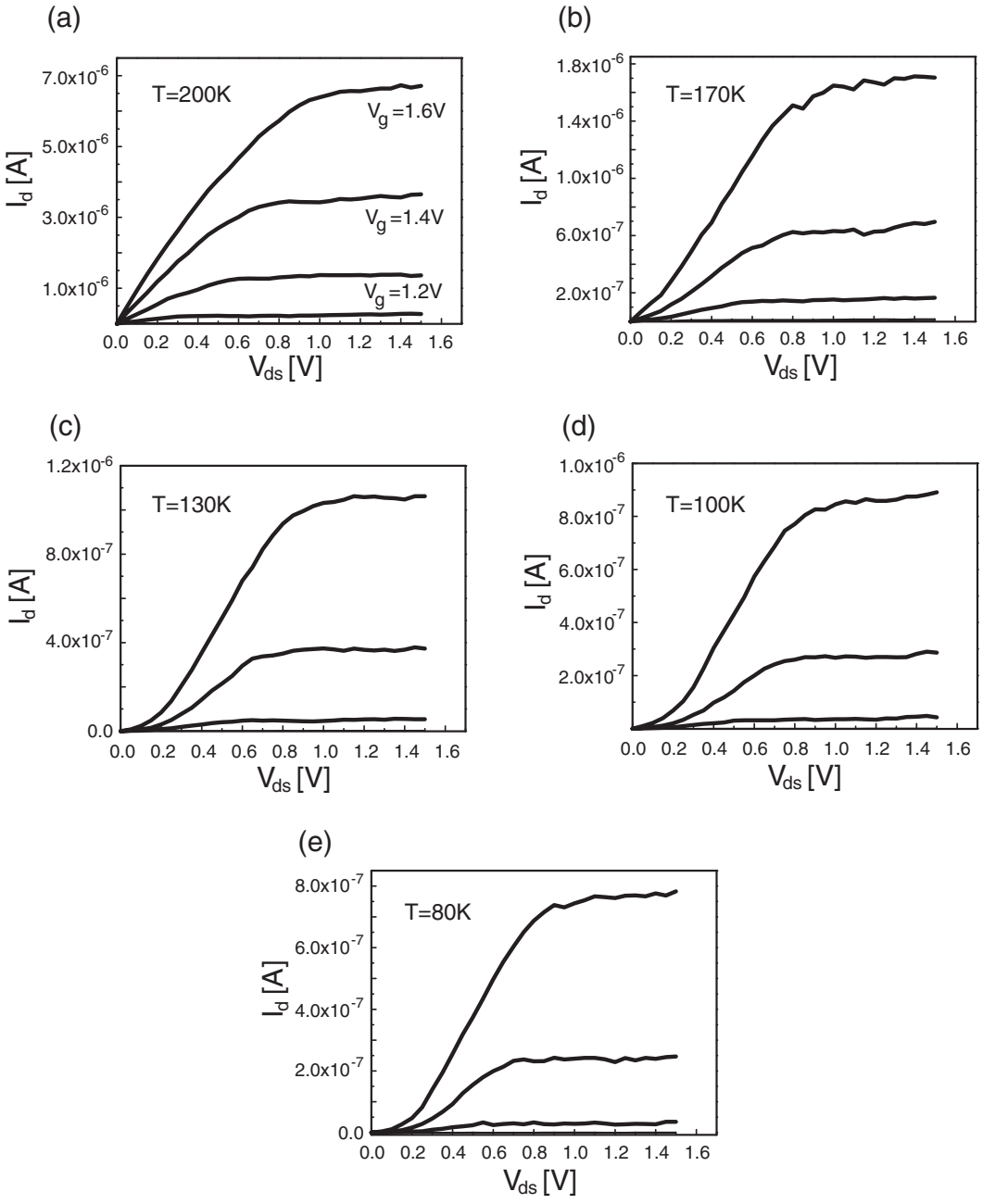

Figure 6.8: Low temperature output characteristics. (a) 200 K, (b) 170 K, (c) 130 K, (d) 100 K and (e) 80 K. The range of gate voltages is as indicated in (a) for all diagrams.

of an exponential increase in I_d for low bias. An exponential increase in I_d for small bias is typical of two reversibly biased SCHOTTKY diodes [91]. Hence, this is a clear signature for the existence of potential barriers at source and drain. These barriers are such that they do not play a significant role at higher temperatures since the exponential behavior in the output characteristics vanishes for temperatures exceeding ∼170 K. In addition, for gate voltages $V_g > 2\,\mathrm{V}$ the exponential behavior vanishes as well.

The output characteristics can be used to estimate the potential barrier height V_{barr}. If the logarithm of I_d is plotted versus $1/T$ for a fixed V_{ds} in the range of the approximately exponential increase of I_d, the plot consists of two straight lines with different slopes. The slope belonging to the higher temperature part of the plot is due to thermionic emission over the potential barrier whereas the other part is due to tunneling through it [73]. The first slope is equal to V_{barr}. This procedure is repeated for several V_{ds}. The obtained barrier heights are plotted versus $\sqrt{V_{ds}}$ which should lie on a straight line. Thus, the linear extrapolation to $V_{ds} = 0$ gives the actual value of V_{barr} [73]. If this analysis is done for different gate voltages it can be seen that the barrier height depends on V_g. This dependence is displayed in figure 6.9 (a) which shows a decrease of V_{barr} with increasing gate voltage. Such a behavior can be a result of the variation in the thickness of the gate oxide along the channel. Due to the enhanced oxidation rate of the n^{++} on the V-groove flanks and due to the special geometry the vertical distance between gate and channel underneath the oxidized n^{++} is increased to about 6 nm in comparison to 2.6 nm on the p^- in the channel. Therefore, the gate drive is significantly reduced in these areas. This is illustrated in figure 6.9 (b) where the black region represents the density of carriers in the channel region which is much smaller at source and drain. The carrier density per unit area in the channel is usually assumed to

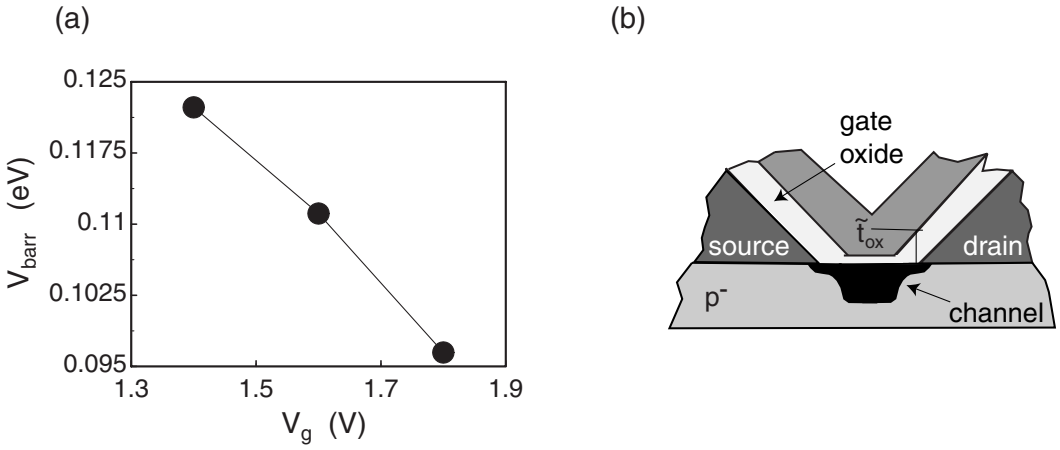

Figure 6.9: (a) Dependence of barrier height on V_g. (b) Magnification of the channel region showing the inversion layer (black) in equilibrium. Due to the enhanced oxidation rate and geometry the gate drive is reduced at source and drain.

be proportional to the oxide capacitance and gate overdrive $n_s = C_{ox}(V_g - V_{th})$ with $C_{ox} = \varepsilon_{ox}/t_{ox}$ [76] so that n_s is decreased by the factor $6/2.6 \approx 2.3$. Hence, the fundamental contact resistance (see equation (6.1)) is increased by a factor 1.5. Simulations have shown that a potential barrier of $\sim 120\,\text{meV}$ does not influence the electrical behavior of the simulated devices at room temperature because these barriers are of the order of the potential barrier represented by the channel. For room temperature the amount of carriers which overcome this barrier and enter the channel is so high that the potential profile in the channel is not much altered when compared to the situation without the barriers. At lower temperatures less carriers are able to enter the channel due to the barriers which significantly influences the potential profile in the channel and therefore the electrical behavior. Hence, the existence of potential barriers does not collide with the analysis of the previous section where we have taken the resistance R_p to be independent of V_g, since the potential barriers apparently do not have an impact at room temperature. However, the exact nature of the potential barriers is to date an open question we are still working on. In particular, four terminal measurements at lower temperatures than 80 K will give new insights.

6.3 Results

The results of the present chapter can be summarized as follows:

- The V-groove approach is capable of producing ultra-short channel MOSFETs with state-of-the-art electrical characteristics.
- MOSFETs with the same nominal channel length show equal electrical behavior. This proves the reproducibility of the V-groove approach. Furthermore, it shows that MOSFETs with defined channel lengths can be fabricated which is an important issue concerning the applicability in circuits.
- Simulations show that the device characteristics can drastically be improved if the SOI film and gate oxide thickness is reduced for the next batch of samples.
- Low temperature measurements give evidence for the presence of potential barriers at source and drain. These barriers are low enough so that they are irrelevant at room temperature but appear at temperatures below $\sim$170 K. Barrier heights of about 0.12 eV are found and are attributed to the reduced gate drive in the regions between channel and source/drain contacts.

7 What Comes Next?

The continuously decreasing size of today's MOSFET devices makes new concepts for both, theoretical simulations and experimental realizations of ultra-short channel devices necessary. In the present thesis both aspects have been addressed.

Quantum mechanical effects play an increasingly important role in today's state-of-the-art devices. Modeling of a single-gated, ballistic SOI-MOSFET has therefore been performed with a self-consistent solution of the POISSON and the quantum transport equations. Additionally, a concept for the fabrication of ultra-short channel MOSFETs has been introduced and the technology in order to realize this concept has been developed.

Fully quantum mechanical simulations become increasingly important as channel lengths reach the regime with $L < 30\,\mathrm{nm}$ because even at room temperature electrons move ballistically through the channel on these length scales. Without scattering the wave nature of the electrons becomes important and influences the transport in ultra-short channel devices. The quasi two-dimensional approach presented here enables the computation of the quantum transport equations of carriers traversing the channel. Additionally, vertical quantization due to the confinement of the carriers in ultra-thin SOI films as well as tunneling through the gate is taken into account. The simulations presented so far show that in principle MOSFET operation is possible for single-gated devices with 10 nm channel length. However, extremely thin SOI and oxide layers are needed. Whether such transistors can be experimentally realized depends crucially on the feasibility to fabricate extremely smooth ultra-thin SOI films. The theoretical analysis of such ultimately scaled devices has shown that a roughness of 10% of the SOI thickness yields threshold voltage fluctuations which disable circuit applications. The formalism used here, allows to incorporate inelastic scattering. Future work has to address the inclusion of inelastic scattering what is inevitable in order to correctly describe electron transport in devices with larger channel lengths.

The technology developed in this thesis enables the reproducible fabrication of ultra-short channel MOSFETs. It has been proven that source-drain separations of 10 nm are feasible with the V-groove approach. Hence, devices with such short channels can in principle be realized in order to explore the limits of today's CMOS technology. The main advantages of the V-groove approach are the abrupt doping profile accessible with solid phase epitaxy and the freedom to adjust the fabrication process to every desired lithographic technique. Thus, it is possible to produce devices with significantly smaller channel

lengths than the present lithographic technique is able to produce. Due to the quasi-self limiting etch behavior during fabrication a reliable and reproducible technology has been developed. Ultra-shallow source and drain contacts due to epitaxially grown highly doped silicon together with the use of thin SOI films efficiently suppresses short-channel effects. MOSFETs with a minimal channel length of 36 nm have been processed and characterized by measuring their $I-V$ characteristics at room temperature and at low temperatures. State-of-the-art electrical behavior could be observed for these samples when compared to the most recent devices fabricated by other research groups. However, the present approach still has an enormous potential for improving the characteristics and for a further miniaturization. The SOI film thickness can be reduced by a factor of three, the oxide thickness by a factor of 1.5. In addition, the doping concentration of the epitaxial n^{++} material can be increased by a factor of 2. With these alterations major improvements of the electrical characteristics are expected which is supported by simulations. Additionally, it is believed that by reducing the SOI and gate oxide thickness a significant reduction of the channel length is possible without corrupting the device's performance by short-channel effects. The main difficulty remains the production of appropriate epitaxial material. The challenges lie in the generation of the high doping level combined with a good crystalline quality. In addition, a moderate interface layer between n^{++} and p^{-} has to be generated which plays an important role when extremely thin SOI films are employed.

Future work will deal with the reduction of all geometrical dimensions of the V-groove devices and with the fabrication of appropriate gates in order to facilitate high-frequency applications. An approach how this can be accomplished has been given but has not yet been fully worked out. The technology has proven to be reliable and reproducible so that it can be used unchanged for the production of further miniaturized devices. Besides that, low temperature measurements have to be performed in order to determine the transport mechanisms in the V-groove MOSFET. In particular, the comparison with the quantum mechanical simulations developed here will provide insight into the nature of carrier transport in ultimately scaled MOSFETs.

A Computational Issues

In this appendix a detailed description of the theory used for the quantum mechanical simulations in chapter 2 is given. The first part is devoted to the quantum mechanical aspects of the computations, the second gives some details of the numerical procedure.

A.1 Nonequilibrium Green's Function Formalism

Starting point is a discretized version of the HAMILTON-operator H on a tight-binding grid with lattice constant a. For the one-dimensional problem considered here H is given in matrix notation by the expression [17]

$$\mathsf{H}_{ij} = \begin{cases} V_i + 2t & i = j \\ -t & i,j \quad \text{nearest neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

where the coupling parameter $t = \frac{\hbar}{2m_l a^2}$ and V_i is the potential energy at grid point i. The light effective electron mass is used for the calculations since the electric quantum limit is assumed [68]. In this case, transport effective mass equals m_l and the transverse mass is m_h. The discretization scheme is depicted in figure A.1 showing the potential distribution of the device, too. The shaded regions, i.e. the device and the intermediate regions, are the computational domain. The attached leads L_l are taken into account analytically as described below. The Green's function of H in matrix notation is[1]

$$(\mathsf{H} \cdot \mathsf{G})_{ij} = \frac{1}{a}\delta_{ij} \quad \Rightarrow \quad \mathsf{G}^{R,A} = \frac{1}{a}[E \times 1 - \mathsf{H} \pm i\eta]^{-1} \tag{A.2}$$

where the '+' sign stands for the retarded and the '−' sign for the advanced Green's function. A '$\cdot$' denotes a matrix multiplication which is equivalent to an integration over the internal spatial variable of the continuous representation. Note, that $\mathsf{G}^{R,A}$ is normalized to δ_{ij}/a which as $a \to 0$ results in the usual $\delta(x - x')$ of the continuous formulation. Hence, $\mathsf{G}^{R,A}$ can be calculated simply by inverting the matrix $[E \times 1 - \mathsf{H} \pm i\eta]$.

For the calculation of an open system the coupling of the conductor C (i.e. the device and the intermediate regions in figure A.1) to semi-infinite leads L_l can be accounted for *exactly* by a retarded self-energy function Σ^R of the following form [17]:

$$\Sigma_k^R = t^2 g_L^R(i,j) \qquad \text{with} \qquad \Sigma^A = \left(\Sigma^R\right)^\dagger \tag{A.3}$$

[1] Matrices are written in Sans Serif

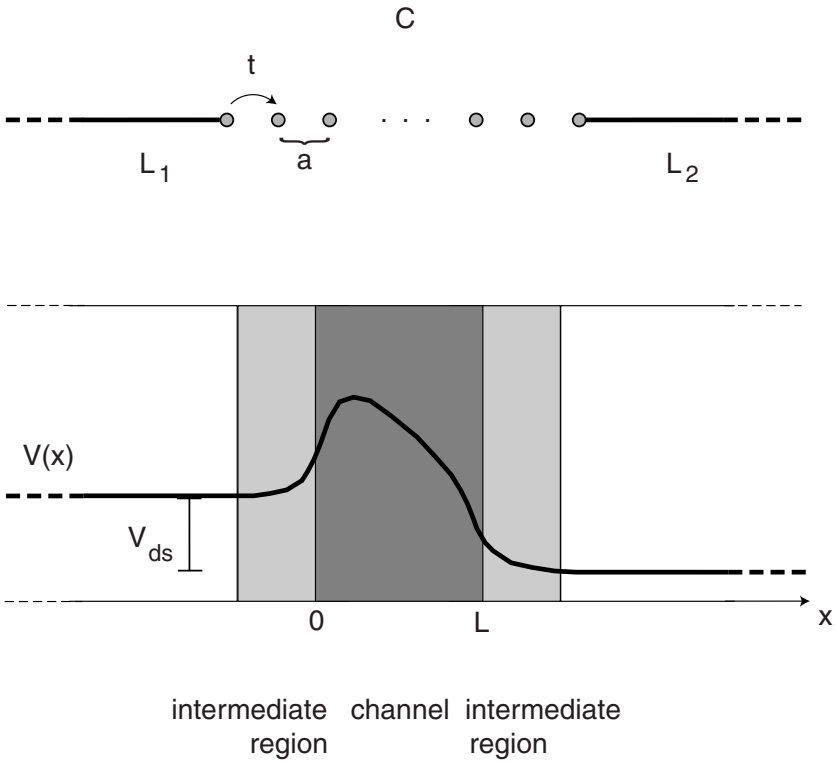

Figure A.1: Sketch of the computational domain C. The domain is partitioned onto a grid with lattice spacing a, nearest neighbor hopping parameter t. Two semiinfinite leads $L_{1,2}$ are attached to the device.

where (i, j) are points in the lead k adjacent to the conductor C. Since the Green's function of the semi-infinite lead g_L^R can be computed analytically one is able to replace the infinite problem by a finite one taking the effects of the leads into account through the self-energy function $\Sigma^{R,A}$. For the $1D$ case of a conductor connected to two leads the retarded self-energy has a rather simple form

$$\Sigma^R = -t\exp(ik_1a) \times \mathsf{M}_1 - t\exp(ik_Na) \times \mathsf{M}_N \tag{A.4}$$

where the only non-zero element of the matrices $\mathsf{M}_{1,N}$ are the $(1,1)$ and (N,N) element, respectively. $k_1 = \sqrt{2mE/\hbar^2}$ and $k_N = \sqrt{2m(E-V_{ds})/\hbar^2}$ with V_{ds} being the drain-source voltage.

The local density of states is proportional to the imaginary part of the retarded Green's function $\mathsf{N}_C(i,E) = -\frac{1}{\pi}\frac{1}{a}\mathrm{Im}\mathsf{G}_C^R(i,i) = \frac{1}{2\pi}\mathsf{A}_C(i,i)$. The function A_C is called the spectral function which can be written in the following form

$$\mathsf{A}_C = \mathsf{G}_C^R \cdot \Gamma \cdot \mathsf{G}_C^A = \mathsf{G}_C^R \cdot \left[i(\Sigma^R - \Sigma^A)\right] \cdot \mathsf{G}_C^A \tag{A.5}$$

In equilibrium the calculation of G_C^R would be sufficient to deduce any observable from this function. Since we deal with nonequilibrium transport we usually

need the functions $\mathsf{G}_C^{n,p}$ and $\Sigma^{n,p}$ (throughout the text the notation of DATTA [17] is used), too, which are the in/out-scattering and electron/hole correlation functions, respectively. In the following the subscript C is suppressed for clarity. These functions are given in matrix notation by the following expressions

$$\mathsf{G}^n = \mathsf{G}^R \cdot \Sigma^{in} \cdot \mathsf{G}^A \qquad \text{and} \qquad \mathsf{G}^p = \mathsf{G}^R \cdot \Sigma^{out} \cdot \mathsf{G}^A \tag{A.6}$$

However, as long as fully coherent transport is considered a simplification can be done. As is well known from the studies of mesoscopic transport phenomena [17] carriers flowing into the conductor C from the left and right lead, respectively, can be considered as being distributed according to the distribution function of the respective lead. If the contacts are in thermal equilibrium the functions $\Sigma_{s,d}^{in} = f_{s,d}^0(E - E_f^{s,d})\Gamma_{s,d}$ and $\Sigma_{s,d}^{out} = (1 - f_{s,d}^0(E - E_f^{s,d}))\Gamma_{s,d}$ with $\Gamma_{s,d} = i(\Sigma_{s,d}^R - \Sigma_{s,d}^A) = -2\mathrm{Im}(\Sigma_{s,d}^R)$ can be split up into two parts referring to carriers originating from source and drain. Hence, G^n can be written as $\mathsf{G}^n = \mathsf{G}^R \cdot \Sigma_s^{in} \cdot \mathsf{G}^A + \mathsf{G}^R \cdot \Sigma_d^{in} \cdot \mathsf{G}^A$ and the electron density is [18, 17]

$$n_i = \frac{1}{2\pi} \int dE\, f_0^s(E - E_f^s)\mathsf{A}_s(i,i) + f_0^d(E - E_f^d)\mathsf{A}_d(i,i) \tag{A.7}$$

with $\mathsf{A}_{s,d}(i,i)$ being the diagonal elements of the local density of states for carriers originating from left (source) and right (drain). The FERMI energies in source and drain are fixed by the terminal voltages and $E_f^s - E_f^d = V_{ds}$. Equation (A.7) is intuitive since the density is simply a product of an occupation factor times the local density of states.

The terminal-current of contact l (i.e. source or drain) per unit energy is given by [17]

$$i_l(E) = \frac{2e}{h}\mathrm{Tr}\left[\Sigma_l^{in} \cdot \mathsf{G}^p - \Sigma_l^{out} \cdot \mathsf{G}^n\right] \tag{A.8}$$

where the factor of 2 is due to the spin degeneracy and $\frac{e}{\hbar}$ is the current carried per unit energy by an occupied mode of energy E.

Up to now all calculations were merely one-dimensional. In order to solve the POISSON equation and the equation for the current we need the three-dimensional expressions of the Green's functions computed so far. As stated in section 4.1.2 for the calculation of the carrier density translational invariance in y,z-direction is valid only in the contacts. In this case the averaging is straightforward. The energy variable in the 1D expressions is simply the total energy minus the transversal part $E = E_{tot} - E_{k_\perp}$. Therefore, in order to get a 3D averaging one simply has to integrate over the transverse momentum variable, i.e. one simply has to replace the FERMI function by the so-called supply-function [18].

$$\sum_{k_\perp} f_0(E + E_{k_\perp} - E_f^{s,d}) = \frac{6m_{dos}}{\pi\hbar^2} kT \ln\left(1 + \exp\left(\frac{E_f^{s,d} - E}{kT}\right)\right) = F_0(E_f^{s,d} - E_{tot}) \tag{A.9}$$

where m_{dos} is the $2D$ density of states effective mass and the factor of 6 is due to the six equivalent valeys [5]. In order to get a three-dimensional expression

for the electron density in the channel and the current the $1D$ expressions obtained so far have to be averaged over the transversal coordinates. We begin by rewriting the expressions for the terminal current in the source contact in the following way with source as contact s and drain as contact d

$$i(E) = \frac{2e}{h}\mathrm{Tr}\left[f_s^0\Gamma_s \cdot \mathsf{A}_d - f_d^0\Gamma_s \cdot \mathsf{A}_d\right] = \frac{2e}{h}(f_s^0 - f_d^0)\,\mathrm{Tr}[\Gamma_s \cdot \mathsf{A}_d] \tag{A.10}$$

Three-dimensional expressions are obtained if $\mathsf{A}_{s,d}$ and $\Gamma_s \cdot \mathsf{A}_d$ in equation (A.7) and (A.10) are replaced by their $3D$ counterparts. The $3D$ expressions of the spectral functions $\mathsf{A}_{s,d}$ are also needed for the computation of the electron density according to equation (A.7) which will be demonstrated now.

The assumption is made that in the channel region of the device translation invariance is present along the y-direction (i.e. along the width of the transistor). In the z-direction only the first two-dimensional subband is taken into account which is a reasonable approximation if the silicon thickness t_{si} is small enough [50]. Higher subbands are accounted for by a numerical factor in the fashion of reference [49]. To estimate the energetic position and the separation of the first and the succeeding subband a rectangular potential well is considered with the eigenfunctions being the well-known sin or cos expressions. The energy of the first subband E_1 is given in equation (4.11). The exact form of the corresponding wave-function is not important as long as it is normalized because the z-coordinate will be integrated over as is shown below.

The function $\mathsf{A}_{s,d}$ can be expanded as $\mathsf{A}_{s,d} = \mathsf{G}^R \cdot \Gamma_{s,d} \cdot \mathsf{G}^A = \mathsf{G}^R \cdot [i(\Sigma^R_{s,d} - \Sigma^A_{s,d})] \cdot \mathsf{G}^A$ as already stated above in equation (A.5). The averaging procedure is therefore demonstrated only for the product $\mathsf{G}^R \cdot \Sigma^R_d \cdot \mathsf{G}^A$. The continuous representation is used in the following and the Greens functions can be expanded into a complete set of orthonormal eigenfunctions.

$$G^R(\vec{r},\vec{r}') = \sum_{k_y,k_z,l} \frac{\phi_{k_y}(y)\psi_{k_z}(z)\phi^*_{k_y}(y')\psi^*_{k_z}(z')\chi_l(x)\eta^*_l(x')}{E - \varepsilon_{k_y} - \varepsilon_{k_z} - \varepsilon_l}$$

$$= \sum_{k_y,k_z} \phi_{k_y}(y)\psi_{k_z}(z)\phi^*_{k_y}(y')\psi^*_{k_z}(z')\chi_l(x)\eta^*_l(x') \times G^R_{1D}(x,x',E - \varepsilon_{k_y} - \varepsilon_{k_z}) \tag{A.11}$$

The advanced Greens function is expanded in a similar fashion. The retarded self-energy function of terminal $l = s, d$ is given as

$$\Sigma^R_l = -t \sum_{k_y,k_z} \phi_{k_y}(y)\psi_{k_z}(z)\phi^*_{k_y}(y')\psi^*_{k_z}(z')\chi_l(x)\eta^*_l(x')e^{ik^l_x a}\,\delta(x - x') \tag{A.12}$$

where $k^l_x = 1/\hbar \times \sqrt{(2m(E - \varepsilon_{k_y} - \varepsilon_{k_z} + V_l))}$, a is the lattice constant of the tight-binding grid and V_l the terminal voltage. Since only the first subband in vertical direction is considered the summation over k_z reduces to a factor being the absolute square of the first vertical subband wave-function (note, that merely the diagonal elements are needed for the density and the current

and therefore $\vec{r} = \vec{r}'$) and $\varepsilon_{k_z} \to E_1$. It is assumed that the potential profile varies in z-direction much more than in x-direction, so that the Hamiltonian can be considered separable. The actual averaging is an integration over the y-coordinate (cf. [4, 41])

$$\frac{1}{W_y}\int dy\, \mathsf{G}^R \cdot \Sigma_l^R \cdot \mathsf{G}^A = \langle \mathsf{G}^R \cdot \Sigma_l^R \cdot \mathsf{G}^A \rangle$$

$$= \frac{2}{t_{si}}\cos^2(\frac{\pi z}{t_{si}})\frac{-t}{W_y}\int dy\, dx_1 \times$$

$$\times \sum_{k_y^1 k_y^2 k_y^3} G_{1D}^R(x, x_1, E - \varepsilon_{k_y^1} - E_1) G_{1D}^A(x, x_1, E - \varepsilon_{k_y^3} - E_1) e^{ik_x^l a} \phi_{k_y^1}(y) \phi_{k_y^3}^*(y) \delta_{k_y^1 k_y^2} \delta_{k_y^2 k_y^3}$$

$$= \frac{2}{t_{si}}\cos^2(\frac{\pi z}{t_{si}})\frac{1}{W_y}\sum_{k_y} \underbrace{\int dx_1 G_{1D}^R(x, x_1, E - \varepsilon_{k_y} - E_1)\left(-te^{ik_x^l a}\right) G_{1D}^A(x, x_1, E - \varepsilon_{k_y} - E_1)}_{= (\mathsf{G}^R \cdot \Sigma_p^R \cdot \mathsf{G}^A)_{1D}(E - \varepsilon_{k_y} - E_1)} \times$$

$$\times \underbrace{\int dy \phi_{k_y}(y)\phi_{k_y}^*(y)}_{=1}$$

$$\overset{E-\varepsilon_{k_y}-E_1 \to \tilde{E}}{=} \frac{2}{t_{si}}\cos^2(\frac{\pi z}{t_{si}})\frac{1}{W_y}\sum_{k_y}\left(\mathsf{G}^R \cdot \Sigma^R \cdot \mathsf{G}^A\right)_{1D}(\tilde{E})$$

$$= \frac{2}{t_{si}}\cos^2(\frac{\pi z}{t_{si}})\int_{-\infty}^{E-E_1} d\tilde{E}\; N_{1D}^0(E - \tilde{E} - E_1) \times (\mathsf{G}^R \cdot \Sigma^R \cdot \mathsf{G}^A)_{1D}(\tilde{E}) \tag{A.13}$$

N_{1D}^0 is the homogeneous one-dimensional density of states. The expression for the current density equation (A.10) contains the matrix product $\Gamma_s \cdot \mathsf{G}^R \cdot \Sigma_d^R \cdot \mathsf{G}^A$ which is averaged in the same fashion as equation (A.13). Replacing the calculated $1D$ expressions by the appropriate $3D$ entities the final electron density is taken as an average over the z-direction, i.e. the integration over z and division by t_{si} yields the electron density necessary for the self-consistent computation of the potential profile.

Although the entire computation includes the space charge layers in the contacts, i.e. the intermediate regions, to calculate the total current only the transmission of carriers between channel beginning and end must be computed. Inserting the expressions of equation (A.13) and noting that $\int dz \frac{2}{t_{si}}\cos^2(\frac{\pi z}{t_{si}}) = 1$ one ends up with the following formula for the current density per unit energy:

$$i_1(E)^{3D} = W\frac{2e}{h}t(f_1 - f_2)\left[\int_{-\infty}^{E-E_1} d\tilde{E} N_{1D}^0(E - \tilde{E} - E_1)\tilde{\Gamma}_s(i_c - 1, \tilde{E}) \cdot \left(\mathsf{G}^R \cdot \Gamma_2 \cdot \mathsf{G}^A\right)(i_c, \tilde{E})\right] \tag{A.14}$$

where i_c denotes the first tight-binding grid point in the channel and $\tilde{\Gamma}_s = 2t^2 \times \mathrm{Im}\mathsf{G}^R(i_c - 1, i_c - 1)$. After self-consistency has been achieved the current is calculated according to equation (A.14).

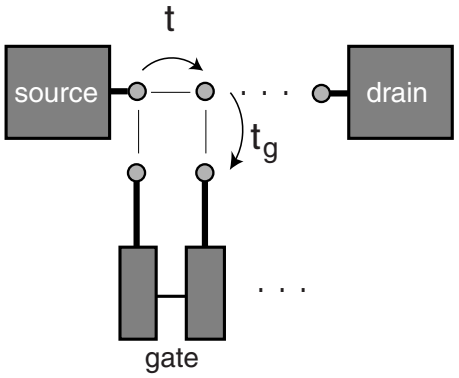

Figure A.2:

The gate leakage current can easily be implemented in the present formalism. The gate consists of additional contacts all having the same FERMI level $E_f - V_g$ attached to each site of the tight-binding grid in the channel region. This is depicted in figure A.2. Attaching a 'gate contact' at each site is necessary so that the tunneling current properly scales with the length of the channel. E_f in these gate contacts is taken to be the same as in the actual source/drain contacts. The difference between the ohmic contacts, i.e. source and drain, and the gate contact(s) is that the hopping parameter t is much smaller depending on the oxide thickness and its potential barrier height (the height is taken as constant V_0 =3.1 eV). The actual hopping parameter for the gate-contact(s) is given according to

$$t_g = t \times \exp\left(-\frac{2}{\hbar}\sqrt{2m_l(V_0 - E)}\, t_{ox}\right) \tag{A.15}$$

where the exponential factor is the WKB solution of the transmission amplitude through a potential barrier (see figure A.2) of height V_0. The gate leakage current component of each gate contact is computed in the same fashion as the source-drain current and is added to the current from source to drain. The influence of gate leakage becomes important only for very thin gate oxides.

A.2 Numerics

Because the matrix has to be inverted for every energy-channel E the computational burden can be immense and one has to make use of sparse matrix routines [58]. To be explicit, the following matrix must be inverted for every

energy channel E:

$$\mathsf{G}^R = \left[\begin{pmatrix} E-(V_1+2t)+t\cos(k_1a) & t & 0 & \cdots & \cdots & 0 \\ t & E_2 & t & 0 & \cdots & 0 \\ 0 & t & E_3 & t & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & t & E_{N-1} & t \\ 0 & \cdots & 0 & 0 & t & E-(V_N+2t)+t\cos(k_Na) \end{pmatrix} \right. \\ \left. +i \begin{pmatrix} t\sin(k_1a)+\eta & 0 & \cdots & \cdots & 0 \\ 0 & \eta & 0 & \cdots & 0 \\ 0 & 0 & \eta & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \eta & 0 \\ 0 & \cdots & \cdots & \cdots & t\sin(k_Na)+\eta \end{pmatrix} \right]^{-1} \tag{A.16}$$

where $E_i = E-(V_i+2t)$ and a small η is inserted to get a stable numerical inversion. The real and imaginary parts of the Green's function G^R are calculated separately according to

$$\mathrm{Im}\mathsf{G}^R = -[\mathsf{A}\cdot\mathsf{B}^{-1}\cdot\mathsf{A}+\mathsf{B}]^{-1} \tag{A.17}$$

$$\mathrm{Re}\mathsf{G}^R = -\mathsf{B}^{-1}\cdot\mathsf{A}\cdot\mathrm{Im}\mathsf{G}^R \tag{A.18}$$

where $\mathsf{A} = \mathrm{Re}[E\times 1-\mathsf{H}-\Sigma^R]$ and $\mathsf{B} = \mathrm{Im}[E\times 1-\mathsf{H}-\Sigma^R]$. To obtain $\mathrm{Im}\mathsf{G}^R$ a pentagonal matrix has to be inverted which is most efficiently done by partitioning the matrix in the following way [58]

$$\mathsf{A} = \begin{pmatrix} \mathsf{P} & \mathsf{Q} \\ \mathsf{R} & \mathsf{S} \end{pmatrix} \longrightarrow \mathsf{A}^{-1} = \begin{pmatrix} \tilde{\mathsf{P}} & \tilde{\mathsf{Q}} \\ \tilde{\mathsf{R}} & \tilde{\mathsf{S}} \end{pmatrix} \tag{A.19}$$

where the submatrices $\tilde{\mathsf{P}}, \tilde{\mathsf{Q}}, \tilde{\mathsf{R}}, \tilde{\mathsf{S}}$ can be found by the formulas

$$\begin{aligned} \tilde{\mathsf{P}} &= (\mathsf{P}-\mathsf{Q}\cdot\mathsf{S}^{-1}\cdot\mathsf{R})^{-1} \\ \tilde{\mathsf{Q}} &= -(\mathsf{P}-\mathsf{Q}\cdot\mathsf{S}^{-1}\cdot\mathsf{R})^{-1}\cdot(\mathsf{Q}\cdot\mathsf{S}^{-1}) \\ \tilde{\mathsf{R}} &= -(\mathsf{S}^{-1}\cdot\mathsf{R})\cdot(\mathsf{P}-\mathsf{Q}\cdot\mathsf{S}^{-1}\cdot\mathsf{R})^{-1} \\ \tilde{\mathsf{S}} &= \mathsf{S}^{-1}+(\mathsf{S}^{-1}\cdot\mathsf{R})\cdot(\mathsf{P}-\mathsf{Q}\cdot\mathsf{S}^{-1}\cdot\mathsf{R})^{-1}\cdot(\mathsf{Q}\cdot\mathsf{S}^{-1}) \end{aligned} \tag{A.20}$$

The equations can be solved iteratively by first successively calculating $\tilde{\mathsf{P}}_i = \mathsf{A}_{i+1}$ for $i = N..2$ where A_{i+1} is in turn decomposed as in equation (A.19). Because S is always taken as a 1×1 submatrix consisting of the $\mathsf{A}_{i,i}$ element and due to the fact that A is a sparse matrix, all matrix multiplications can be determined directly. The whole inversion is thus a process of the order N^2+N with N being the dimension of the matrix A. This algorithm is similar to the recursive Green's function method [25] which has proven to be the most efficient for matrix inversion for this particular purpose.

The NEWTON-RAPHSON-method to solve the POISSON equation together with the equations for the Green's function is adopted from reference [42]. This method leads to self-consistency within a few iterations. In each iteration a correction $\delta\Phi^{m+1}$ to the potential profile Φ^m of the mth iteration is calculated. The POISSON equation is written as

$$\frac{d^2\Phi}{dx^2} - \frac{\rho(x)}{\varepsilon_s i} = \delta\rho(x) \tag{A.21}$$

The correction $\delta\Phi$ is the solution of the following equation given in matrix notation

$$\sum_j \frac{\partial\delta\rho_i^m}{\partial\Phi_j^m}\delta\Phi_j^{m+1} = -\delta\rho_i^m \tag{A.22}$$

where i, j denote points on the tight-binding grid. The JACOBIAN $\sum_j \frac{\partial\delta\rho_i^m}{\partial\Phi_j^m}$ requires the calculation of $\frac{\partial n_i}{\partial\Phi_j}$ where n_i is the quantum charge density. To compute this the semi-classical charge density is used instead of the quantum charge [42]. The particular form of a screened POISSON equation used here is accounted for in the channel region such that the JACOBIAN looks like

$$\frac{\partial\delta\rho_i}{\partial\Phi_j} = \delta_{i,j}\frac{2}{a^2} - \delta_{i,i\pm1}\frac{1}{a^2} + 2\delta_{i,j}\left(\frac{2m_l kTe}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{1}{kT}\frac{F_{1/2}(E_f-\Phi_i+0.001) - F_{1/2}(E_f-\Phi_i)}{0.001} \tag{A.23}$$

in the contacts (with a being the lattice spacing) and

$$\frac{\partial\delta\rho_i}{\partial\Phi_j} = \delta_{i,j}\left(\frac{2}{a^2} + \frac{1}{\lambda^2} + \frac{V_{bi}-V_g}{\lambda^2}\right) - \delta_{i,i\pm1}\frac{1}{a^2} \tag{A.24}$$

in the channel. Notice that the last fraction in equation (A.23) is just the derivative of the FERMI-integral. Equation (A.21) yields a tridiagonal matrix which is solved with a standard recursion method [82]. NEUMANN boundary conditions are used. Thus, the applied voltage V_{ds} fixes the energetic difference of the FERMI-levels in source and drain; the potential takes on whatever value is needed to ensure overall charge neutrality and a flat potential profile at the contact beginning [18]. The potential profile is then given as $\Phi^{m+1} = \Phi^m + \delta\Phi^{m+1}$.

Self-consistency is achieved when the maximum potential variation between the ith and the $(i+1)$st iteration is smaller than a specified value.

B What Etches What?

	Pan-Etch	Cr-Etch	BOE	KOH (30° C)	H_2O_2	RCA1/2	H_2O_2 (60° C)	Cr-Etch (50° C)	H_3PO_4 (80° C)
Al	++	rough surface after 30 min, OK for 6-7 min	++	++	rough surface after 30 min				
Cr	X(1/2 h)	++	X	X(1/2 h)	X	X	X(1/2 h)	++	
Nb	X(1/2 h)	X(1 h)	rough surface after 3 min	X(1 h)	X(1/2 h)		X(1/2 h)	X(1/2 h)	
W	X(1 h)	rough surface after 30 min	X(5 min)	X(1/2 h)	++		++	~6 nm/min	X(1/2 h)
Si	X	X	X	++[1]	X	X	X	X(1 h)	
Resist 7500/107		X		++		+		after 30 min(?)	
Resist AZ5214			X(2 min) with HMDS	++	after 6 min	+			
SiO_2	X	X	++	very small etch rate[2]	X	X	X	X(1 h)	X

[1,2] see ref. [65, 22]

X → not etched in specified time

++ → appropriate etch

C Recipes

KOH solution(40%, 30° C)

- mix 112.5 g H_2O with 100 g KOH pellets
- wait until the temperature is under 40° C
- add 12.5 g isopropyl alcohol (IPA)

SiO_2 Mask Definition

- cleaning of the sample in acetone, IPA and deionized (DI) water
- resist: AR 7500/107, 6000 rpm for 30 s (without pre-acceleration)
- prebake 5 min, 90°
- optical lithography to pre-pattern the resist: 7 s, 10 mW/cm^2
- develop not longer than 6 s (!)
- e-beam lithography: rather high dose of approximately 400 μC/cm^2; for much higher doses the resist cannot be solved anymore. Additionally, the adherence to the substrate is deteriorated.
- watering for 3 min
- 3 s flood exposure, 10 mW/cm^2
- develop 30 s in AR 300-47
- post-bake 1 h, 120° C
- approximately 12 s BOE(AF-91) for 7 nm oxide
- removal of resist in 1-methyl-dipyrolidon (NMP) for at least 3 h at 80° C, then immerse in acetone, IPA and DI water

KOH Etching

- immediately before KOH etching: Dip in BOE to remove native oxide
- KOH as stated above always with stirring to prevent concentration gradients in the etch solution which lead to an altered etch behavior
  - etch rate({100} planes): ∼30 nm/min for undoped silicon, ∼10 nm/min for the highly Sb doped silicon
  - etch rate({111} planes): ∼1 nm/min for undoped silicon, ∼2.5 nm/min for highly Sb doped silicon
- water for at least 5 min

Si-Cr Mask Definition

- after mask generation dip in BOE to remove native oxide
- mount in the UHV chamber immediately after BOE dip
- evaporate ∼20 nm Cr
- bake 1h, 120° C
- remove Cr in Cr-etch and remove mask
- KOH as stated above

SC1/SC2

- heat 150 ml H_2O to 60° C
- add 30 ml NH_4OH/HCl and 30 ml H_2O_2
- heat up to 65° C

D Remarks

This appendix gives some details of the fabrication process which were left out in the main text for reasons of clarity.

D.1 The wrong Mask Geometry

If one chooses the wrong mask geometry, i.e. one which does not encounter the underetching of convex corners, the mask can be totally underetched. If this happens the structure is not bounded by {111} planes any more and the silicon under the mask is etched away completely. This is shown in figure D.1. As

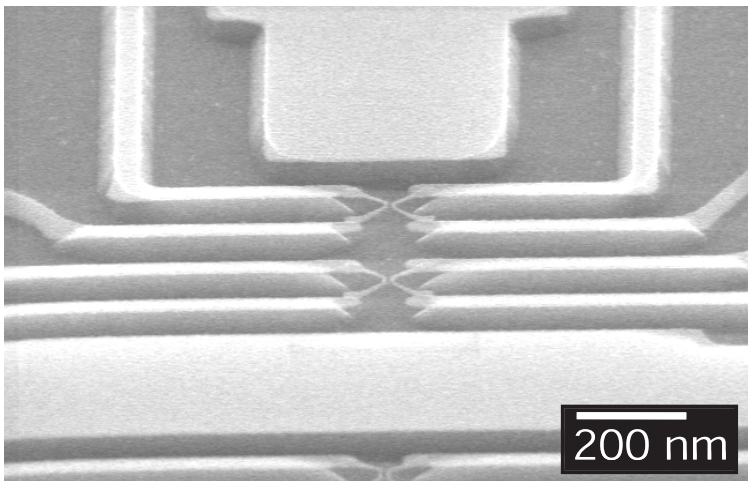

Figure D.1: SEM image of transistor structures with an inadequate mask geometry. The silicon underneath the SiO_2 mask is completely etch away.

can be seen the silicon underneath the SiO_2 mask is etched away and the mask stands freely. Notice the tremendous stability of the SiO_2 mask which is only 5 nm in thickness but about one micron long.

D.2 Wet Chemical Etching of Tungsten

The effect of wet chemically etching tungsten was already mentioned in section 5.2.4. At convex bends the tungsten is - although covered by a resist mask

- cut through. Although the quasi self-aligned process of the gate is based on this phenomenon it can be disturbing since the gate can thus be separated into electrically isolated parts. Figure D.2 (a) shows an electron micrograph of the

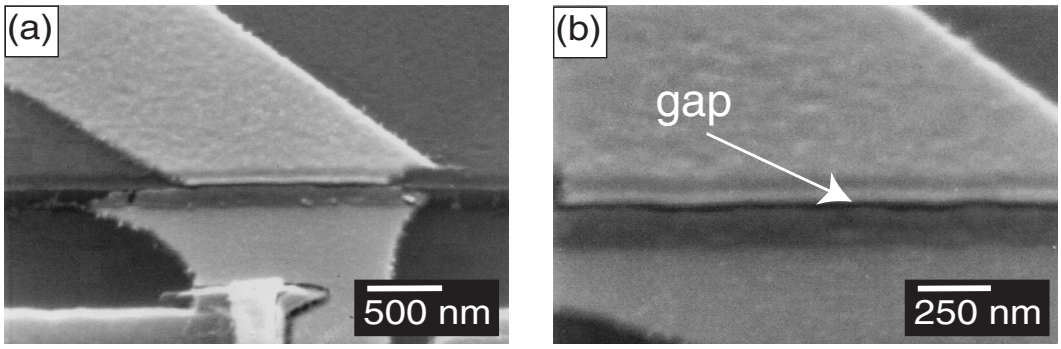

Figure D.2: Wet chemical patterning of the tungsten gate. (a) shows the behavior of the tungsten when running over convex bends. At the bottom of the image the transistor structure can be seen. (b) shows a magnification. The cut off of the tungsten can clearly be seen.

tungsten gate after wet chemical patterning which runs over a convex corner. (b) is a magnification and reveals the effect of the H_2O_2-treatment which manifests itself in the gap between the two parts of the tungsten gate. In order to shorten the two isolated parts an additional aluminum contact layer was evaporated and patterned using a standard lift-off process.

E Used Chemicals

BOE	AF-91	Riedel de Haën	No. 19408
HCl	32%	Merck	No. 1.00319.1000
Cr-Etch 3144		Merck	No. 17553
H_2O_2	30%	Merck	No. 1.07209.1000
KOH	85%	Merck	No. 1.05021.1000
HF	50%	Merck	No. 1.00373.1000
NH_3	25%	Merck	No. 1.05432.1011
E-Beam resist	AR 7500/107	Allresist GmbH	
Developer	AR 300-47	Allresist GmbH	

Bibliography

[1] A. Abou-Elnour and K. Schuenemann, *A Comparison Between Different Numerical Methods Used to Solve Poisson's and Schroedinger's Equations in Semiconductor Heterostructures*, J. Appl. Phys., **74**, 3273, 1993.

[2] V. Aggarwal, M. Khanna, R. Sood, S. Haldar and R. Gupta, *Analytical Two-Dimensional Modeling for Potential Distribution and Threshold Voltage of the Short-Channel Fully Depleted SOI (Silicon-on-Insulator) MOSFET*, Solid-State Electron., **37**, 1537 – 1542, 1994.

[3] E. Ahlers, R. Ostrom, P. Datta and F. Allen, *Crystal Quality of Solid Phase Epitaxy Grown Silicon*, J. Vac. Sci Technol. B, **6**, 723– 726, 1988.

[4] M. Alam, R. Morrisey and A. Khondker, *Self-Consistent Analysis in the Presence of Phase-Randomizing Processes for Double-Barrier Structures*, J. Appl. Phys., **71**, 3077 – 3090, 1992.

[5] T. Ando, A. Fowler and F. Stern, *Electronic Properties of Two-Dimensional Systems*, Rev. Mod. Phys., **54**, 437 – 594, 1982.

[6] J. Appenzeller, J. del Alamo, R. Martel, K. Chan and P. Solomon, *Ultrathin 600 degrees C Wet Thermal Silicon Dioxide*, Electrochem. & Solid-State Lett., **3**, 84 – 86, 2000.

[7] J. Appenzeller, R. Martel, P. Solomon, K. Chan, P. Avouris, J. Knoch, J. Benedict, M. Tanner, S. Thomas, K. Wang and J. del Alamo, *Scheme for the Fabrication of Ultrashort Channel Metal-Oxide-Semiconductor Field-Effect Transistors*, Appl. Phys. Lett., **77**, 298 – 300, 2000.

[8] F. Assad, Z. Ren, S. Datta and M. Lundstrom, *Performance Limits of Silicon MOSFETs*, in *International Electron Devices Meeting 1999. Technical Digest*, 547 – 550, 1999.

[9] G. Baccarani and G. Sai-Halasz, *Spreading Resistance in Submicron MOSFETs*, IEEE Electron Dev. Lett., **4**, 27 – 29, 1983.

[10] Y. Backlund and L. Rosengren, *New Shapes in (100) Si Using KOH and EDP Etches*, J. Micromech. Microeng., **2**, 75 – 79, 1992.

[11] K. Bean, *Anisotropic Etching of Silicon*, IEEE Trans. Electron Dev., **ED-25**, 1185 – 1193, 1978.

[12] C. Bowen, G. Klimeck, M. Goodwin and D. Chapman, *Dopant Fluctuations and Quantum Effects in Sub-0.1μm CMOS*, 1997, www.cfdrc.com/nemo/pubs/isdrs_html/isdrs_html.html.

[13] R. Chau, J. Kavalieros, B. Roberds, R. Schenker, D. Lionberger, D. Barlage, B. Doyle, R. Arghavani, A. Murthy and G. Dewey, *30nm Physical Gate Length CMOS Transistor with 1.0 ps n-MOS and 1.7 ps p-MOS Gate Delays*, in *International Electron Devices Meeting 2000. Technical Digest*, 45 – 48, 2000.

[14] G. Chinalore, S. Hareland, S. Jallepalli, A. Tasch, C. Maziar, V. Chia and S. Smith, *Experimental Determination of Threshold Voltage Shifts Due to Quantum Mechanical Effects in MOS Electron and Hole Inversion Layers*, IEEE Electron Dev. Lett., **18**, 206 – 208, 1997.

[15] J.-P. Colinge, *Silicon-On-Insulator Technology: Materials to VLSI*, Kluwer Academic Publishers, 1991.

[16] S. Datta, *A Simple Kinetic Equation for Steady-State Quantum Transport*, J. Phys.: Condens. Matter, **2**, 8023 – 8052, 1990.

[17] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge University Press, 1995.

[18] S. Datta, *Nanoscale Device Modeling: The Green's Function Method*, Superlatt. Microstructures, **28**, 253 – 278, 2000.

[19] S. Datta, F. Assad and M. Lunstrom, *The Silicon MOSFET from a Transmission Viewpoint*, Superlatt. Microstructures, **23**, 771 – 780, 1998.

[20] B. Davari, R. Dennard and G. Shahidi, *CMOS Scaling, the Next Ten Years*, Proc. IEEE, **83**, 595, 1995.

[21] R. Dennard, F. Gaensslen, H.-N. Yu, V. Rideout, E. Bassous and A. LeBlanc, *Design of Ion-Implanted MOSFETs with Very Small Physical Dimensions*, IEEE J. Solid-State Circuits, **SC-9**, 256 – 268, 1974.

[22] M. Elwenspoek and H. Jansen, *Silicon Micromachining*, Cambridge University Press, 1998.

[23] T. Ernst, D. Munteanu, S. Cristoloveanu, T. Ouisse, S. Horiguchi, Y. Ono, Y. Takahashi and K. Murase, *Investigation of SOI MOSFETs with Ultimate Thickness*, Microelectronic Engineering, **48**, 339 – 342, 1999.

[24] D. Ferry, *Semiconductor Transport*, Taylor & Francis, 2000.

[25] D. Ferry and S. Goodnick, *Transport in Nanostructures*, Cambridge University Press, 1997.

[26] A. Franciosi, D. Peterman and J. Weaver, *Silicon-Refractory Metal Interfaces: Evidence of Room-Temperature Intermixing for Si-Cr*, J.Vac. Sci. Technol., **19**, 657 – 660, 1981.

[27] A. Franciosi, D. Peterman, J. Weaver and V. Moruzzi, *Structural Morphology and Electronic Properties of the Si-Cr Interface*, Phys. Rev. B, **25**, 4981 – 4993, 1982.

[28] D. Frank, S. Laux and M. Fischetti, *Monte Carlo Simulation of a 30nm Dual-Gate MOSFET: How Short can Si go?*, in *International Electron Device Meeting 1992. Technical Digest*, 553 – 556, 1992.

[29] W. Frensley, *Boundary Conditions for Open Quantum Systems Driven Far From Equilibrium*, Rev. Mod. Phys., **62**, 745 – 791, 1990.

[30] A. Ghetti, J. Bude, P. Silverman, A. Hamad and H. Vaidya, *Modeling and Simulation of Tunneling Current in MOS Devices Inlcuding Quantum Mechanical Effects*, IEICE Trans. Electron., **E83-C**, 1175 – 1182, 2000.

[31] P. Grunthaner, F. Grunthaner, R. Fathauer, T. Lin, M. Hecht, L. Bell, W. Kaiser, F. Schowengerdt and J. Mazur, *Hydrogen-Terminated Silicon Subtrates for Low-Temperature Molecular Beam Epitaxy*, Thin Solid Films, **183**, 197 – 212, 1989.

[32] P. Grunthaner, F. Grunthaner, R. Fathauer, T. Lin, F. Schowengerdt, B. Pate and J. Mazur, *Low Temperature Substrate Cleaning Technology for Si MBE*, in *Proceedings of the Second International Symposium on Silicon Molecular Beam Epitaxy*, 375 – 391, The Electrochemical Society, 1988.

[33] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor and C. Hu, *FinFET-A Self-Aligned Double-Gate MOSFET Scalable to 20 nm*, IEEE Trans. Electron Dev., **47**, 2320 – 2325, 2000.

[34] X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor and C. Hu, *Sub 50-nm FinFET: PMOS*, in *International Electron Devices Meeting 1999. Technical Digest*, 67 – 70, IEEE, 1999.

[35] R. Ikeno, H. Ito and K. Asada, *Device Parameter Estimation of SOI MOSFET Using One-Dimensional Numerical Simulation Considering Quantum Mechanical Effects*, IEICE Trans. Electron., **E80-C**, 806 – 811, 1997.

[36] H. Kawaura, T. Sakamoto and T. Baba, *Observation of Source-to-Drain Direct Tunneling Current in 8nm Gate Electrically Variable Shallow Junction Metal-Oxide-Semiconductor Field-Effect Transistors*, Appl. Phys. Lett., **76**, 3810 – 3812, 2000.

[37] H. Kawaura, T. Sakamoto, T. Baba, Y. Ochiai, J. Fujita and J. Sone, *Transistor Characteristics of 14-nm-Gate-Length EJ-MOSFET's*, IEEE Trans. Electron Dev., **47**, 856 – 860, 2000.

[38] J. Kedzierski, P. Xuan, E. Anderson, J. Bokor, T.-J. King and C. Hu, *Complementary Silicide Source/Drain Thin-Body MOSFETs for the 20nm Gate*

Length Regime, in *International Electron Devices Meeting 2000. Technical Digest*, 57 – 60, 2000.

[39] W. Kern and D. Puotinen, *Cleaning Solutions Based on Hydrogen Peroxide for Use in Silicon Semiconductor Technology*, RCA Review, **31**, 187 – 206, 1970.

[40] G. Klimeck, R. Lake, R. Bowen, W. Frensley and T. Moise, *Quantum Device Simulation with a Generalized Tunneling Formula*, Appl. Phys. Lett., **67**, 2539 – 2541, 1995.

[41] R. Lake and S. Datta, *Nonequilibrium Green's-Function Method Applied to Double-Barrier Resonant-Tunneling Diodes*, Phys. Rev. B, **45**, 6670 – 6685, 1992.

[42] R. Lake, G. Klimeck, R. Bowen and D. Jovanovic, *Single and Multiband Modeling of Quantum Electron Transport Through Layered Semiconductor Devices*, J. Appl. Phys., **81**, 7845 – 7869, 1997.

[43] S.-H. Lon, D. Buchanan, Y. Taur and W. Wang, *Quantum-Mechanical Modeling of Electron Tunneling Current from the Inversion Layer of Ultra-Thin-Oxide nMOSFET's*, IEEE Electron Dev. Lett., **18**, 209 – 211, 1997.

[44] B. Majkusiak, T. Janik and J. Walczak, *Semiconductor Thickness Effects in the Double-Gate SOI MOSFET*, IEEE Trans. Electron Dev., **45**, 1127 – 1134, 1998.

[45] A. Merlos, M. Acero, M. Bao, J. Bausells and J. Esteve, *TMAH/IPA Anisotropic Etching Characteristics*, Sensors Actuators, **A37-38**, 737 – 743, 1993.

[46] T. Mikolajick, V. Haeublein and H. Ryssel, *The Effect of Random Dopant Fluctuations on the Minimum Channel Length of Short-Channel MOS Transistors*, Appl. Phys. A, **64**, 555 – 560, 1997.

[47] J. Moers, *Entwicklung und Charakterisierung von vertikalen Silizium-MOSFETs auf der Basis von selektiver Epitaxie*, Ph.D. thesis, RWTH Aachen, 1998.

[48] G. Moore, *Cramming more Components onto Integrated Circuits*, Electronics, **38**, 114 – 117, 1965.

[49] K. Natori, *Ballistic Metal-Oxide-Semiconductor Field Effect Transistor*, J. Appl. Phys., **76**, 4879 – 4890, 1994.

[50] Y. Naveh and K. Likharev, *Shrinking Limits of Silicon MOSFETs: Numerical Study of 10nm Scale Devices*, Superlatt. Microstruct., **27**, 111 – 123, 2000.

[51] E. Nicollian and J. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*, Wiley-Interscience, New York, 1982.

[52] H. Noda, H. Sakiyama, Y. Goto, T. Kure and S. Kimura, *Tungsten Gate Technology for Quarter-Micron Application*, Jpn. J. Appl. Phys., **35**, 807 – 811, 1996.

[53] Y. Omura, *Features of Ultimately Miniaturized MOSFETs/SOI: A New Stage in Device Physics and Design Concepts*, IEICE Trans. Electron., **E80-C**, 394 – 406, 1997.

[54] Y. Omura, S. Horiguchi, M. Tabe and K. Kishi, *Quantum-Mechanical Effects on the Threshold Voltage of Ultrathin-SOI nMOSFETs*, IEEE Electron Dev. Lett., **14**, 569 – 571, 1993.

[55] Y. Omura, K. Kurihara, Y. Takahashi, T. Ishiyama, Y. Nakajima and K. Izumi, *50-nm Channel nMOSFET/SIMOX with an Ultrathin 2- or 6-nm Thick Silicon Layer and Their Significant Features of Operation*, IEEE Electron Dev. Lett., **18**, 190 – 193, 1997.

[56] F. Pikus and K. Likharev, *Nanoscale Field-Effekt Transistors: An Ultimate Size Analysis*, Appl. Phys. Lett., **71**, 3661 – 3663, 1997.

[57] W. Potz, *Self-Consistent Model of Transport in Quantum Well Tunneling Structures*, J. Appl. Phys., **66**, 2458 – 2466, 1989.

[58] W. Press, S. Teukolky, W. Vetterling and B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.

[59] B. Puers and W. Sansen, *Compensation Structures for Convex Corner Micromachining in Silicon*, Sensors and Actuators, **A21-A23**, 1036 – 1041, 1990.

[60] P. Rai-Choudhury, *Microlithography. Handbook of Microlithography, Micromachining and Microfabrication*, volume 1, SPIE, 1997.

[61] E. Rauly, O. Potavin, F. Balestra and C. Raynaud, *On the Subthreshold Swing and Short Channel Effects in Single and Double Gate Deep Submicron SOI-MOSFETs*, Solid-State Electronics, **43**, 2033 – 2037, 1999.

[62] W. Richards and M. Shen, *Extraction of Two-Dimensional Metal-Oxide Semiconductor Field Effect Transistor Structural Information from Electrical Characteristics*, J. Vac. Sci. Technol. B, **18**, 533 – 539, 2000.

[63] S. Saha, *Effects of Inversion Layer Quantization on Channel Profile Engineering for NMOSFETs with 0.1 μm Channel Lengths*, Solid-State Electronics, **42**, 1985 – 1991, 1998.

[64] H. Schroeder, *Modell des anisotropen Aetzens von einkristallinem Silizium in waessrigen KOH-Loesungen*, Ph.D. thesis, Technische Universitaet Berlin, 2000.

[65] H. Seidel, L. Csepregi, A. Heuberger and H. Baumgaertel, *Anisotropic Etching of Crystalline Silicon in Alkaline Solutions. I. Orientatin Dependence and Behavior of Passivation Layers*, J. Electrochem. Soc., **137**, 3612 – 3626, 1990.

[66] J. Shiely and H. Massoud, *Simulation of the Drain-Current Characteristics of MOSFETs with Ultrathin Oxides in the Presence of Direct Tunneling*, Microelectronic Engineering, **48**, 101 – 104, 1999.

[67] A. Spinelli, A. Benvenuti and A. Pacelli, *Self-Consistent 2-D Model for Quantum Effects in n-MOS Transistors*, IEEE Trans. Electron Dev., **45**, 1342 – 1349, 1998.

[68] F. Stern, *Self-Consistent Results for n-Type Si Inversion Layers*, Phys. Rev. B, **5**, 4891 – 4899, 1972.

[69] D. Streit, E. Ahlers and F. Allen, *Sharp Profiles and Low-Temperature Diffusion of Ga and Sb in Silicon Modulation-Doped Superlattices*, J. Vac. Sci. Technol. B, **5**, 752– 756, 1987.

[70] D. Streit, R. Metzger and F. Allen, *Doping of silicon in Molecular Beam Epitaxy Systems by Solid Phase Epitaxy*, Appl. Phys. Lett., **44**, 234 – 236, 1984.

[71] J. Sune, P. Olivo and B. Ricco, *Self-Consistent Solution of the Poisson and Schroedinger Equations in Accumulated Semiconductor-Insulator Interfaces*, J. Appl. Phys., **70**, 337 – 343, 1991.

[72] E. Suzuki, K. Ishii, S. Kanemaru, T. Maeda, T. Tsutsumi, T. Sekigawa, K. Nagai and H. Hiroshima, *Highly Suppressed Short-Channel Effects in Ultrathin SOI n-MOSFET's*, IEEE Trans. Electron Dev., **47**, 354 – 359, 2000.

[73] S. Sze, *Physics of Semiconductor Devices*, Wiley, New York, 2nd edition, 1981.

[74] Y. Taur, *MOSFET Channel Length: Extraction and Interpretation*, IEEE Trans. Electron Dev., **47**, 160 – 170, 2000.

[75] Y. Taur, D. Buchanan, W. Chen, D. Frank, K. Ismail, S.-H. Lo, G. Sai-Halasz, R. Viswanathan, H.-J. Wann, S. Wind and H.-S. Wong, *CMOS Scaling into the Nanometer Regime*, Proc. IEEE, **85**, 486 – 504, 1997.

[76] Y. Taur and T. Ning, *Fundamentals of modern VLSI Devices*, Cambridge University Press, 1998.

[77] G. Timp, K. Bourdelle, J. Bower, F. Baumann, T. Boone, R. Cirelli, K. Evans-Lutterodt, J. Garno, A. Ghetti, H. Gossmann, M. Green, D. Jacobson, Y. Kim, R. Kleiman, F. Klemens, A. Kornbit, C. Lochstampfer, W. Mansfield, S. Moccio, D. Muller, L. Ocola, M. O'Malley, J. Rosamilia, J. Sapjeta, P. Silverman, T. Sorsch, D. Tennant, W. Timp and B. Weir, *Progress Toward 10nm CMOS Devices*, in *International Electron Devices Meeting 1998. Technical Digest*, 615 – 618, IEEE, 1998.

[78] H. Trieu and W. Mokwa, *A Generalized Model Describing Corner Undercutting by the Experimental Analysis of TMAH/IPA*, J. Micromech. Microeng., **8**, 80 – 83, 1998.

[79] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, WCB/McGraw-Hill, 2. edition, 1999.

[80] T. Tsuchiya, Y. Sato and M. Tomizawa, *Three Mechanisms Determining Short-Channel Effects in Fully-Depleted SOI MOSFET's*, IEEE Trans. Electron Dev., **45**, 1116 – 1121, 1998.

[81] S. Veeraraghavan and J. Fsooum, *Short-Channel Effects in SOI MOSFET's*, IEEE Trans. Electron Dev., **36**, 522 – 528, 1989.

[82] F. Vesely, *Computational Physics. An Introduction*, Plenum Press, New York, 1994.

[83] H. Wakabayashi, M. Ueki, M. Narihiro, T. Fukai, N. Ikezawa, T. Matsuda, K. Yoshida, K. Takeuchi, Y. Ochiai, T. Mogami and T. Kunio, *45-nm Gate Length CMOS Technology and Beyond Using Steep Halo*, in *International Electron Devices Meeting 2000. Technical Digest*, 49 – 52, 2000.

[84] H.-S. Wong, D. Frank and P. Solomon, *Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25 nm Channel Length Generation*, in *International Electron Devices Meeting 1998. Technical Digest*, 407 – 410, IEEE, 1998.

[85] H.-S. Wong, D. Frank, P. Solomon, C. Wann and J. Welser, *Nanoscale CMOS*, Proc. IEEE, **87**, 537 – 569, 1999.

[86] H.-S. Wong, D. Frank, Y. Taur and J. Stork, *Design and Performance Considerations for Sub-0.1μm Double-Gate SOI MOSFET's*, in *IEDM Technical Digest*, 1994.

[87] H.-S. Wong, Y. Taur and D. Frank, *Discrete Random Dopant Distribution Effects in Nanometer-Scale MOSFETs*, Microelectronics Reliability, **38**, 1447 – 1456, 1998.

[88] R.-H. Yan, A. Ourmazd and K. Lee, *Scaling the Si MOSFET: From Bulk to SOI to Bulk*, IEEE Trans. Electron Dev., **39**, 1704 – 1710, 1992.

[89] K. Young, *Analysis of Conduction in Fully Depleted SOI MOSFET's*, IEEE Trans. Electron Dev., **36**, 504 – 506, 1989.

[90] K. Young, *Short-Channel Effect in Fully Depleted SOI MOSFET's*, IEEE Trans. Electron Dev., **36**, 399 – 402, 1989.

[91] Q. Zhao, F. Klinkhammer, M. Dolle, L. Kappius and S. Mantl, *Nanometer Patterning of Epitaxial $CoSi_2$/Si(100) for Ultrashort Channel Schottky Barrier Metal-Oxide-Semiconductor Field Effect Transistors*, Appl. Phys. Lett., **74**, 454 – 456, 1999.

Acknowledgments

I am indebted to the following people who contributed in various ways to this thesis.

- First of all, I would like to thank Prof. Dr. B. Lengeler for providing the opportunity to work in his institute and for his continuous interest in the physical *and* technological aspects of the V-groove MOSFET. Furthermore, I am very grateful for his support concerning my post-doc plans.

- I am also indebted to Prof. Dr. H. Lüth for undertaking the Korreferat of this work and for the opportunity to use the facilities of his institute at the FZ Jülich.

- My dearest thanks go to Dr. J. Appenzeller who pushed me right into this successful and very exciting project. He gave me the freedom to work independently but has always been the one who put me back on the right track when I was about to lose the thread. Jörg, it was a great pleasure working together with you and sharing your experience and enthusiasm although we were most of the time separated by a marginal puddle called the Atlantic Ocean. Long live the WWW and DHL. Hope, we can continue soon.... Thanks for all the private conversations, too, which I enjoyed a lot.

- Thanks to Dr. R. Martel, Dr. P. Solomon and Dr. Ph. Avouris from IBM T.J. Watson Research Center, Yorktown Heights for their confidence to let me take part in this project. And thanks go to Prof. J. del Alamo and J. Scholvin from the Massachusetts Institute of Technology for their advice and for contributing one of the most delicate parts of the MOSFET fabrication.

- Furthermore, I would like to thank M. Tanner, S. Thomas, Y. Lu and K.L. Wang from the University of California at Los Angeles for providing the silicon MBE samples.

- Thanks to F. Günzler, my late room-mate, who installed a webcam so that I can check the level of the coffee pot behind me without turning around. Great invention.... I will miss the conversations about..., well everything and, of course, your PowerPoint presentations. And thanks to Dr. Ch. Schroer, who was always ready to discuss physical and numerical issues with an incredible energy and enthusiasm.

- Moreover, I am indebted to my colleagues Th. Hunger, Ch. Zimprich, M. Kuhlmann, B. Benner and A. Gerhardus who shared an office or, during the last weeks, their computers with me. Thomas, I was so disappointed that you do not have a Saxonian accent. And certainly I am grateful to my former colleagues Dr. H. Stahl and Dr. M. Jacob who not only set up and maintained most of the equipment but also provided a lot to the nice atmosphere in the institute. Heiko, we miss your cakes. Hope, that we manage to keep in touch.

- I gratefully acknowledge the continuous help and support of the colleagues in Jülich, in particular A. Steffen, J. Müller, F.-J. Zillikens, M. Nonn, A. Pracht.

- Special thanks go to Prof. Dr. S. Datta of Purdue University who probably will never read this thesis but gave me the crucial hint to solve the self-consistency problem of my computer program.

- Last but not least I thank my wife and my family for their patience, support and all the time they let me off instead of spending it with them during the last three years...

Curriculum Vitae

- born on March, 14th, 1972 in Aachen, Germany
- 1978 - 1982: elementary school
- 1982 - 1991: secondary school, Bischöfliches Piusgymnasium, Aachen
- 1991: Abitur (A-levels)
- 1991 - 1992: civilian service

- 1992 - 1994: study of physics at the Aachen University of Technology (RWTH Aachen)
- 1994 - 1995: associate student at Queen Mary and Westfield College, University of London, UK (now Queen Mary, University of London)
- 1995 - 1998: resumption of the studies at the RWTH Aachen
- 1997 - 1998: masters thesis at the II. Physikalisches Institut B, investigations of superconductor-semiconductor contacts using III-V semiconductor-heterostructures
- 1998: Diplom (masters) in physics, summa cum laude
- since 1998: Ph.D. student at the II. Physikalisches Institut B