Mesoscopic Phenomena in Nanometer Scale MOS Devices

von der

Fakultät für Elektrotechnik der

Universität Dortmund

genehmigte Dissertation

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

von

Gilson Inácio Wirth

Dortmund 1999

Tag der mündlichen Prüfung: 4. August 1999
Hauptreferent: Prof.Dr.-Ing. K. Goser
Korreferent: Prof.Dr.rer.nat. W. R. Fahrner

Contents

List of Symbols

B	Magnetic field
c	Speed of light in vacuum
C_{ox}	Oxide capacitance per unit area
C_{ID}	Capacitive coupling between island and drain
C_{IG}	Capacitive coupling between island and gate
C_{IS}	Capacitive coupling between island and source
C_{IS}	Total island capacitance
e	Magnitude of the electron charge
E_B	Energy barrier
E_C	Conduction band edge
E_{ch}	Charging energy
E_F	Fermi level
E_{Fi}	Intrinsic Fermi level
E_T	Trap energy level
E_V	Valence band edge
FET	Field Effect Transistor
h	Planck's constant
$\hbar$	Planck's constant divided by 2π
g_m	Drain conductance of a transistor
g_m	Transconductance of a transistor
I_B	Bulk current
I_D	Drain current
I_G	Gate current
I_S	Source current
k_B	Boltzmann constant
l	Mean free path
L	Channel length
L_{design}	Design channel length
L_{eff}	Effective channel length
MBE	Molecular Beam Epitaxy
MOS	Metal-oxide-semiconductor
$MOSFET$	Metal-oxide-semiconductor field effect transistor
m_e	Free electron mass
m_*	Charge carrier effective mass

m_e^*	Electron effective mass
m_h^*	Hole effective mass
n	Electron concentration in the conduction band
n_i	Intrinsic electron concentration in the conduction band
NDR	Negative differental resistance
N_a	Acceptor density
N_d	Donor density
N_e	Density of states of electrons in the conduction band
N_a	Acceptor density
N_{inv}	Inversion layer charge density
N_{trap}	Trapped charge density
p	Hole concentration in the valence band
p_i	Intrinsic hole concentration in the valence band
q	Magnitude of electron charge
Q_{inv}	Inversion layer charge per unit area
Q_{dep}	Depletion charge per unit area
Q_I	Net electrical charge at the island
RTS	Random telegraph signal
SOI	Silicon-on-insulator
T	Temperature
T_e	Electron temperature
T_l	Lattice temperature
t_{ox}	Gate oxide thickness
v	Velocity of the electron
$\overline{v}$	Mean velocity of the electrons
v_s	Saturation velocity of the electron
v_{th}	Thermal velocity of the electron
V_B	Bulk voltage
V_D	Drain voltage
V_{FB}	Flat-band voltage
V_G	Gate Voltage
V_{GB}	Gate to bulk bias
V_{GS}	Gate to source bias
V_I	Voltage at the island
V_S	Source voltage
V_T	Threshold voltage
W	Channel width
ε_0	Free space permittivity
ε_{sc}	Static dielectric constant of the semiconductor
ε_{ox}	Static dielectric constant of the oxide
ϕ_S	Surface potential
ϕ_B	Bulk potential

$\vec{k}$	Wave vector
k_x	Component of the wave vector in the x-direction
k_y	Component of the wave vector in the y-direction
k_z	Component of the wave vector in the z-direction
λ	Wave length
λ_F	Fermi wave length
μ	Charge carrier mobility
μ_n	Electron mobility
μ_p	Hole mobility
ω_c	Cyclotron frequency
$\vec{p}$	Momentum of a particle
ψ	Electron wavefunction
τ_c	Mean electron capture time
τ_e	Mean electron emission time

1. Introduction

The key point in the success of the semiconductor industry is its ability to continuously provide electronic products with decreasing cost per function as well as, at the same time, increasing performance. This is a result of a steady reduction in the feature size combined with a steady rise in density. In this framework, MOSFETs are being aggressively scaled down to dimensions well below 100 nm. However, at a given device dimension, the scaling of the physical processes breaks down and new phenomena that are absent in larger structures can dominate device behavior. The SIA roadmap (SIA - Semiconductor Industry Association 1997) predicts a minimum feature size of 35 nm and 10^8 transistors per cm^2 for the commercial CMOS technology around 2012. It is now quite obvious that the behavior of the devices that will build up the circuits produced with this technology will significantly deviate from the behavior of their counterparts of greater dimensions. This means that the modeling and characterization community has to quickly respond to the challenges related to the further shrinking of device size. Subsequently, if devices are not properly characterized and modeled, this can turn out to be a barrier for further development in the semiconductor industry. The main scope of this work is to help and speed up the proper characterization and modeling of sub-100 nm MOSFETs.

The main purpose of modeling and characterization is to make possible the quantitative investigations needed for optimization and diagnosis, but it is also important to provide insight into technology directions and potentials. Even if the requirements for the commercial application of a particular technology node are not met, modeling and characterization is still a very useful activities since it provides many important technical insights.

Characterization and modeling are known to consist of three steps that are intimately related:

1. Elaboration of Models: Models are mental abstractions of reality, created with the help of mathematical tools.

2. Simulation: Taking advantage of modern computational facilities, models can be implemented in computer codes, which can then be used to predict the behavior of objects under study in an efficient way.

3. Calibration and Validation: The results obtained in the simulation step are compared to representative experimental data to validate the model and to obtain adequate numerical values for parameters.

These three activities compose a cycle that can be interactively repeated until a suitable model is achieved.

In the scope of the work presented in the next chapters, electrical measurements were carried out on samples processed at the facilities available at the *Lehrstuhl für Baulemente der Elektrotechnik*. The measured data represents the reality to be modeled. This work deals with physical models. Physical models intend to represent mathematically the understanding of the physical properties of the devices being studied.

In the subsequent steps, the experimental data obtained is compared to the predictions of available models, some of them implemented in simulators. Numerical values for parameters are obtained and the adequacy of the model is checked. If the model shows deficiencies, a new approach to overcome the deficiencies is proposed. Eventually the cycle is repeated, with new measurements and simulations being performed. This is the methodology used in this work.

The term mesoscopic phenomena, the main subject of this work, has been introduced to describe the characteristics of systems that are neither microscopic (one or few atoms) nor macroscopic. Meso is borrowed from the Greek, meaning middle. In such systems the wave nature of electron transport or the discrete charge nature of electrons may become relevant.

Before the main results obtained are reported, the basic theory needed to follow the discussions undertaken is reviewed in chapter 2. A brief description of the particular technology used to process the devices studied is also presented in this chapter. In the next chapter, sample preparation and the experimental setup used for electrical characterization are described.

In chapter 4 experimental results showing mesoscopic phenomena in the electrical behavior of sub-100 nm MOSFETs at low temperature are presented.

Unexpected periodic oscillations with negative differential resistance (NDR) are present in the $I_D \ x \ V_G$ characteristics. The oscillations, present from sub-threshold up to strong inversion, are reproducible from sample to sample and with temperature cycling. Investigations are undertaken to verify if there is a relation between device geometry an oscillation pattern. Measurements are carried out at temperatures between 300 mK and 35 K, and magnetic fields up to 15 T are applied.

This is the first time that such phenomena are reported for a conventional bulk MOS system, suggesting that the electrical behavior of ultra short channel devices may still be an open question.

In chapter 5 various electric transport models for small size MOS systems are studied. For several reasons, Coulomb blockade seems to be a rather plausible explanation for the mesoscopic phenomena presented in chapter 4, although some questions, discussed in sections 5.5 and 5.6, remain open.

In chapter 6 another single electron switching phenomena is discussed. Namely, the possibility of studying the behavior of individual defect sites in the oxide is demonstrated. The capture and emission of charge carriers from a single defect site brings up discrete changes in the drain current. This is known as the random telegraph signal (RTS). Studying the bias-voltage dependence of the RTS it is possible to determine the geometrical and energetical location of these defects. It is shown that a trap can be used as a probe into the local surface potential. It is also shown that the behavior of the RTS does depend on the properties of the channel electrons close to the defect state, making RTS analysis a valuable way to study effects as Coulomb scattering, electron gas heating and the mechanisms of relevance for electrical channel formation in very small area devices.

If the active device area is decreased, the relative importance of the channel region affected by the defect increases. The study of the random telegraph signal in very small area devices is thus of significance. Noise in general, and the RTS noise in particular, may become an issue for advanced circuitry. The steady lowering of the supply voltages combined with the decrease in device active area will make systems more sensitive to fluctuations in device characteristics.

The investigations carried out in this work intend to contribute to the better understanding of the properties of ultra short MOS devices, a necessary issue to produce optimized and reliable systems, leading ultimately to better products.

2. The MOSFET

This chapter starts with a brief review of the operation principles of the Metal Oxide Semiconductor Field Effect Transistor (MOSFET), followed by a description of the temperature dependence of the parameters relevant to this work. After that the effects of the quantization of the transverse electron motion in the inversion layer are discussed.

The last section of this chapter describes briefly the particular technology used to process the devices at nanoscale.

2.1 Principles of Operation

The MOSFET is a four terminal device, as shown in Fig. 2.1. The gate and the bulk terminals define a capacitor where the gate is insulated from Si by a dielectrica, such as SiO_2. The region beneath the gate oxide and between source and drain is called the channel. First, it will be assumed that bulk, drain and source are at the same potential. Depending on whether V_G, the voltage between gate and the other terminals, is equal to, less than, or greater than the flat-band voltage V_{FB}, the channel can be in the flat-band condition, in accumulation, in depletion or inversion. The flat-band voltage is the external voltage used between the gate and the bulk terminals to keep the channel neutral by canceling the effects of the contact potentials and the ”parasitic” charge that exists within the oxide as well as at the oxide-semiconductor interface.

For the NMOS device of Fig. 2.1, if a voltage less than the flat-band voltage is applied between gate and bulk a positive charge is induced at the oxide-semiconductor interface. As the bulk is p-type, accumulation of excess holes occur at the interface.

If V_G increases above V_{FB} the positive charge on the gate will induce a negative charge in the channel. If V_G is not much higher than V_{FB}, the positive potential at the surface with respect to the bulk will simply drive holes away from the surface, leaving it depleted from mobile carriers. This condition is called depletion. The charge in the channel consists of ionized acceptor atoms.

If V_G increases further, more acceptor atoms are uncovered and the potential in the channel becomes sufficiently positive to attract a significant number of free electrons to the surface. Eventually, with a sufficiently high V_G, the density of electrons will exceed that of holes at the surface. Now one has the so called inversion layer.

The inversion layer is contacted electrically at the two ends by the source and the drain. By applying a voltage between these ends, a current can be caused to flow in the layer. Since the number of carriers available for conduction depends on the gate potential, the latter can be used to create or eliminate the inversion layer as well as to modulate its conduction.

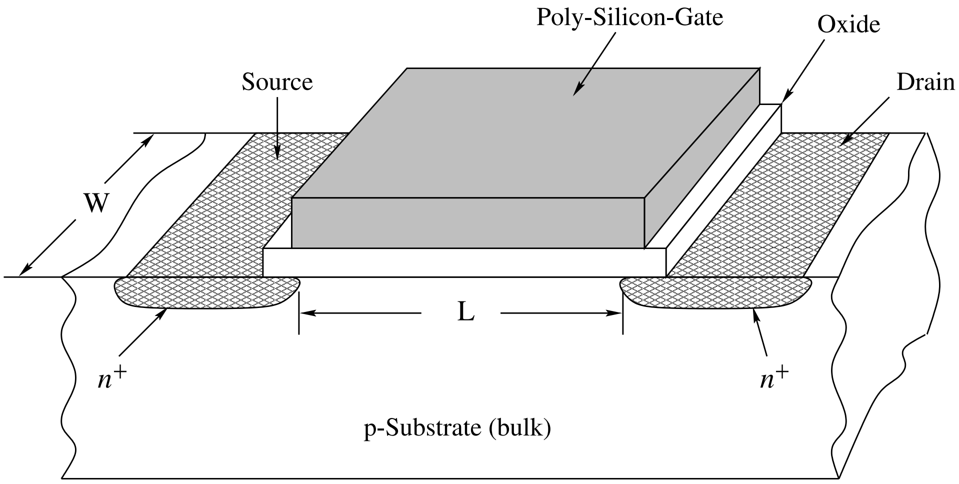


Figure 2.1: Simplified structure of a n-channel MOS transistor.

2.1.1 The Inversion Layer

Due to the importance of the inversion layer in this work, it will be now examined in quantitative detail. In Fig. 2.2 the bandbending of the semiconductor in the channel region is shown. The bandbending is described by the quantity $e\phi(z)$, which measures the position of the intrinsic Fermi level with respect to the bulk intrinsic Fermi level, and its value at the oxide-semiconductor interface (where $z = 0$) is $e\phi_S$, as shown in Fig. 2.2. The $z-axis$ is assumed perpendicular to the oxide-semiconductor interface, increasing in the direction into the semiconductor. Let $E_{Fi}(z)$ be the intrinsic Fermi level, i.e. the Fermi level in an undoped semiconductor, and $e\phi_B$ be the difference between the intrinsic Fermi level and the Fermi level in the bulk. Now a quantitative definition of the Flat-Band, Accumulation, Depletion and Inversion conditions qualitatively described above is possible:

- $\phi_S < 0$: Accumulation.
- $\phi_S = 0$: Flat-band.
- $0 < \phi_S < \phi_B$: Depletion.
- $\phi_S > \phi_B$: Weak inversion.
- $\phi_S > 2\phi_B$: Strong inversion.

Because of bandbending, the electron distribution in the channel depends upon the position z and upon the gate voltage. In the bulk region the gate potential does not affect the electron concentration. This electron concentration, called n_0, is (Singh 1994)

$$n_0 = n_i exp\left(\frac{-e\phi_B}{k_B T}\right) \tag{2.1}$$

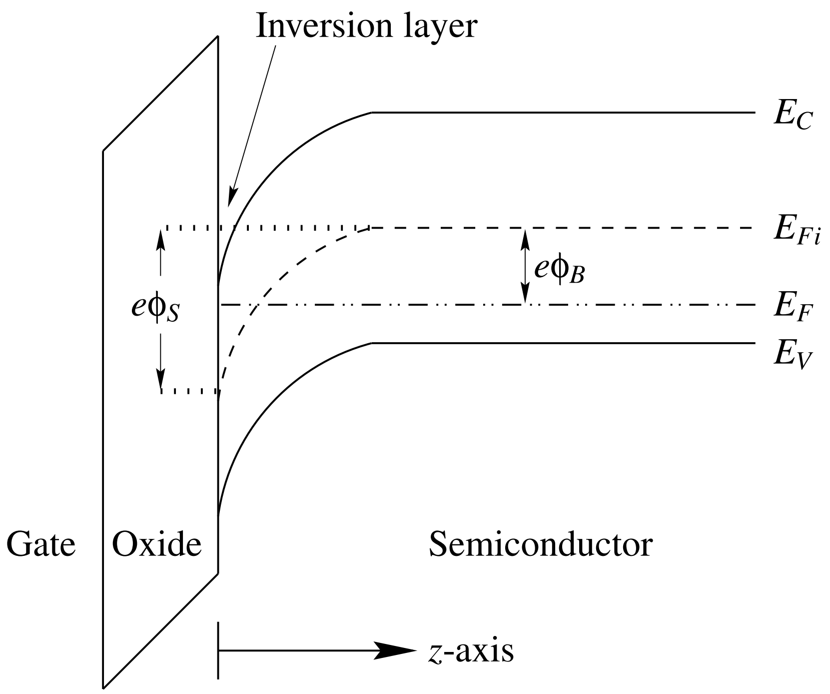


Figure 2.2: The bandbending in the p-semiconductor due to a positive gate bias V_G, leading to an inversion layer.

where n_i is the intrinsic carrier concentration, T the temperature and k_B the Boltzmann constant.

Using Boltzmann statistics, the position dependent electron concentration in the inversion-layer can be calculated as being

$$n(z) = n_0 exp\left(\frac{e\phi(z)}{k_B T}\right) \tag{2.2}$$

The hole concentration is similarly given by

$$p(z) = p_0 exp\left(\frac{-e\phi(z)}{k_B T}\right) \tag{2.3}$$

where p_0 is the equilibrium hole concentration in the bulk, that can be calculated similarly as n_0 in eq. 2.1.

Solving the Poisson equation using these charge-potential relations one can relate the charge to the externally applied potential. It comes out that once the strong inversion condition is satisfied, the depletion width does not change because the free carriers induced after inversion start preventing further depletion. In fact, because of the large density of states in the conduction band, once strong inversion is achieved the surface potential ϕ_S gets virtually pinned to the value $2\phi_B$. The width of the depletion region at this point is given by

$$W_{max} = \sqrt{\frac{4\varepsilon_{sc}\phi_B}{eN_a}} \tag{2.4}$$

where ε_{sc} is the permittivity of the semiconductor and N_a is the channel doping. The corresponding value of V_G (at the point where strong inversion is reached) is called the threshold voltage (Singh 1994)

$$V_T = V_G(\phi_S = 2\phi_B) = V_{FB} + 2\phi_B + \frac{\sqrt{4e\varepsilon_{sc}N_a\phi_B}}{C_{ox}} \tag{2.5}$$

For $V_G > V_T$, the charge in the channel consists of the free carriers in the inversion layer and the fixed charge of the dopants in the depletion region. For this bias the depletion layer charge is constant, and a small change δV_G in the gate potential will induce a change δQ_{inv} in the inversion layer charge given by

$$\delta Q_{inv} = \delta V_G \frac{\varepsilon_{ox}}{t_{ox}} = \delta V_G \cdot C_{ox} \tag{2.6}$$

where t_{ox} is the gate oxide thickness.

2.1.2 Current-Voltage Characteristics

Now a quantitative analysis of what happens if a potential is applied at the drain contact of a MOSFET is given. It will be assumed that source and bulk are at the same potential, as it is usually the case in practical applications.

The first step is to establish the relation between the channel charge at any point x along the channel and the gate and drain potentials. As the inversion layer is contacted at the drain side its potential will no longer be pinned at $2\phi_B$. The extension of the depletion layer is then not limited to the value implied by 2.4. Assuming that both ends of the channel are in strong inversion, the surface potential, now a function of x, the direction along the channel, can be written as

$$\phi_S(x) = 2\phi_B + V_c(x) \tag{2.7}$$

where $V_c(x)$ is the x-dependent channel potential, assumed to zero at the source and equal to V_D at the drain.

Solving the Poissons equation again, using the charge-potential relations given by 2.2 and 2.3, the position dependent inversion layer charge in the channel comes out to be (Singh 1994)

$$Q_{dep}(x) = -\sqrt{2\varepsilon_{sc}eN_a[V_c(x) + 2\phi_B]} \tag{2.8}$$

As long as both ends of the channel are in strong inversion the current I_D between drain and source is mainly due to carrier drift. Making the usual constant mobility μ_n approximation it will be given by

$$I_D = -Q_{inv}\mu_n W \frac{\partial V_c(x)}{\partial x} \tag{2.9}$$

where W is the channel width.

The position dependent charge in the inversion layer can be calculated subtracting the depletion layer charge from the total charge in the channel

$$Q_{inv}(x) = C_{ox}[V_G - 2\phi_B - V_{FB} - V_c(x)] - \sqrt{2\varepsilon_{sc} e N_a [V_c(x) + 2\phi_B]} \tag{2.10}$$

Integrating 2.9 from source ($x = 0$) to drain ($x = L$) using the above value for $Q_{inv}(x)$ leads to

$$I_D = \frac{W}{L}\mu_n C_{ox}[(V_G - 2\phi_B - V_{FB})V_D - \frac{V_D^2}{2} - \frac{\sqrt{8\varepsilon_{sc} e N_a}}{3C_{ox}}((V_D + 2\phi_B)^{3/2} - (2\phi_B)^{3/2})] \tag{2.11}$$

where L is the length of the device channel.

Equation 2.11 is valid as long as both ends of the channel are in strong inversion. For a sufficiently high drain bias, the potential difference between the drain end of the channel and the bulk will be dropped completely over the depletion layer and no inversion layer will be available at that end, i.e., the channel *pinches off* at the drain side. This defines the saturation drain voltage $V_D sat$, as the value of V_D for which $Q_{inv}(x = L) = 0$.

The value of V_{Dsat} can be derived equating $Q_{inv} = 0$ at $x = L$ in 2.10:

$$V_{Dsat} = V_G - 2\phi_B - V_{FB} + \frac{\varepsilon_{sc} e N_a}{C_{ox}^2}\left[1 - \sqrt{1 + \frac{2(V_G - V_{FB})Cox^2}{\varepsilon_{sc} e N_a}}\right] \tag{2.12}$$

The above derivation of I_D (Eq. 2.11) remains valid only as long as $V_D < V_{Dsat}$, called the linear region. Beyond pinch off the treatment of the current problem is quite complex (Chen, Ma, Kuo, Yu, and Dutton 1995). As all relevant measurements in this work are carried out in the linear region, this problem will not be discussed in detail here. However, for long enough devices, beyond pinch off the current essentially remains constant.

For very small drain bias $V_D << (V_G - V_{FB} - 2\phi_F)$ one can consider the level of inversion being the same along the whole channel (Q_{inv} independent of x) and write the current-voltage relation in terms of the threshold voltage as

$$I_D \approx \frac{W}{L}\mu_n C_{ox}(V_G - V_T)V_D \tag{2.13}$$

In this low drain bias region the current-voltage relations are linear and the device is said to be operating in the *linear region*.

As mentioned earlier, an increasing drain bias causes the channel to pinch off at the drain end. In a simple approximation, it can be assumed that after V_D exceeds V_{Dsat} the current keeps constant (Brews 1978). Another simplification generally accepted is to assume the onset of pinch off to occur when the potential applied between gate an drain equals threshold, i.e., $V_{Dsat} = V_G - V_T$. This leads to the traditional form for the current in the saturated region (Tsividis 1987)

$$I_{Dsat} = \frac{W}{2L}\mu_n C_{ox}(V_G - V_T)^2 \tag{2.14}$$

According to this equation, in the saturated region, the current has a square law dependence upon the gate bias and does not depend upon the drain bias.

The above current-voltage relations are based on 2.9, which assumes that the current is mainly due to drift. In weak inversion this is not the case. In weak inversion most of the current is caused by carrier diffusion.

Diffusion occurs whenever the charge carriers are not uniformly distributed along the channel, that is, when carrier concentration gradients exist; then the thermal random motion of the carriers tends to cause the spreading out from regions of high concentration to regions of low concentration. The diffusion current can be written as

$$I_{diff}(x) = \mu_n W \frac{k_B T}{q}\frac{\partial Q_{inv}}{\partial x} \tag{2.15}$$

Assuming that in weak inversion the only significant component of the current is due to diffusion (Brews 1979) and noting that in dc steady state, the case of relevance for this work, the total current in the channel must be the same for all x along the channel and equal to the drain current. Therefore, it follows from the form of 2.15 that $\partial Q_{inv}/\partial x$ must be a constant, i.e., Q_{inv} *versus* x must be a straight line, leading to

$$I_D = W\mu\frac{k_B T}{q}\frac{Q_{inv}(x=0) - Q_{inv}(x=L)}{L} \tag{2.16}$$

The above two values of Q_{inv} can be found solving Poissons equation again, using an inversion layer concentration of the form 2.2 that depends on the coordinate x along the channel, and considering that the difference between the surface potential at the drain end and at the source end of the inversion layer must be equal to the drain voltage [$\phi_S(x=L) - \phi_S(x=0) = V_D$]. This leads to the following expression for drain current (Brews 1979)

$$I_D = \frac{W}{L}\mu Q_{inv}(0)\left(1 - e^{-\frac{qV_D}{k_B T}}\right) k_B T \tag{2.17}$$

where $Q_{inv}(0)$ is the inversion layer charge at the source end of the channel (Brews 1978),

$$Q_{inv}(0) = qN_A L_B \sqrt{2} \left(\sqrt{\frac{q\phi_S(0)}{k_B T} - 1 + \left(\frac{n_i}{N_A}\right)^2 e^{\frac{q\phi_S(0)}{k_B T}}} - \sqrt{\frac{q\phi_S(0)}{k_B T} - 1} \right) \tag{2.18}$$

where L_B is the bulk Debye length, $L_B = (k_B T \varepsilon_{sc}/q^2 N_A)$, and the band bending at the source end of the channel $\phi_S(0)$ can be related to the gate bias by (Brews 1979)

$$\phi_S(0) = V_G - V_{FB} - \frac{k_B T}{q}\left(\frac{\varepsilon_{sc} t_{ox}}{\varepsilon_{ox} L_B}\right)^2 \left(\sqrt{1 + \frac{2\varepsilon_{ox} L_B}{\varepsilon_{sc} t_{ox}} \left((V_G - V_{FB}) \frac{q}{k_B T} - 1 \right)} - 1 \right) \tag{2.19}$$

These equations lead to a drain current that increases exponentially with V_G. For drain biases V_D larger than a few $(k_B T/q)$ the current in the sub threshold region will saturate (i.e., is independent of V_D). Therefore, it is a matter of continuity to suppose that the pinch off condition also prevails in the sub threshold region when the current has saturated.

2.2 Temperature Dependence

MOSFET characteristics are known to be strongly temperature-dependent (Ytterdal, Hurt, Shur, Park, Tsai, and Peatman 1996; Borzdov and Petrovich 1997; Chen, Ma, Kuo, Yu, and Dutton 1995). One of the parameters whose temperature dependence has been extensively studied is the mobility μ of the charge carriers in the channel.

According to the Bloch theorem, in a perfect periodic potential, the electrons show a free electron-like behavior where no scattering is allowed. But in real semiconductors there are imperfections and scattering does occur. The most important sources of scattering are Coulomb centers, surface roughness and lattice vibrations (phonons). Each of them depicts a very particular temperature dependence.

Lattice vibrations are an inevitable source of scattering and dominate the scattering process near room temperature for high quality samples, i.e. with small interface roughness. In the range of inversion layer electron concentrations of $N_{inv} = 0.5 - 5x10^{12} cm^{-2}$ and around room temperature, the mobility μ is known to be proportional to T^{-a}, with $1 < a < 1.5$. This behavior of mobility is generally assumed to be determined by phonon scattering (Ando, Fowler, and Stern 1982).

Surface roughness at the Si/SiO_2 interface are expected to constitute a major cause of scattering at high electron concentrations. The surface roughness contributes a perturbation $\delta V(x,y,z)$ to the potential energy in the inversion layer, which becomes more important as the transversal electrical field increases, i.e. at higher V_G and consequently larger inversion layer charge concentrations. After the calculations of Cheng and Sullivan (Cheng and Sullivan 1973) scattering by surface roughness is only weakly dependent on temperature.

Coulomb scattering is originated by charged centers near or in the inversion layer, for example, ionized dopant impurities or oxide charges near or at the Si/SiO_2 interface. This scattering mechanism is known to lead to a scattering probability that increases as the kinetic energy of the carriers decrease (Ando, Fowler, and Stern 1982). As the carrier density in the inversion layer increases, however, the screening effect becomes important. The screening effect is reduced as the kinetic energy of the carriers increases. Therefore, the actual temperature dependence of Coulomb scattering is determined by the relative importance of these two effects.

As a result, the scattering mechanism that will dominate at a given temperature depends on the quality of the sample and on the amount of inversion (Lee, Choi, Sim, and Kim 1991). For undoped or lightly doped samples with negligible oxide charge the mobility increases monotonically as the temperature decreases, since Coulomb scattering is negligible and the lattice scattering decreases with temperature. However, for doped samples the mobility shows a peak at low temperatures and then decreases, showing that coulomb scattering at ionized impurities is the dominant scattering mechanism for doped samples at low temperatures. If the charge concentration in the oxide is high, coulomb scattering at the oxide charge near the Si/SiO_2 interface may become important. For high inversion layer concentrations, screening becomes effective. In this situation there are large transversal electrical fields, and if the Si/SiO_2 interface is of poor quality, surface roughness can become the dominating scattering mechanism (Takagi, Toriumi, Iwase, and Tango 1994).

Another important point is that at low temperatures, Boltzmann statistic does not hold anymore and Fermi-Dirac statistics must be used. Fermi-Dirac statistics leads to very complex analytic expressions and is prohibitive for practical numerical analysis. One simplification usually done, called the 0 Kelvin approximation, is to approximate the Fermi-Dirac distribution by a step function at the Fermi level. At 0 Kelvin the carriers are frozen out, leading to the Fermi level lying between the valence (conduction) band edge and the acceptor (donors) energy levels, and the acceptor (donor) atoms not ionizing. However, at high doping concentrations the acceptor (donor) energy levels come very close the valence (conduction) band and the energy needed to excite the doping atoms becomes vanishingly small. Hence, for degenerated semiconductors mobile carriers are available even at vanishingly small temperatures (Wu and Anderson 1974).

For an nMOSFET in the 0 Kelvin temperature region, if sufficient gate bias is applied, the energy of the conduction band edge can be reduced below that of the edge of the quasi-Fermi level of the degenerate source and/or drain. In this case, electrons are injected into the inversion layer. Furthermore, at low temperatures, the potential in the bulk ϕ_B is closer to the valence band edge than at room temperature. Therefore, the total band bending needed to create the inversion layer is improved. This process leads to an increasing threshold voltage V_T with decreasing temperature. On the other hand, the channel turns on more abruptly, decreasing the sub threshold slope (Hanamura, Aoki, Masuhara, Minato, Sakai, and Hayashida 1986).

Although these very important differences between the operation of MOSFETs at room and cryogenic temperatures do exist, Hu and Anderson (Wu and Anderson 1974) have found that the expression for the drain current has the same form for the two temperature ranges. This is true as long as the following conditions are fulfilled:

1. The impurity concentrations at the source and drain are very high and thus these are degenerated. This insures that electrons are available, i.e. not frozen out, even in the 0 K approach.

2. There are no potential barriers between the channel and the drain and source regions.

In contrast, if the gate does not overlap source and drain or if the gate oxide thickness is increased in the overlap regions, potential hills will occur in the channel. Although these barriers will be small to affect current-voltage characteristics at room temperature, at low temperature they can prevent current flow until a minimum drain voltage (V_{Td}) is applied. At voltages very near V_{Td}, the $I-V$ characteristics is expected to be quite non-linear, but monotonically increasing with increasing V_D, as shown in Fig. 2.3 (Gutierrez 1995). This is valid as long as the channel is long enough, the quantization of energy states in the x-direction along the channel can be neglected, and the capacitance of the channel is large enough so that the shift in channel surface potential by charging it with one extra electron is vanishingly small. If this does not hold, the picture can change dramatically, as discussed in section 5.1 (resonant tunneling) and section 5.5 (Coulomb Blockade).

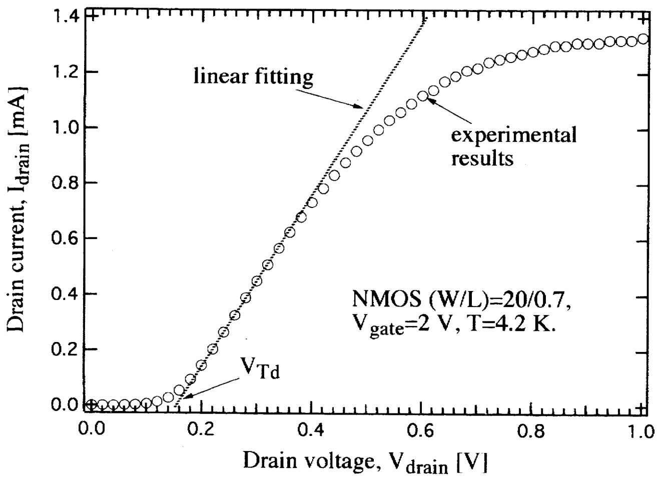


Figure 2.3: Measured I_D x V_D characteristics of nMOS LDD transistor with L=0.7 μm operated at 4.2 K showing the Drain Threshold Voltage V_{Td} due to lack of overlapping between gate and drain and/or source. After (Gutierrez, 1995).

2.3 Channel Carrier Quantization

As shown in Fig. 2.4, the bandbending of the semiconductor in the inversion layer is strong enough to produce a potential well. Its width in the direction perpendicular to the surface is small compared to the wavelengths of the carriers. Thus, the energy levels of the electrons are grouped into electric subbands, each of which corresponds to a quantized level for motion in the direction perpendicular to the surface, with a continuum for motion in the plane parallel to the surface.

A rigorous treatment of electron states in the inversion layer requires complex numerical and self-consistent solutions of the Poisson and Schrödinger equations.

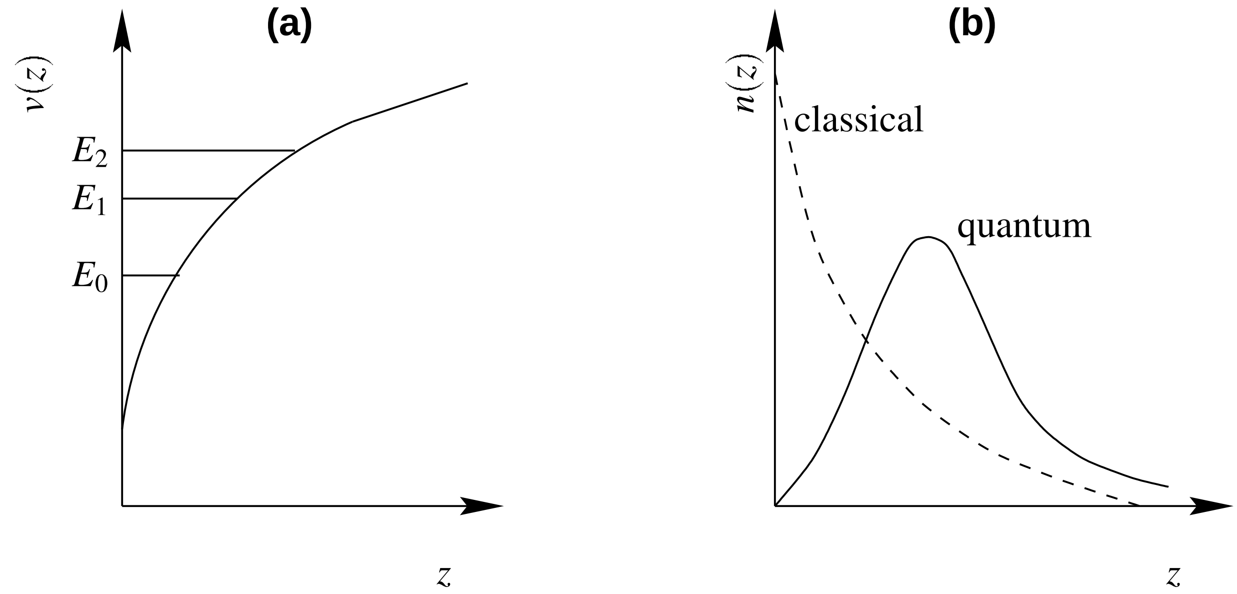


Figure 2.4: Quantization of the inversion charge. (a) shows the potential well and the quantized energy levels for a representative Si surface. (b) shows the corresponding classical and quantum-mechanical charge densities.

One of the simplest approximations, the effective-mass approximation, treats the electrons as though they had masses characteristic of a conduction-band minimum, neglecting nonparabolicity and coupling to other band extrema. Then one can use a kinetic energy operator of the form (Siggia and Kwok 1970)

$$\hat{T} = \frac{\hbar^2}{2} \sum_{i,j} w_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \tag{2.20}$$

where the $w_{i,j}$ are the elements of the reciprocal effective-mass tensor for the particular conduction band minimum being observed.

If no bias is applied between drain and source, the potential is a function of z only, and the wave function can be written as the product of a Bloch function, a z-dependent factor, and a plane-wave factor representing the free motion in the xy-plane. This leads to constant-energy parabolas above the level E_i which is the bottom of the $i-$th subband, where the electronic wave function for the $i-$th subband is of the form (Stern and Howard 1967)

$$\psi_i(x,y,z) = \zeta_i(z) e^{i\theta z} e^{ik_1 x + ik_2 y} u_\alpha \tag{2.21}$$

where k_1 and k_2 are measured relative to the band edge, θ depends on k_1 and k_2, u_α is the Bloch function for the bottom of the conduction band valley being considered, and $\zeta_i(z)$ is the solution of the Schrödinger equation

$$\frac{d^2\zeta_i}{dz^2} + \frac{2m_3}{\hbar^2}[E_i + e\phi_i(z)] = 0 \tag{2.22}$$

where m_3 is the effective mass in the direction perpendicular to the Si/SiO_2 interface (z-direction). As the electrons are bound in the inversion layer, it is required that $\zeta_i(\infty) = 0$. In addition, it is required that $\zeta_i(z)$ vanish at the surface, where $z = 0$. This is a good approximation for the Si/SiO_2 interface, for which the potential barrier for electrons is approximately 3 eV. More realistic boundary conditions which do not require that $\zeta_i(z)$ vanish at the surface will lead to energy levels with reduced energies and increased energy-level splittings, but the deviations are expected to be small. More recent calculations, which do not impose this restriction, do not show significant deviations in the results. See, for instance, (Janik and Majkusiak 1998); (Spinelli, Benvenutti, and Pacelli 1998) or (Jallepalli, Shih, Pinto, Maziar, and Jr. 1997).

Each eigenvalue E_i found from the solution of 2.21 is the bottom of a continuum of levels called a subband, with energy levels given by

$$E_i(\vec{k}) = E_i + \frac{\hbar^2 {k_1}^2}{2m_1} + \frac{\hbar^2 {k_2}^2}{2m_2} \tag{2.23}$$

where m_1 and m_2 are the principal effective masses for motion parallel to the surface, which can be obtained in a straightforward way from the bulk masses (Stern and Howard 1967). Values for these quantities and for the mass m_3 for motion perpendicular to the surface, used in 2.21, are given in Table 2.1 for the (100) surface orientation, the surface orientation of the devices used in this work.

Valley		Lower	Higher
Degeneracy	n_v	2	4
Normal mass	m_3	0.916	0.190
Longitudinal	m_1	0.190	0.190
masses	m_2	0.190	0.916

Table 2.1: Parameters for the (100)-Si-Surface.

Since in this approach the kinetic energy of the motion perpendicular to the surface is of the form $\hbar^2 k_z^2/2m_3$, the valleys which present the highest mass for motion perpendicular to the surface have the lowest kinetic energy and the lowest energy levels. The subbands arising from these valleys are labeled with indices 0,1,2, ... The second ladder of subbands, arising from the valleys with the lower mass for motion perpendicular to the surface, are labeled 0', 1', 2', ...

The potential $\phi(z)$ which appears in 2.22 is the solution of Poissons equation

$$\frac{d^2\phi(z)}{dz^2} = -\frac{1}{\varepsilon_{sc}}\left[\rho_{depl}(z) - e\sum_i N_i \zeta_i^2(z)\right] \tag{2.24}$$

where ε_{sc} is the permittivity of the semiconductor, N_i is the carrier concentration in the ith subband, given by

$$N_i(\vec{k}) = \frac{n_{vi} m_{di} k_B T}{\pi\hbar^2} F_0[\frac{E_F - E_i}{k_B T}] \tag{2.25}$$

$F_0(x) = ln(1+e^x)$, n_{vi} and m_{di} are the valley degeneracies and the density-of-states effective mass per valley, given in Table 2.1. ρ_{depl} is the charge density of the depletion layer.

The second term on the right hand side of 2.24 is the charge density in the inversion layer

$$\rho_{inv}(z) = -e\sum_i N_i \zeta_i^2(z) \tag{2.26}$$

where ζ_i is the normalized eigenfunction corresponding to the ith solution of 2.22. The function $g_0(z) = |\zeta_0(z)|^2$ is shown for a representative case in Fig. 2.4(b).

Because of the way in which the Schrödinger and Poisson equations are coupled, it is necessary in general to solve these equations numerically to obtain self-consistent results, as described below.

It is, however, possible to find approximate analytical results for some simple limiting cases. The simplest approximation is to replace the potential $\phi(z)$ in 2.22 by $-F_s z$ for $z > 0$ and by an infinite barrier for $z < 0$. F_s is the surface electrical field. This is the so called "triangular-potential approximation". It leads to Airy equation with the solutions (Stern 1972):

$$\zeta_i(z) = Ai\left[\left(\frac{2m_3 e\phi_s}{\hbar^2}\right)^{\frac{1}{3}}\left(z - \frac{E_i}{eF_s}\right)\right] \tag{2.27}$$

$$E_i \cong \left(\frac{\hbar^2}{2m_3}\right)^{\frac{1}{3}}\left(\frac{2}{3}\pi e\phi_s(i+\frac{3}{4})\right)^{\frac{2}{3}} \tag{2.28}$$

The exact eigenvalues for E_i have $i + \frac{3}{4}$ in 2.28 replaced by 0.7587, 1.7540 and 2.7525, respectively (Stern 1972).

The triangular-potential approximation may be a reasonable approximation when there is a little or no charge in the inversion layer (weak inversion), but fails when the charge density per unit area in the inversion layer is comparable to or exceeds that in the depletion layer (strong inversion) (Stern 1972).

When only one subband is occupied, that is, in the electric quantum limit, a variational approach gives a good estimate for the energy of the lowest subband. Fang and Howard (Fang and Howard 1966) used the trial eigenfunction

$$\zeta_0(z) = \left(\frac{1}{2}b^3\right)^{\frac{1}{2}} z e^{\frac{-bz}{2}} \tag{2.29}$$

with a single undetermined parameter b. The energy of the lowest state is found to be

$$E_0 = \frac{\hbar^2 b^2}{8m_3} + \frac{3e^2}{\varepsilon_{sc} b}\left[N_{depl} + \frac{11}{16}N_{inv} - \frac{2}{b}(N_A - N_D)\right] \tag{2.30}$$

Approximate energy levels for the excited states can be obtained in the electric quantum limit by treating the inversion-layer potential and the curvature of the depletion potential as perturbations. The inversion-layer charge density is assumed to be a sheet located a distance z_0 from the surface.

The approximated analytical results described above may be used as an initial estimate for the potential $\phi(z)$ to solve Eqs. 2.22 and 2.24 successively with a numerical algorithm until the output potential from 2.24 agrees with the input potential in 2.22 to within specified limits. This is the so called self-consistent method in the effective mass approach.

Because of the form of the kinetic energy operator in 2.20, the results obtained with the effective mass approach do not apply to inversion-layer energy levels arising from a warped (non-parabolic) bandstructure like the valence band of Silicon.

One method to overcome the limitations of the effective mass approach is to include non-parabolicity corrections for the dispersion relation in the self-consistent solution of 2.22 and 2.24 (Troger, Kosina, and Selberherr 1997). Following the effective mass approach and including a non-parabolicity correction term in the bulk dispersion relation

$$\varepsilon_i(1 + \alpha\varepsilon_i) = \frac{\hbar^2 {k_3}^2}{2m_3} + \frac{\hbar^2 {k_1}^2}{2m_1} + \frac{\hbar^2 {k_2}^2}{2m_2} \tag{2.31}$$

one has to solve the Schrödinger equation, which now, unlike 2.20, has a kinetic energy operator that does not allow to separate the energy term due to the in-plane transport. It will introduce a dependence on the in-plane wave vectors k_1, k_2 of both the eigenenergies and the wave functions (Troger, Kosina, and Selberherr 1997):

$$\hat{T} = \frac{1}{2\alpha}\left[-1 + \sqrt{1 + 2\alpha\hbar^2\left(\frac{{k_1}^2}{m_1} + \frac{{k_2}^2}{m_2} + \hat{G}\right)}\right] \tag{2.32}$$

$$\hat{G} = -\frac{\partial}{\partial z}\frac{1}{m_3}\frac{\partial}{\partial z} \tag{2.33}$$

This complicates the problem and increases the computational power needed to solve the equations.

The group of Jallepalli (Jallepalli, Shih, Pinto, Maziar, and Jr. 1997) has done full band calculations and did observe remarkably good agreement between the full band and the effective-mass calculations of electron subband dispersions and wave functions in n-MOS inversion layers. The only significant difference between the two was found to be a small doping-dependent shift between the subband edges.

One of the most important aspects of the inversion layer quantization in the scope of this work is the possibility of population of excited subbands. In the 0 K approximation, only the ground subband is occupied at moderate inversion layer charge densities (up to $3x10^{12}cm^{-2}$ for $N_A - N_D = 10^{15}cm^{-3}$) (Stern 1972). For higher inversion layer charge densities the first excited subband can become occupied. As the separation between the subbands increases with increasing inversion layer density (Moglestue 1986), the population of a second excited subband at low temperature is not expected. Another important feature is the pinning of the Fermi level to the bottom of the first excited subband after they cross. The pinning is a consequence of the increased density of states after the second subband is occupied (Ando, Fowler, and Stern 1982; Hänsch, Vogelsang, Kircher, and Orlowski 1989). As the temperature is increased, inversion-layer electrons will be excited to higher subbands, leading to an increasing inversion layer contribution to the potential well and a larger average z_{av} distance of the electrons from the surface.

At room temperature, an increasing space charge (increasing V_G), will lead to a decreasing z_{av}, because of the increasing average field seen by the electrons, which increases the band-bending and overpowers the increase that might have been expected when the higher subband is occupied.

If a large source drain bias V_{DS} is applied, electrons can obtain sufficient energy in the longitudinal electric field to be transferred to an excited subband, as observed by the group of Neugebauer (Neugebauer, Landwehr, and Hess 1978). These authors observed Negative Differential Resistance (NDR) in nMOSFET of (100) surface orientation and suggested that the origin are heated electrons transferred from the twofold to the fourfold degenerate subband system, which has a greater effective mass in the channel direction and hence a lower mobility.

Other important aspects of inversion layer quantization are its impact on the threshold voltage V_T, transconductance, and on the capacitance of the MOS system. The shift of the carrier concentration peak away from the interface results in an increase in the effective oxide thickness (Jallepalli, Shih, Pinto, Maziar, and Jr. 1997; Ip and Brews 1998). This increase is also manifested as a lower transconductance and a lower quasi-static capacitance in strong inversion. Quantization also predicts an increased V_T, when compared to the classical theory. The first reason is that, as can be seen from Fig. 2.4, the ground state for electrons, E_0, is formed above the bottom of the conduction band. The second reason for the phenomenon is the reduction in the density of states in the inversion layer. In the 3-D transport system, the density of states increases continuously from the bottom of the conduction band. In the 2-D transport system, on the other hand, it changes in a stepwise manner, and there are fewer states even above the ground state than those in the 3-D transport system (Hareland, Jallepalli, Chindalore, Shih, Jr., and Maziar 1997). So the Fermi level has to be closer to the ground state in order to induce the inverted charges.

It is important to note that the preceding discussion of the subband structure is within the

Hartree approximation, which does not account for many-body effects such as exchange and correlation. This approximation is valid when the average kinetic energy of electrons is much larger than the average interaction energy, that is, when the electron concentration is sufficiently high (Ando, Fowler, and Stern 1982).

2.4 Process Technology

The devices used in this research work are processed using an unique approach, developed along the years at the university of Dortmund. This method applies only conventional optical lithography and enables the fabrication of structures with a minimum line-width down to 30 nm without the requirement of expensive and time consuming techniques like e-beam writing or molecular beam epitaxy (MBE). Below an overview of the technique and the parameters of relevance to this work are given. A more detailed description of the process can be found in (Horstmann, Hilleringmann, and Goser 1996) or (Horstmann, Hilleringmann, and Goser 1997).

The process is based on a sidewall-etchback technique similar to the one used for the definition of the spacers in conventional MOS-Transistors, and runs on a conventional 0.8 μm-CMOS-technology line.

The first step consists in depositing a polysilicon layer on the top of the thermally grown thin gate-oxide-film. After that, a sacrificial oxide layer is deposited and structured by dry etching, applying conventional optical lithography as shown in Fig. 2.5(a). The conformal deposition of a nitride layer follows, as depicted in Fig. 2.5(b). This nitride layer is etched back anisotropically, leaving a nitride spacer surrounding the oxide (step (c) in Fig. 2.5). This nitride spacer is used as a masking material for the definition of the transistor gates. Since the thickness of the nitride layer determines the width of the nitride spacer, the gate length is given by the nitride layer thickness. After the selective removal of the remaining sacrificial oxide layer, the nitride mask is transferred into the polysilicon layer, as depicted in part (d) of Fig. 2.5. This step is performed using anisotropic dry etching in the Reactive Ion Etching (RIE)-mode with chlorine chemistry.

For the Lightly Doped Drain regions, antimony is implanted, as shown in Fig. 2.5(e). Antimony is the material of choice because of its high mass, which allows to a low implantation range and diminished dopant diffusion in the subsequent thermal annealing steps. As a 7 degree tilt angle implantation method is used, the wafers are rotated four times during the implantation process to compensate the shadowing effect of the gate electrode. Finally, the oxide spacers are defined and the source/drain arsenic implantation is performed (step (f) in Fig. 2.5). Arsenic is chosen because of its high solubility in silicon.

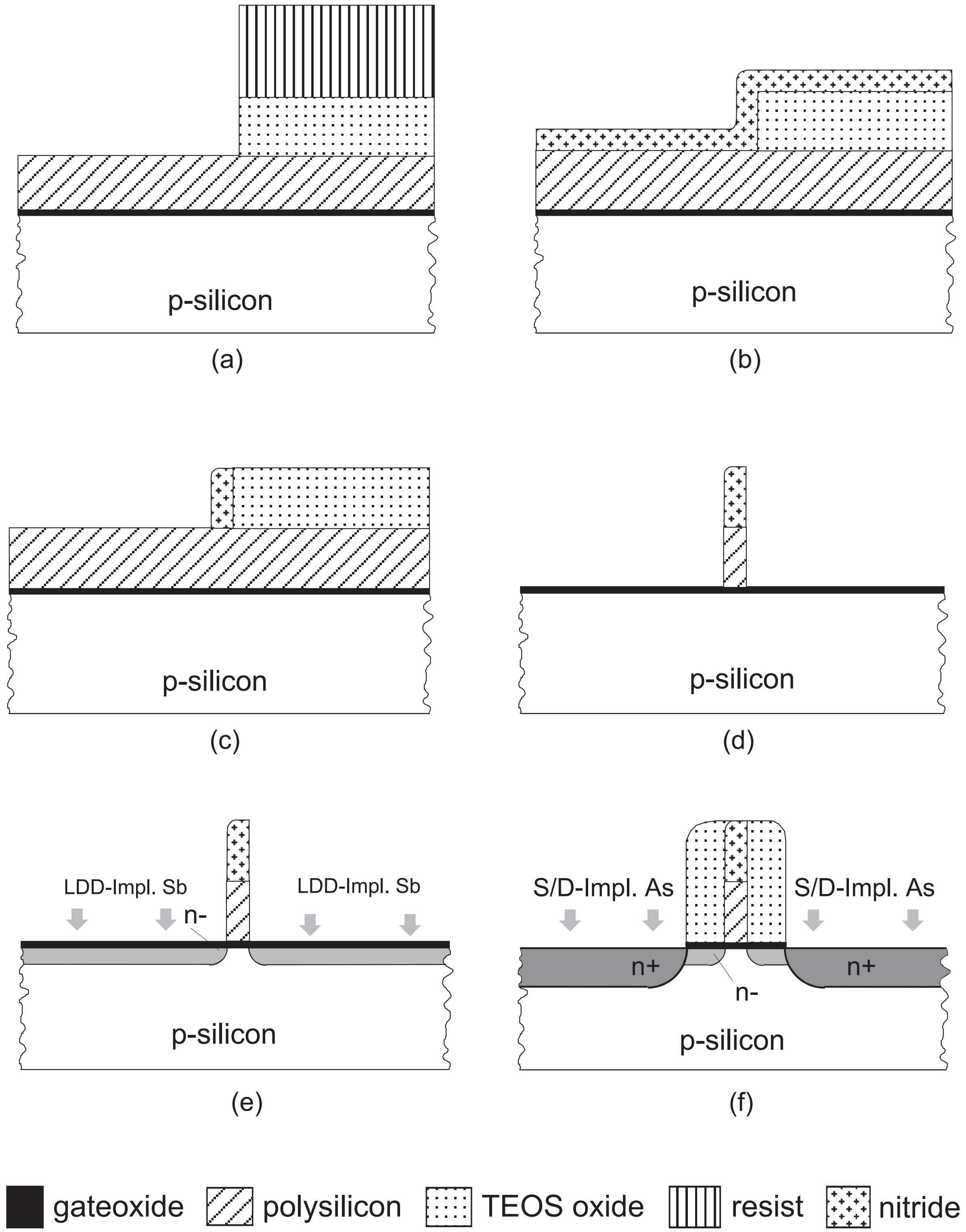


Figure 2.5: The processing of sub-100nm-devices at the Department for Electrical Engineering, University of Dortmund. For a detailed explanation of the individual steps, please refer to the text.

3. Samples and Experimental Setup

In this chapter the layouts of the samples investigated are presented and the sample preparation is described. The main technological process parameters are given and the experimental setup used for carrying out the measurements is introduced.

Measurements have been carried out at temperatures between 300 mK and 35 K, and at room temperature. Low temperature measurements were carried out using a ^4He cryostat with a single ^3He insert from Oxford Instruments, or using simple, self-made ^4He bath cryostats. The measurements in high magnetic fields were carried out using the $^3He/^4He$ cryostat together with a superconducting coil magnet. This magnet can provide magnetic fields up to $B = 17$ T.

3.1 Sample Layout and Parameters

The layout of the transistors studied is shown in figures 3.1, 3.2 and 3.3. The first test structure (Fig. 3.1) is composed by 8 transistors with 4 different channel widths. For the mask set CTC 17 the channel widths are 1.0 μm, 2.5 μm, 6.0 μm and 25 μm. For the mask set CTC 16 one has channel widths of 4.0 μm, 5.5 μm, 12.5 μm and 25 μm. The sources of all transistors are connected to a common source pad. Similarly, the gates are all connected to a common pad. Each transistor is individually operated using separate drain pads. The second test structure (Fig. 3.2) is composed by 4 transistors, each with a different channel width: 1.2 μm, 1.0 μm, 0.8 μm and 0.6 μm. In this arrangement every transistor has one separate pad for driving drain and gate. The source pads are shared between two transistors: one by the 1.2 μm and 1.0 μm wide transistors and the other by the 0.8 μm and 0.6 μm wide transistors. The last test structure (Fig. 3.3) constitutes of 2 transistor with the same channel width of 100μm. If a CMOS process is used, one of them will be a n-type-transistor and the other a p-type-transistor. These transistors share a common gate pad, while other contacts are connected to separate pads.

As described in section 2.4, the channel length is not defined by the layout. The thickness of a nitride layer deposited in the particular process technology used defines the channel length. For the wafers used to obtain the experimental results reported in this work, the technological parameters of relevance are summarized in tables 3.1 and 3.2. The gate oxide thickness (t_{ox}) is measured by an ellipsometer. L_{design} is the design channel length, i. e., the width of the nitride spacer, measured using a Scanning Electron Microscope (SEM). The effective channel length (L_{eff}), the Sb-dopant concentration in the LDD region (LDD Conc., in cm^{-3}) and its depth (LDD Depth), the B-dopant concentration in the channel at the Si/SiO_2 interface (Surf. Conc., in cm^{-3}) as well as the depth of the source and drain regions (D/S Depth) are values estimated by means of process simulation with the two-dimensional process simulator DIOS (ISE - Integrated

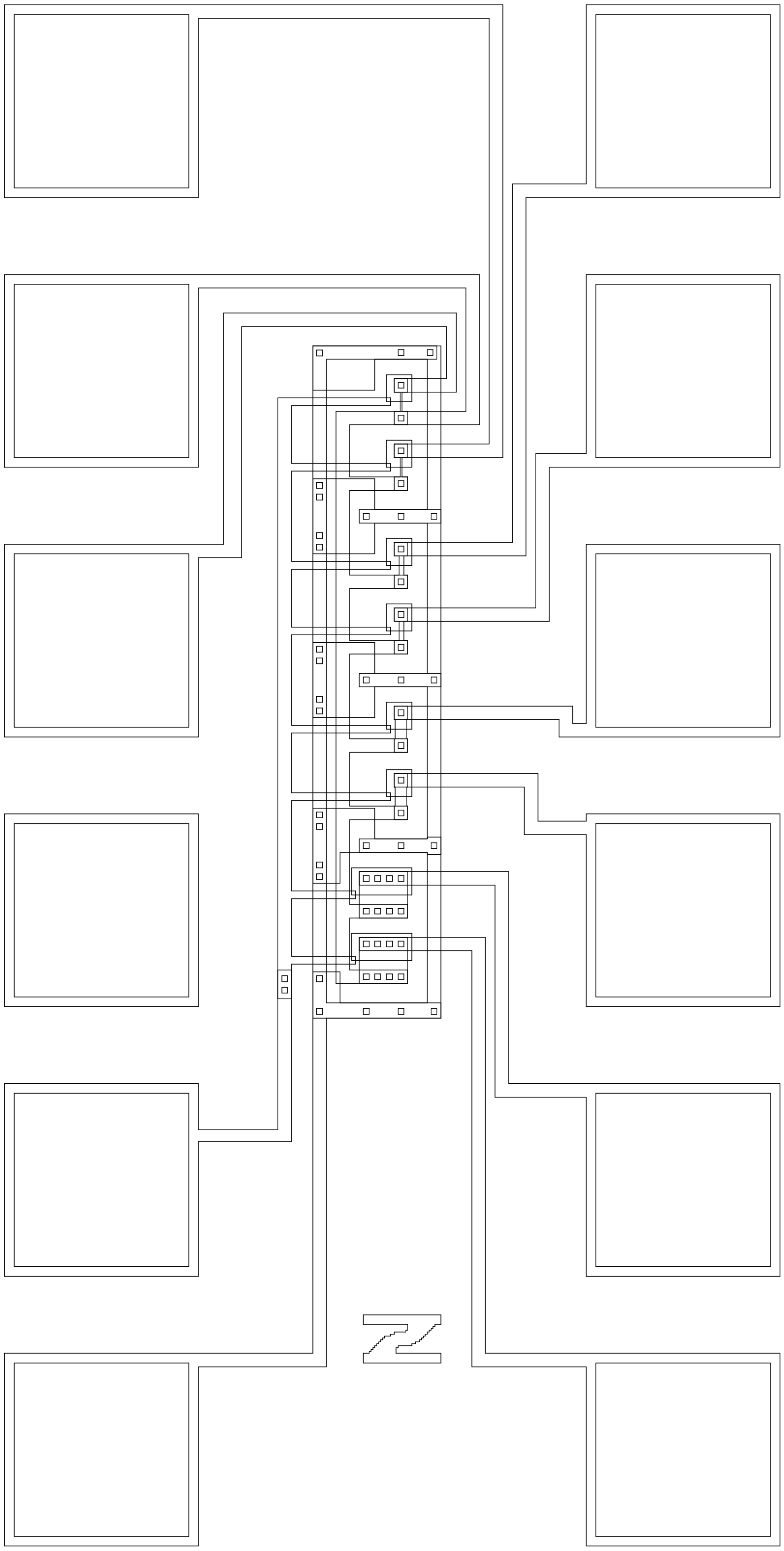


Figure 3.1: Layout of the test structure composed by four transistor pairs with different channel widths. The gates are driven by a common gate pad, and the sources contacted through a common source pad. The transistors can be individually operated over the isolated drain pads.

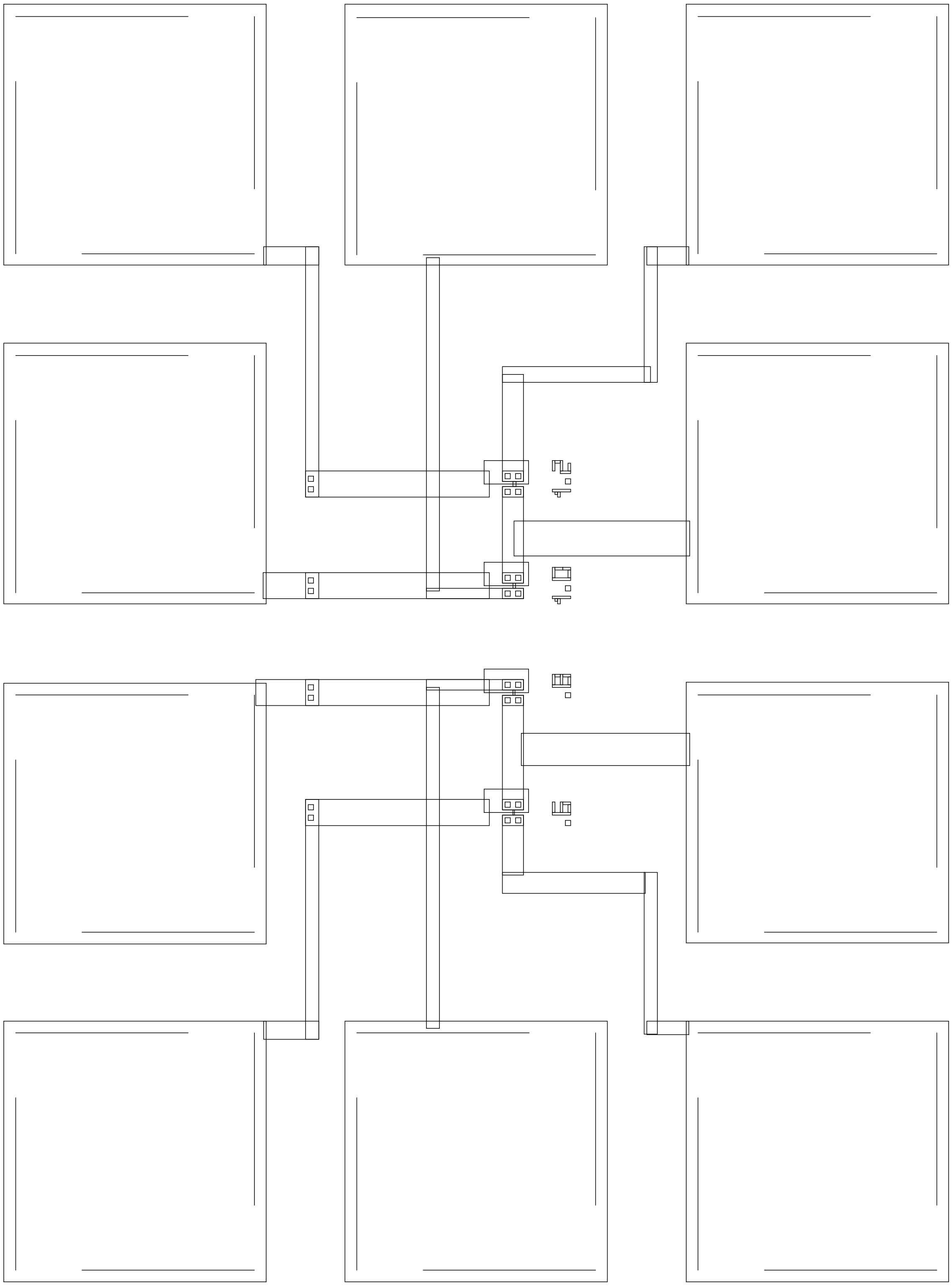


Figure 3.2: Layout of the test structure composed by four transistor with different channel widths. Channel widths are 1.2 μm, 1.0 μm, 0.8 μm and 0.6 μm. There are isolated pads for driving the individual gate and drain contacts. All transistors gates are driven by a common gate pad. The source pads are shared between two transistors.

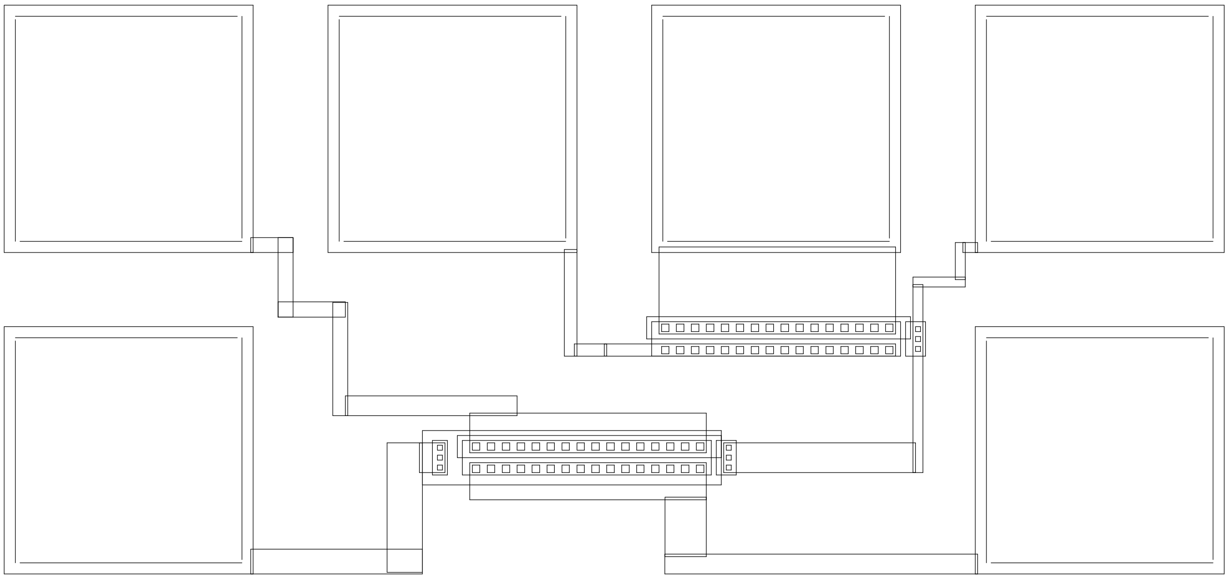


Figure 3.3: Layout of the test structure composed by two transistors with the same channel width W = 100 μm. Transistors gates are driven by a common gate pad. All other contacts are connected to isolated pads.

Systems Engineering AG 1994), for room temperature.

After processing, the individual test structures are split from the wafer using a conventional diamond dicing saw. Next, the dies are cleaned with acetone and submited to a 15 min. furnace annealing at 150 C. The dies are then attached to gold plated ceramic dual in line packages (DIP) using a silver epoxy adhesive. The chip-pads are bonded to the package pins using an ultra-sonic bonder with aluminum wiring. Good bulk contact is ensured by attaching the dies to the gold package plating using a silver basis adhesive. This contact could be verified by measuring the diodes formed by the source-bulk and drain-bulk junctions (at room and at low temperature).

After bonding, samples are only handled with the user connected to ground through an anti-static-bracelet, to assure the integrity of tests structures, specially of the very thin gate oxide.

Devices with design channel lengths L_{design} between 30 nm and 120 nm and channel widths W between 0.6 μm and 100 μm, processed on (100) silicon as described in the previous chapter, are studied at low ($0.3K < T < 35K$) and room temperature. The samples have been processed in different runs, with gate oxide thickness (t_{ox}) ranging from 4.0 nm to 8.7 nm. The depths of the lightly doped drains (LDD) are between 10 nm and 13 nm. For the t_{ox}, L and LDD depth values of the particular samples, please refer to tables 3.1 and 3.2. For the sake of comparison some devices without LDD have also been processed. For these devices the design channel length was $L_{design} = 120nm$, and the effective channel length is estimated to be $L_{eff} \approx 50nm$, by numerical process simulations performed using DIOS (ISE - Integrated Systems Engineering AG 1994).

Devices with micrometer scale channel lengths, processed on the same wafers as the sub-100 nm devices, have also been measured.

A scanning electron microscope (SEM) photography of a device cross section is shown in

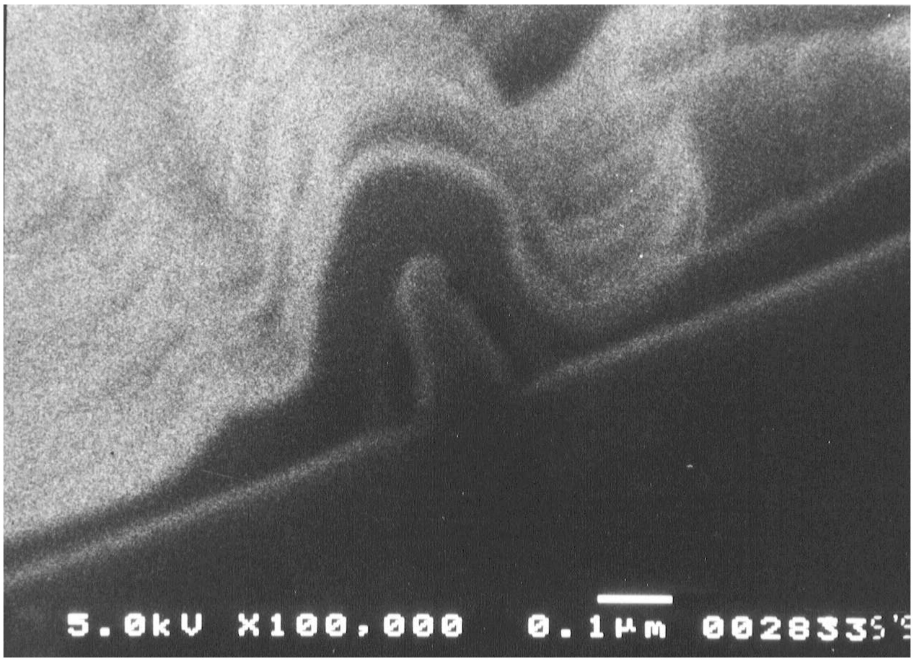


Figure 3.4: Scanning electron micrography showing a device cross section. The polysilicon gate electrode is $\approx 50nm$ in length and $\approx 220nm$ in height. The top layer is SiO_2.

Fig. 3.4. The 50 nm long gate electrode with the passivation oxide on the top can be clearly seen.

3.2 Experimental Setup for Measurements at Low Temperature

For the charge transport measurements at liquid Helium temperatures three different setups were employed. The theory necessary to understand the behavior of MOSFET at cryogenic temperatures was given in section 2.2.

In the first measurement setup, present at the *Lehrstuhl für Werkstoffe der Elektrotechnik, Fakultät für Elektrotechnik, Ruhr-Universität Bochum*, the samples were loaded into a ^4He cryostat with a single ^3He insert from Oxford Instruments, model *Spectromag with Heliox(^3He)-insert*. This cryostat allows top loading into a ^3He or ^4He vessel.

If the sample is immersed in the ^4He bath held at atmospheric pressure, the sample will be in thermal equilibrium with the bath after few minutes (at the ^4He boiling temperature, 4.2 K). The samples are slowly loaded to avoid a thermal shock, and a wait period of at least 15 min. after the sample was immersed into the bath was required, before proceeding with the electrical measurements.

To allow experiments at lower temperatures, the cryostat is equiped with a sorption pump (sorb) for the condensation of the liquid ^3He, making it possible to reach a base temperature

Wafer	366/3	366/13	377/3	397/2	397/6
L_{design}	53 nm	54 nm	72 nm	53 nm	32 nm
L_{eff}	48 nm	49 nm	67 nm	49 nm	28 nm
Spacer Width	110 nm	110 nm	110 nm	100 nm	100 nm
t_{ox}	8.7 nm	7.0 nm	4.4 nm	4.9 nm	4.7 nm
Surf. Conc.	1.0x10^{18}	8.0x10^{17}	8.5x10^{17}	7.7x10^{17}	7.7x17^{17}
LDD Conc.	1.5x10^{18}	1.2x10^{18}	1.5x10^{18}	1.2x10^{18}	1.2x10^{18}
LDD Depth	10 nm	11 nm	13 nm	10 nm	10 nm
D/S Depth	120 nm	120 nm	120 nm	120 nm	120 nm

Table 3.1: Technology Parameters for the NMOS Wafers.

Wafer	413/6	430/1	430/3	430/5	430/7
L_{design}	120 nm	82 nm	80 nm	49 nm	48 nm
L_{eff}	50 nm	60 nm	60 nm	45 nm	45 nm
Spacer Width	n.a.	110 nm	110 nm	110 nm	110 nm
t_{ox}	4.7 nm	3.8 nm	8.3 nm	4.0 nm	8.1 nm
Surf. Conc.	1.1x10^{18}	1.0x10^{16}	1.0x10^{16}	8.2x10^{17}	8.0x10^{17}
LDD Conc.	n.a.	9.0x10^{17}	1.6x10^{18}	1.4x10^{18}	1.6x10^{18}
LDD Depth	n.a.	12 nm	10 nm	12 nm	10 nm
D/S Depth	90 nm	120 nm	120 nm	120 nm	120 nm

Table 3.2: Technology Parameters for the NMOS Wafers (cont.).

lower than 0.3 K. The temperature of the bath is then held constant by a manifold pumping and needle valves system, that controls the vessel pressure. With this system the temperature can be held constant for a few hours, depending on the power dissipated by the sample. A typical value of the power dissipated by the sample during measurement was 1 nW ($100nA$ x $10mV$). The cooling power of the system is much greater (120 μW, minimum).

During the measurements the temperature was continuously monitored by a sensor in contact with the sample socket. The temperature sensed with an accuracy of at least 0.01 K if no magnetic field is applied, and of at least 0.1 K if high magnetic fields was present. As well as offering mechanical valve control, the cryostat also incorporated a still heater control and automated temperature monitoring, making it possible to carry out measurements at a wide temperature range.

The cryostat is equiped with a Nb_3Sn DC superconducting magnet that allows the application of magnetic fields up to 17 T, with a homogeneity better than 0.1% over the whole sample area. The manufacturer guarantees the persistence of the magnetic field to be better than $0.01\%/hour$. The magnetic field strength was always monitored during measurements. The magnet operating current, which brings up the magnetic field, is remotely controlled through a microcomputer connected to the current control unit via the GPIB interface. The accuracy of the applied field is better than 0.001 T.

In order to avoid the damaging of the very sensitive devices, all sample pins are connected to ground during the loading into the cryostat and cooling down procedure.

The equipment for electrical measurements is outside the cryostat, at room temperature, and the cryostat is connected to the same electrical ground as the measurement equipment. Measurements were carried out using standard lock-in techniques, with a dual phase lock-in amplifier supplied by Stanford Research Systems, model SR830 (Stanford Research Systems 1996). This lock-in amplifier has a sensitivity of 2 nV to 1 V, a gain accuracy of $\pm 1\%$, an absolute phase error of $< 1°$ and a relative phase error $< 0.001°$.

The second measurement setup, where most measurements were carried out, is composed by a HP4156A precision semiconductor analyzer (Yokogawa-Hewlet-Packard. Ltd. 1995) and a self-built ^4He bath cryostat. The measurements were carried out taking advantage of the low-temperature gases facility of the *Lehrstuhl Experimentelle Physik II, Institut für Physik, Universität Dortmund*.

The cryostat consists basically of an 100 liter can filled with ^4He and a sample-carrier for top loading. The sample-carrier consists of a steel tube with a 16-pin socket at its bottom end, where the sample, mounted and bonded on a ceramic DIP package, can be attached. The tube is made out of high quality steel with low thermal conductivity. There are two isolated copper wires to each socket pin. BNC connectors are provided at the top-end. A schematic representation of the experimental setup is shown in 3.5. To ensure temperature stability, the sample is immersed into the liquid ^4He bath and the system is held at atmospheric pressure during the whole measurement. Hence, the sample is kept at the stable ^4He boiling temperature (4.2 K). For economic reasons, the evaporated ^4He is recovered and liquefied again. Also in this measurement setup, the sample is slowly loaded and sample pins are kept grounded during the whole cool down procedure.

The sample is then electrically connected to the HP4156A using triaxial cables. There are independent connections for applying the electrical signals (force) and to perform the measurements (sense), that is, Kelvin connections are employed. Cryostat and measurement equipment are connected to the same electrical ground.

The HP4156A offers a measurement resolution of $2\mu V$ and accuracy of $\pm(0.01\% + 200\mu V)$ for the voltage range up to $2.2V$. For the range $2.2V\ < V <\ 22V$ measurement resolution is $20\mu V$ and accuracy $\pm(0.01\% + 1mV)$. Voltage set accuracy is $\pm(0.02\% + 400\mu V)$ for the $0V\ < V <\ 2.2V$ range, and $\pm(0.02\% + 3mV)$ for the $2.2V\ < V <\ 22V$ range. The minimum output resolution is $100\mu V$. Current measurement accuracy is $\pm(0.5\% + 0.4pA)$ for the 1 nA Range. For the 10 nA range, measurement accuracy is $\pm(0.5\% + 2pA)$. The 100 nA range has an accuracy of $\pm(0.1\% + 20pA)$, and finally, the last range of relevance for this work, the $1\mu A$ range, has an accuracy off $\pm(0.1\% + 200pA)$. The High Resolution Source/Monitor Units (HRSMUs) were always used.

In both experimental setups described, the smallest drain voltage that could be applied during measurement was $0.1mV$. This value corresponds to the thermal voltage at a temperature $T \approx 1.1K$. Hence, the electron gas is slightly heated at the lowest temperatures. It should be again noted that, as described in chapter 4, the pattern does not change neither with drain voltage nor with temperature, and this is not expected to have dramatic consequences.

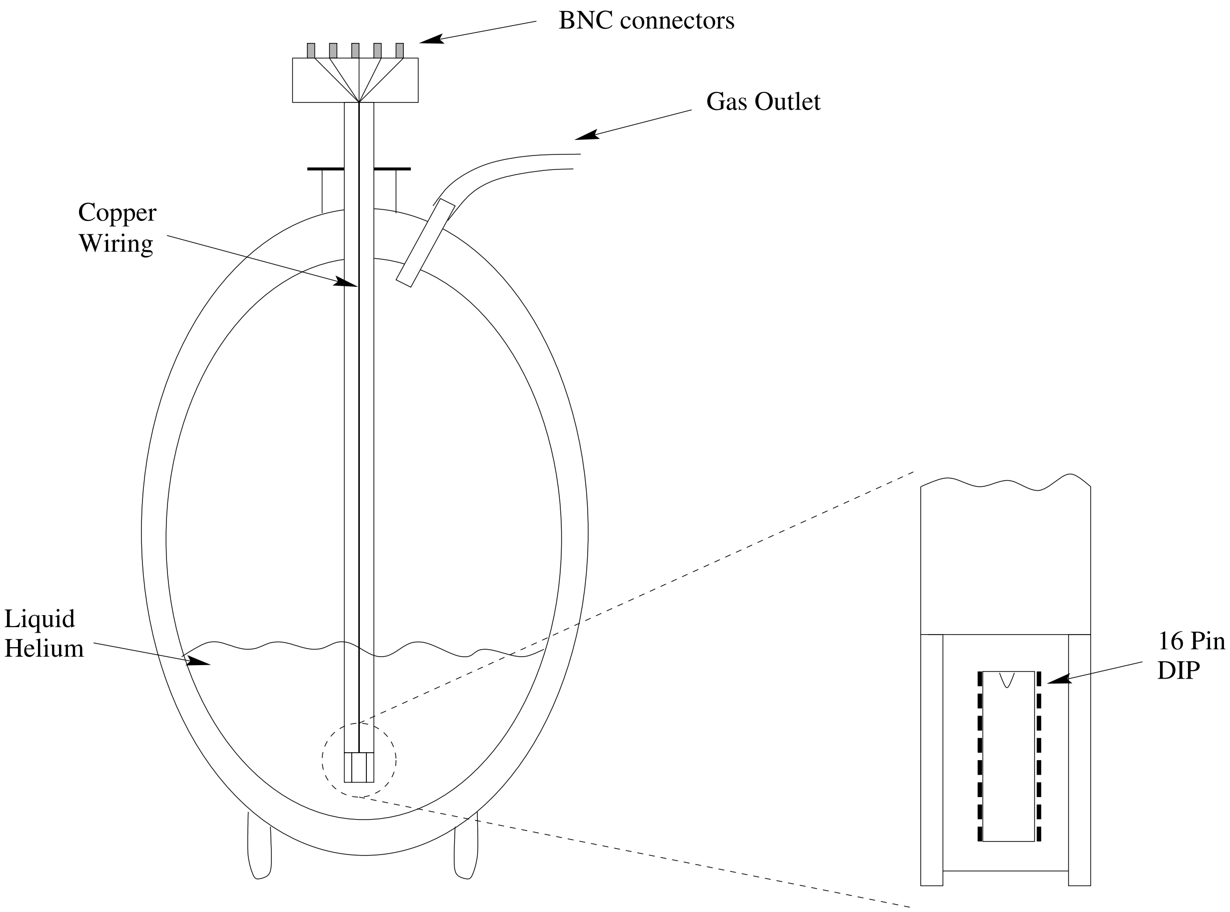


Figure 3.5: Schematic representation of the experimental setup: the sample is immersed into liquid Helium. The gas outlet is used for recovering the evaporated Helium, and the system is kept at atmospheric pressure. The sample is attached to a 16-Pin DIP socket. Each socket pin is connected to two BNC connectors using isolated copper wires, to allow the implementation of Kelvin connections up to the measurement apparatus.

A third measurement setup, very similar to the one just described, was used in some measurements carried out at the *Fakultät für Elektrotechnik, Institut für Physik, Universität der Bundeswehr (München)*. This setup is also composed by a HP4156A precision semiconductor analyzer and a self-built ^4He bath cryostat. The main difference is that this setup does not implement Kelvin connections and a temperature sensor placed on the back side of the sample socked is provided. Measurements at different temperatures can then be carried out if the sample is not immersed into the liquid Helium bath, but held at some distance from the bath surface, and the sample temperature is monitored.

Early measurements were carried out using a fourth measurement setup, composed by an HP4145B (Yokogawa-Hewlet-Packard. Ltd. 1986) and a closed cycle Helium refrigerator. The refrigerator is constituted of a Helium compressor, a vacuum pump system and a test fixture where the samples, attached to gold plated ceramic dual in line packages, are mounted. This systems allows experiments to be carried out at temperatures down to 20 K. This experimental setup was used in the early phase of this research work, and no results obtained using this setup are published here. Nevertheless, the characteristics measured using this setup are in conformity with the data published in the following chapters.

3.3 Experimental Setup for Measurements at Room Temperature

The measurements at room temperature were carried out using the HP4156A precision semiconductor analyzer in connection with a HP 16058A test fixture. This test fixture is equipped with an electrostatic light-shielding cover. Triaxial cables were used to connect the HP 16058A to a HRSMU (High Resolution Source/Monitor Unit) channel of the HP4156A precision semiconductor analyzer. As described above, all samples have been attached to ceramic dual in line packages (DIP) with gold plating and bonded using an ultra-sonic bonder with aluminum wiring. The minimum sampling time (minimum time interval between two successive measurement points) of the HP4156A is $60\mu sec.$ and up to 10000 points can be acquired on a single sampling measurement (Yokogawa-Hewlet-Packard. Ltd. 1995).

This measurement setup was mainly used for the study of the random telegraph signal (RTS), discussed in chapter 6, and for sample characterization prior to cooling down.

4. Periodic Oscillations: Experimental Results

In this chapter the experimental investigation done to characterize the behavior of the sub-100nm MOSFETs at low temperature is described. In the measurements carried out at temperatures between 300 mK and 35 K unexpected periodic oscillations with negative differential resistance (NDR) are present in the I_D x V_G characteristics. At room temperature the periodic oscillations can not be observed. Temperature and magnetic field dependence of sample characteristics are studied. Using the experimental setup described in chapter 3, magnetic fields up to $B = 17$ T could be applied.

Although similar behavior was already observed in silicon-on-insulator (SOI) devices (Peters, den Hartog, Dijkhuis, and Molenkamp 1998; Colinge, Baie, Bayot, and Grivei 1996), in MBE-grown vertical MOSFETs (Hansch, Rao, Fink, Kaesen, and Eisele 1998) and in GaAs FETs (Poole, Pepper, and Myron 1983), this is the first time that such phenomena is reported in a planar bulk MOS technology device.

4.1 I_D x V_G Characteristics

The typical I_D x V_G (input) characteristics of sub-100nm MOSFETs at low temperature are depicted in Fig. 4.1, 4.2 and 4.3. As the gate leakage current is always below 10 pA, tunneling through the gate oxide can be neglected (Kunze, Drebinger, Klehn, and Lindolf 1994). The measured input characteristic shows unexpected periodic oscillations with negative differential resistance (NDR), similar to those found by the group of Eisele (Eisele, Baumgärtner, and Hansch 1995; Eisele, Baumgrtner, and Hansch 1995; Hansch, Rao, Fink, Kaesen, and Eisele 1998) in vertical sub-100 nm MOSFETs grown by silicon molecular beam epitaxy. Although oscillations are always present, for devices with LDD as well as for devices without LDD, not all samples clearly show the periodicity observed in Fig. 4.1 to 4.3. This may be due to superposition of different physical phenomena involved in the electrical transport. Mesoscopic phenomena expected to play a role in the electrical transport of MOSFETs at cryogenic temperatures are discussed in chapter 5. Figure 4.5, for instance, depicts the characteristics of a sample for which transconductance fluctuations that are not periodic in V_G are present. In the higher V_G region (around 2 V) the fluctuations caused by the trapping and detrapping of electrons in localized defect states are also observed. This phenomenon is discussed in further detail in chapter 6.

Devices with effective channel lengths L_{eff} between $\approx 28nm$ and $\approx 120nm$ and channel widths W between $0.6\mu m$ and $100\mu m$, processed on (100) silicon as described in chapter 3 are

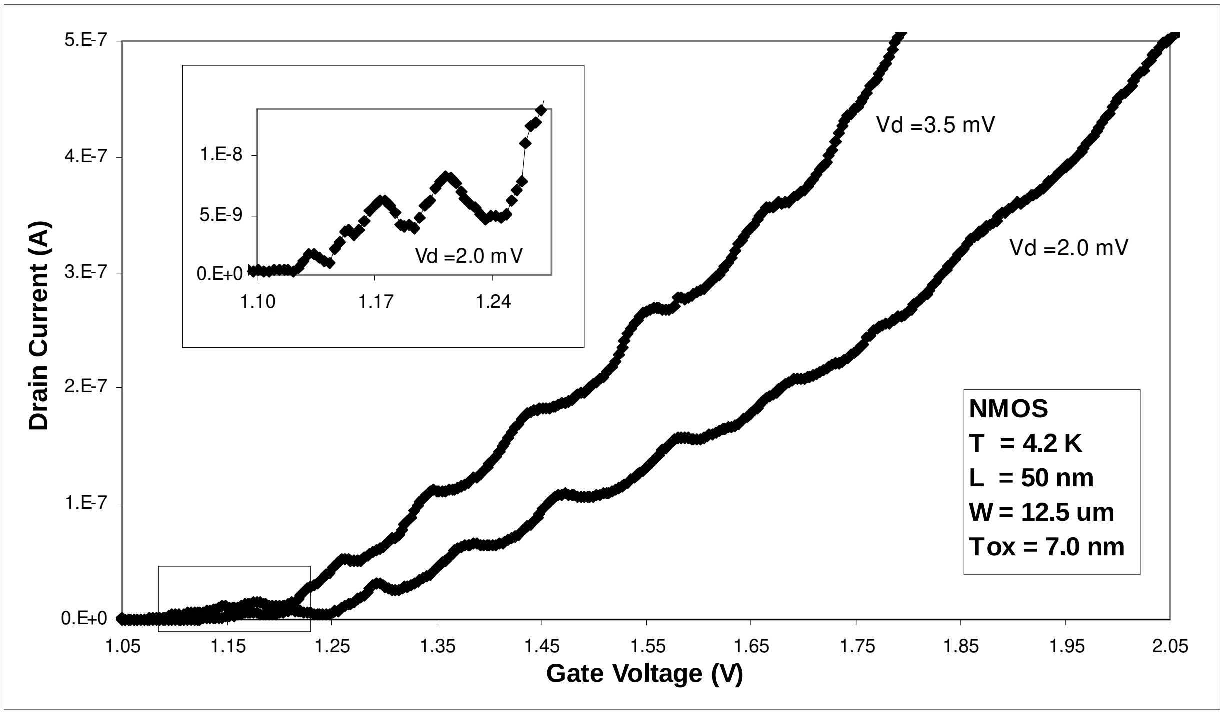


Figure 4.1: Input characteristics of a nMOS transistor showing periodic oscillations. The inset depicts the subthreshold region in a more appropriate scale.

characterized. In the experimental data presented in this work, the effective channel length, estimated by means of process simulation with the two-dimensional process simulator DIOS (ISE - Integrated Systems Engineering AG 1994), is always given. For details on sample preparation and technology parameters, please refer to section 3.1.

The oscillations are fairly reproducible from sample to sample and present from sub-threshold up to strong inversion. These characteristics, namely the presence of oscillation in strong inversion, is worth being emphasized. As discussed in chapter 2, at strong inversion screening of impurity and defect potentials is expected to be effective. No apparent relationship between the period of the oscillations and t_{ox}, W, L or channel doping concentration could be observed. For many samples the oscillation period in the region near the threshold voltage (V_T) is about 35 mV in gate voltage. These are followed by oscillations with a period of nearly 100 mV in the higher V_G region. The 35 mV period oscillations are better resolved for the $L = 30nm$ samples, as can be seen in Fig. 4.2. For these samples, the oscillations with a period of 100 mV in V_G are hard to be recognized at $T > 4.2K$. Samples with the shortest gate lengths show the strongest periodic oscillations, the same as observed by Poole et al. (Poole, Pepper, and Myron 1983) in GaAs FETs. From the 34 samples that did show periodic oscillations (Table 4.1) 22 have $W \leq 2.5\mu m$ and only 4 have $W > 10\mu m$. Up to now we were not able to find periodic transconductance fluctuations in devices with $L_{eff} > 100nm$. Transistor with micrometer channel lengths are also processed on the same wafer with the sub-100 nm MOSFETs. Eleven of these devices were characterized at low temperature. In contrast to the sub-100 nm MOSFETs, the input characteristics of these

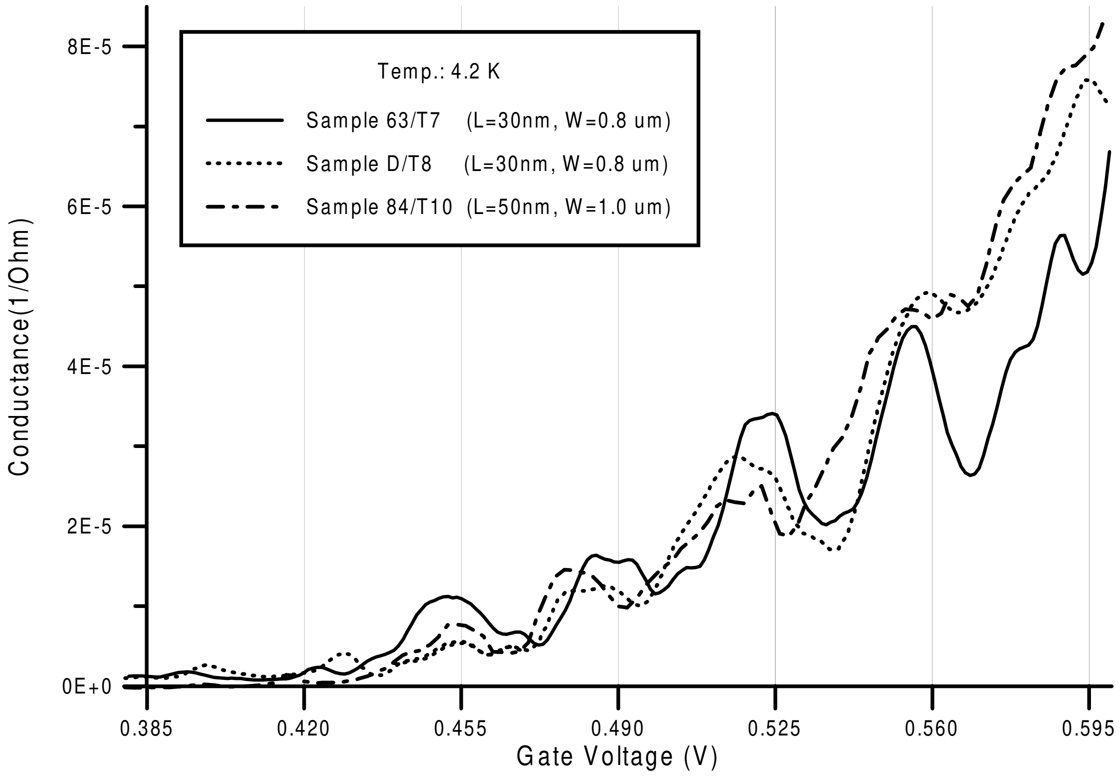


Figure 4.2: Characteristics of three different samples showing oscillation near threshold. Curves are shifted in V_G for better comparison.

micrometer channel length devices do not show any marked structure. The I_D x V_G curve is much smoother, and a periodic component is not present at all.

No dependence of the oscillation period on temperature (Fig. 4.8), drain voltage or magnetic fields up to 17 T could be observed (discussed in section 4.4).

The periodic pattern smears out with rising temperature or increasing drain voltage, but can be found at temperatures up to 35 K and drain source electrical fields up to $2.4kV/cm$, corresponding to a drain bias of $12mV$ for the $L = 50nm$ devices. On rising temperature, both peak and valley conductances do increase, as shown in Fig. 4.8. The valley current increases faster than the peak current, until the oscillations are smeared out. Different phenomena with distinct temperature dependency are present in channel current transport. Diverse parameters, as for instance the threshold voltage V_T, do have a strong temperature dependency. Therefore, it is difficult to isolate the current changes due to the temperature dependency of each phenomenon.

The oscillations are reproducible with temperature cycling. In fact, the conductance characteristics were studied over several months and remained essentially the same, even for samples that are cooled down to 4.2 K, brought back to room temperature and measured again at low temperature in between time intervals as long as half a year. Nevertheless, the first transconductance peak always occurs in the subthreshold region of the transistor, as shown in Fig. 4.4.

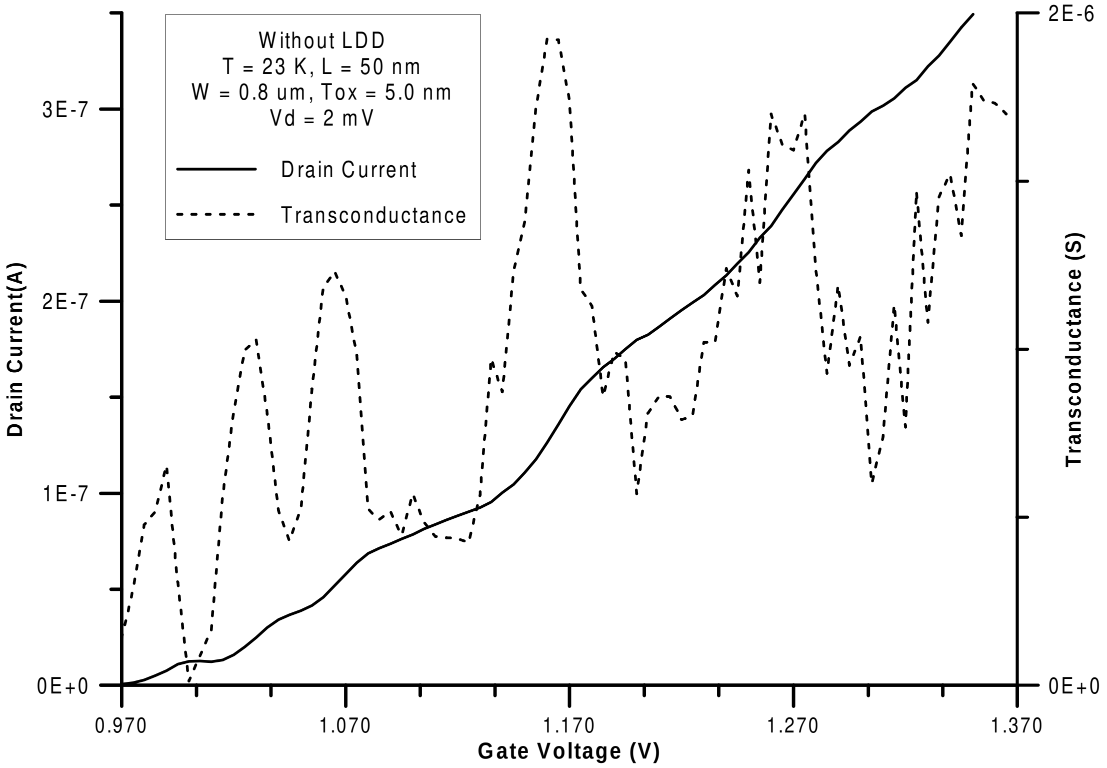


Figure 4.3: Drain current I_D and transconductance g_m dependencies on gate voltage V_G in a device without LDD.

In order to make a quantitative analysis of the periodic components (power spectrum) present in the experimental data discrete Fourier transform was applied. The Fourier transforms were evaluated using the software package called *Mathematica* (Wolfram 1997; Wolfram 1996). In *Mathematica*, the discrete Fourier transform b_s of a data series a_r of length n is taken to be $\frac{1}{\sqrt{n}}\sum_{r=1}^{n} a_r e^{2\pi i(r-1)(s-1)/n}$.

The goal of this Fourier analysis was to bring about quantitative statements regarding the fraction of samples that show periodic oscillations. This can be done looking to the amplitude peaks found at the frequency components in the discrete Fourier transforms of the measured I_D x V_G characteristics.

To perform Fourier analysis in an efficient way, the following methodology was automated using *Mathematica*:

1. Curve fitting is performed: a polynomial fit of the form $I_D = k_3V_G^3 + k_2V_G^2 + k_1V_G + k_0$ is found for the measured I_D x V_G curve. A least-squares fit with a mean value of zero over the V_G interval being considered is sought.

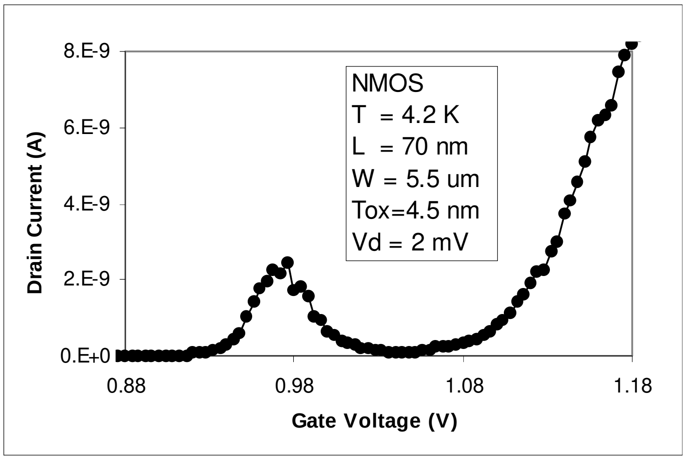


Figure 4.4: The first transconductance peak occurs in the subthreshold characteristics.

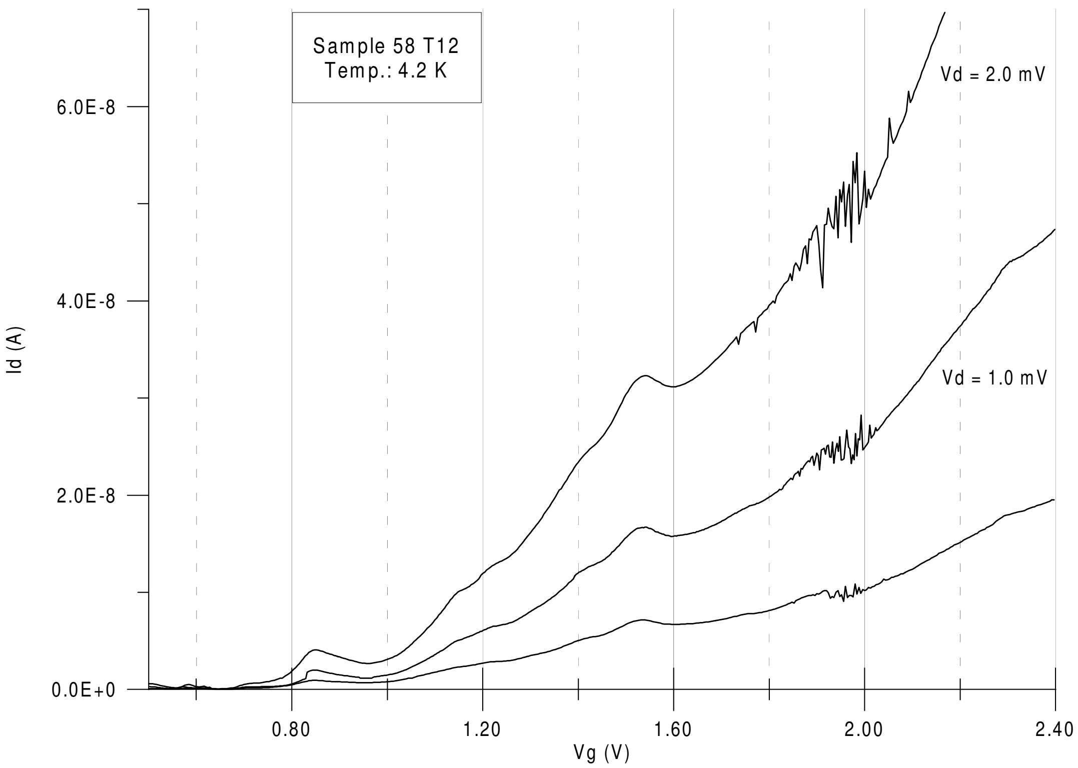


Figure 4.5: Input characteristics showing non-periodic fluctuations in the channel transconductance. Channel width and length are $W = 1.2 \mu m$ and $L = 60 nm$, respectively. Gate oxide thickness t_{ox} is 3.8 nm. Measurements were carried out at 4.2K. In the higher gate voltage region (around 2 V), fluctuations caused by trapping and detrapping of electrons in localized defect states are also observed.

2. The difference between the fitting polynomial and the measured data is calculated.

3. The discrete Fourier transform of the difference calculated in the preceding step is evaluated.

It should not be implied that the low-temperature behavior of the channel conduction of sub-100nm MOSFETs is properly modeled by an expression of the form $I_D = k_3V_G^3 + k_2V_G^2 + k_1V_G + k_0$. This expression is only used to filter out the monotonous low-frequency components of the characteristics.

As the Fourier transform is additive, the signal can be decomposed in a sum of components. In the scope of the Fourier analysis carried out, the characteristics is viewed as being the sum of three components. The first component is the difference evaluated in step 2 of the methodology just described. This part is called *the periodic component*. The other two components arise from the fitting polynomial. One is the mean value of the polynomial (evaluated over the V_G range of the measured characteristics), and the other is the difference between the mean value and the polynomial fit itself. As a result, this third part, called *zero mean polynomial*, has a mean value of zero. The discrete Fourier transform of the original characteristics is then the sum of the transforms of the three components. The Fourier transform of a constant (the mean value of the polynomial) has the form of a Dirac delta function at the zero frequency component. The discrete Fourier transform of the polynomial is a monotonically decreasing function of the frequency. It follows that the peaks present in the spectrum of *the periodic component* are also part of the whole spectrum of the characteristics (the sum of the 3 components). The periodicity of the oscillations is then inferred from the peaks in the discrete Fourier power spectrum of *the periodic component*.

This methodology is exemplified in Figures 4.6 and 4.7. The left side of Fig. 4.6 depicts the original I_D *x* V_G characteristics. This characteristics happens to show weak fluctuations in the drain current as the gate voltage is increased. Initially, no periodic component seemed to be present. The right side of the figure depicts the fitting polynomial and the *the periodic component*. Figure 4.7 presents the discrete Fourier transform of the *zero mean polynomial* (right side), which is a monotonically decreasing spectrum, without maxima nor minima, and the transform of *the periodic component* (left side), where a clear peak can be seen at the 10 Hz frequency component.

Table 4.1 summarizes the fraction of samples that did show a peak in the Fourier spectrum at the (10 ± 1) Hz component (corresponding to the 100 mV period oscillations) and/or at the (28.5 ± 1) Hz component (corresponding to the 35 mV period oscillations). Only peaks with a peak to valley ratio of at least 2:1 are considered. The V_G range analyzed was always more than 600 mV for the oscillations of 100 mV period (that is, at least 6 times the period of interest), and more than 175 mV for the oscillation of 35 mV period (that is, at least 5 times the period of interest).

4.2 I_D *x* V_D Characteristics

Fig. 4.9 shows a typical I_D *x* V_D (output) characteristics at low temperature. The solid curve marked with full circles is for the device biased at a gate voltage corresponding to a peak in the

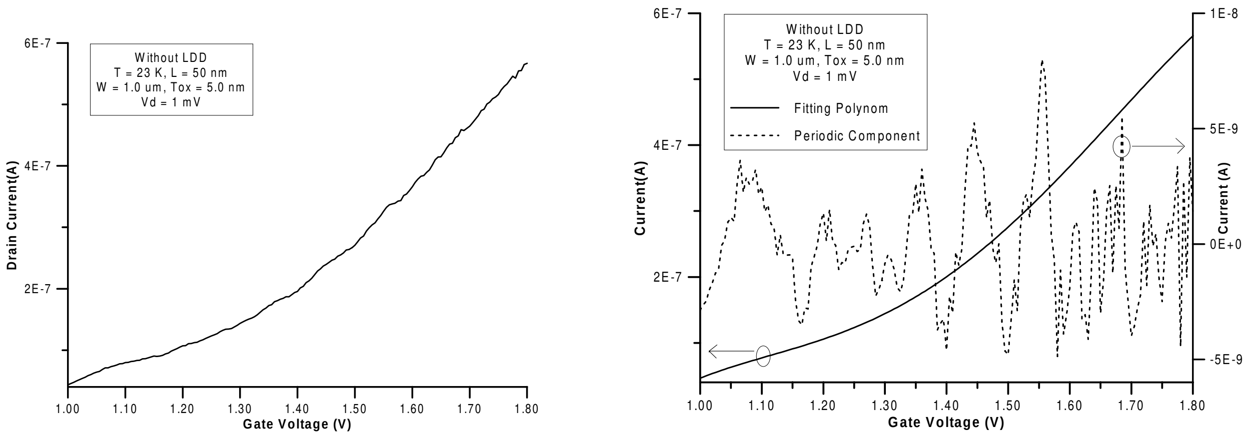


Figure 4.6: Left side: Input characteristics showing transconductance fluctuations. Right side: A polynomial of order 3 fitted to the input characteristics (Fitting Polynomial) and the arithmetic difference between the fitting polynomial and the measured data (Periodic Component). For more details please refer to the text.

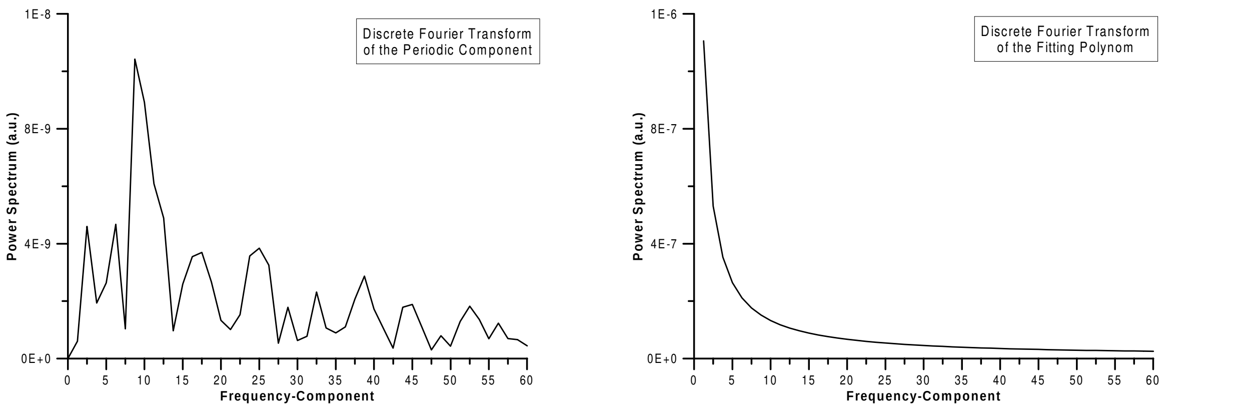


Figure 4.7: The discrete Fourier Transform of the Periodic Component (left side) and of the Fitting Polynomial (right side). For more details please refer to the text.

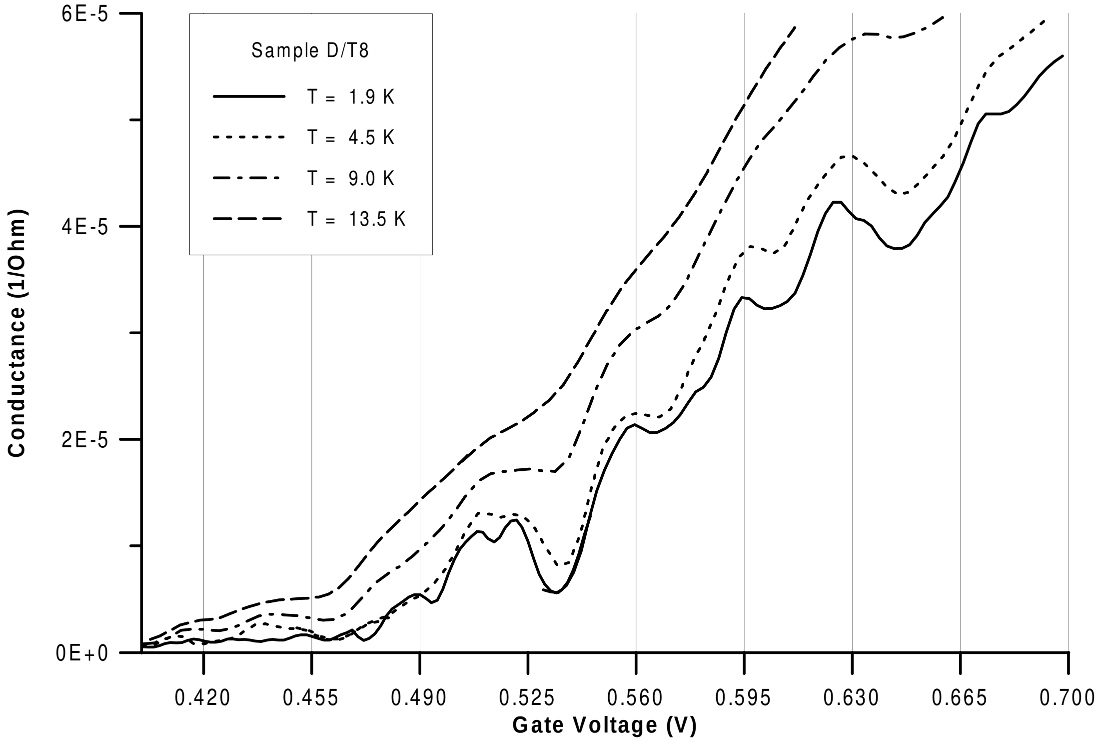


Figure 4.8: Temperature dependence of the oscillations. Drain voltage V_D is 0.15 mV. Channel length and width are $L = 30nm$ and $W = 0.8\mu m$, respectively.

Wafer	Nr. of Samples	100 mV Period	35 mV Period
366/3	15	04 (26.7%)	02 (13.3%)
366/13	14	04 (28.6%)	01 (07.1%)
377/3	08	01 (12.5%)	00 (00.0%)
397/2	14	03 (21.4%)	02 (14.3%)
397/6	22	08 (36.4%)	06 (27.3%)
413/6	11	03 (27.3%)	02 (18.2%)
430/1	14	03 (21.4%)	02 (14.2%)
430/3	09	02 (22.2%)	00 (00.0%)
430/5	10	03 (30.0%)	01 (10.0%)
430/7	09	03 (33.3%)	02 (22.2%)
Total	126	34 (26.9%)	18 (14.3%)

Table 4.1: The occurrence frequency of the periodic oscillations in the studied samples. For further details please refer to the text.

input characteristics. The dashed line marked with open circles corresponds to the gate bias of the subsequent valley, at a higher gate bias. A so-called drain threshold voltage due to lack of source-gate overlapping (discussed in section 2.2) was not observed in our devices.

This smooth characteristics, without noticeable conductance fluctuations, is found for all devices. For the devices that show a negative differential resistance (NDR) in the input characteristics, the current at the peak is greater than the current at the subsequent valley (at a higher gate bias), among the smallest drain biases. With increasing drain bias, the valley current overtakes the peak current and the usual output characteristics is found, i.e., higher drain currents at higher gate biases.

The conclusions that can be achieved by the analysis of this characteristics in the framework of different electric transport models are discussed in chapter 5.

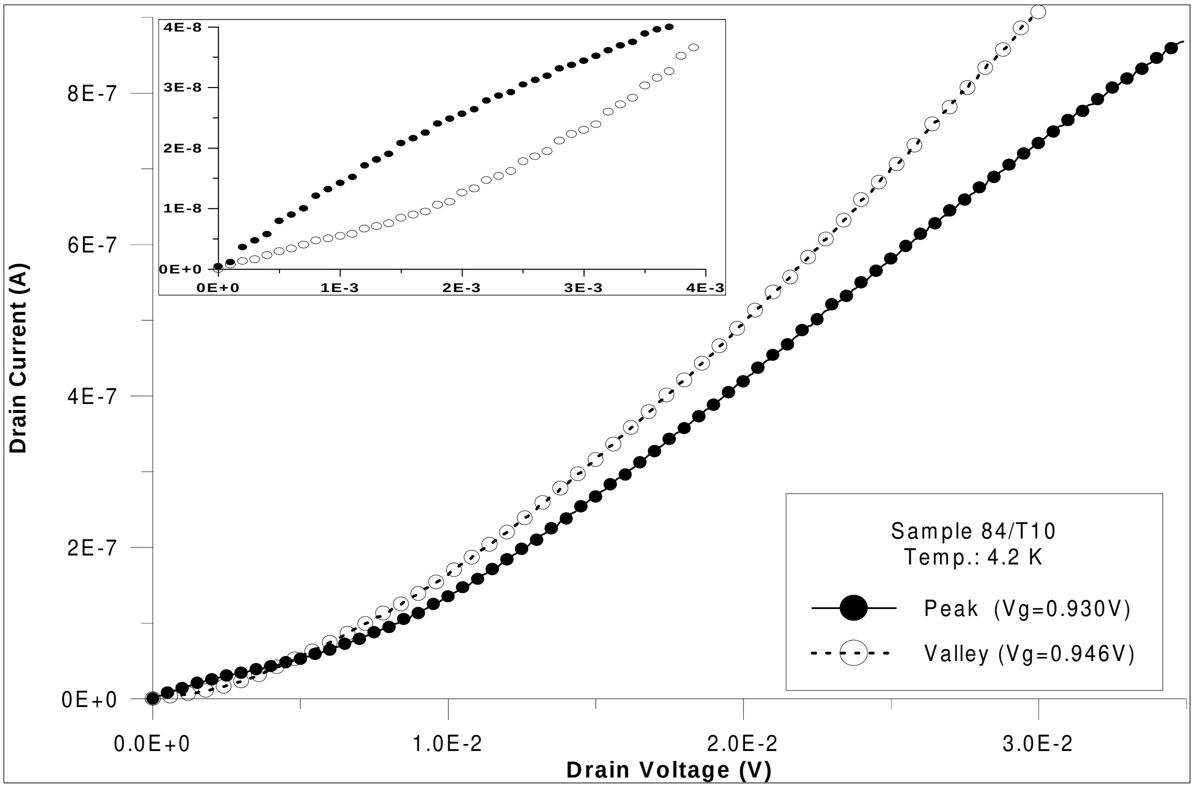


Figure 4.9: Output characteristics for gate bias corresponding to a peak (full symbols) and the subsequent valley (open symbols). The inset depicts the region with the smallest drain bias in a more appropriate scale. Channel length and width are $L = 50nm$ and $W = 1.0\mu m$, respectively.

4.3 Source Voltage Sweep

Curve traces were also taken sweeping the source bias V_S, at constant drain-source bias V_{DS}.

The measurements are carried out with gate and bulk potentials held constant. Drain-source V_{DS} bias is also kept constant, while V_S and V_D are simultaneously swept. With each measurement point the drain and source potentials are incremented by the same amount.

Figure 4.10 depicts the typical characteristics found for these traces. The solid line is for forward operation. The dotted line is for reverse operation, where source and drain contacts are interchanged. The I_D *x* V_G characteristics (dashed line) is also ploted. Note that the gate bias V_G, plotted on the upper $y - axis$, is on the same scale as the source bias (lower $y - axis$). As can be seen in the figure, the pattern found if the source voltage is swept is the same found in the I_D *x* V_G sweeps, and does not change if source and drain contacts are interchanged.

This behavior is discussed in chapter 5, under the framework of the different electric transport models there discussed.

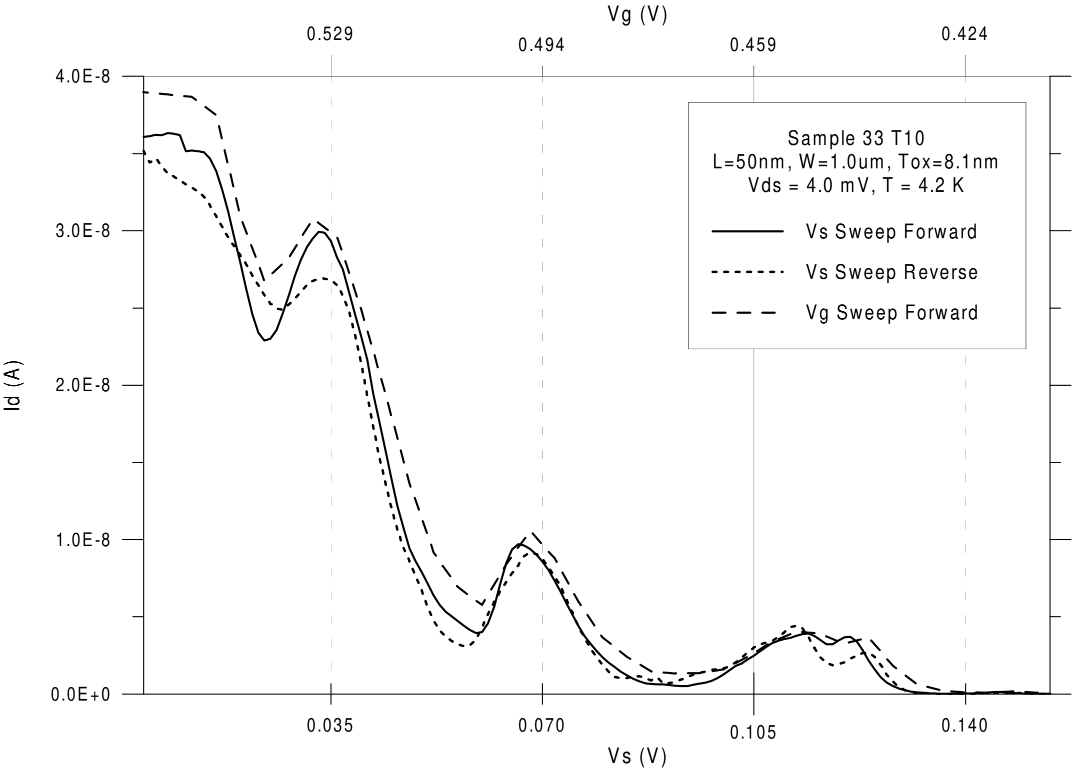


Figure 4.10: Source voltage sweep at 4.2 K. Source bias V_S is swept, while drain-source bias V_{DS} is held constant at 4 mV. The solid line is for forward operation. The dotted line for drain and source contacts interchanged (reverse operation). In both cases V_G is kept constant at 0.564 V. The dashed line depicts the I_D *x* V_G characteristics. For more details, please refer to the text.

4.4 Magnetic Field Dependence

Using the measurement setup described in section 3.2, the behavior of the samples under high magnetic fields is studied. Magnetic fields up to 17 T are applied perpendicular to the Si/SiO_2 interface. As can be seen in Fig. 4.11, the periodicity of the oscillations does not change if a magnetic field is present. Only the overall conductance is seen to decrease with increasing magnetic fields, a well known behavior also found in standard long MOSFET devices (Bagwell, Park, Yen, Antoniadis, Smith, Orlando, and Kastner 1992).

The implications of pattern invariance in the presence of magnetic fields are discussed in detail in chapter 5, in the framework of the different electric transport models there presented.

The following describes a quantum confinement effect expected if high magnetic fields are present.

If a magnetic field is applied perpendicular to the channel plane (x, y-plane) the density of states of the system is affected and one expects to observe the Shubnikov-de Haas effect (Ouisse, Cristoloveanu, and Maude 1993): the quantum mechanical states rearrange in the presence of the magnetic field, and various points in the k_x, k_y plane condensate into points on circles which represent constant energy surfaces, called Landau levels. As the magnetic field is altered, the separation of the Landau levels changes and these move across the Fermi surface (Berggren, Thornton, Newson, and Pepper 1986), resulting in oscillations in channel conductivity. If only one subband is occupied, the oscillations are expected to be periodic if plotted as a function of (1/B), with a period

$$\Delta\left(\frac{1}{B}\right) = \frac{e}{\hbar\pi^2 N_{inv}} \tag{4.1}$$

where N_{inv} is the inversion layer charge density and B is the magnetic field.

Theoretically, Shubnikov-de Haas oscillations can be observed if the separation between Landau levels exceeds the thermal energy and if the relaxation time τ is greater than the cyclotron period. The separation between Landau levels Δ_L is given by

$$\Delta_L = \hbar\omega_c \tag{4.2}$$

The cyclotron frequency ω_c is:

$$\omega_c = \frac{eB}{m_c c} \tag{4.3}$$

where m_c is the cyclotron mass and c the speed of light in vacuum. For electrons in a parabolic band one has $m_c = m_e^*$, the effective electron mass (Schoenberg 1984).

Subsequently, one can expect to observe the Shubnikov-de Haas oscillations if ($k_B T << \hbar\omega_c$) and ($\omega_c\tau << 1$). The first condition is expected to be satisfied for magnetic fields $B > 4T$. The second condition is related to τ, which depends mainly on the channel carrier mobility.

As can be seen in Fig. 4.12, it is not possible to observe the Shubnikov-de Haas effect in our probes. This can be the consequence of level broadening, meaning that the constrain ($\omega_c\tau << 1$) is not satisfied by our probes.

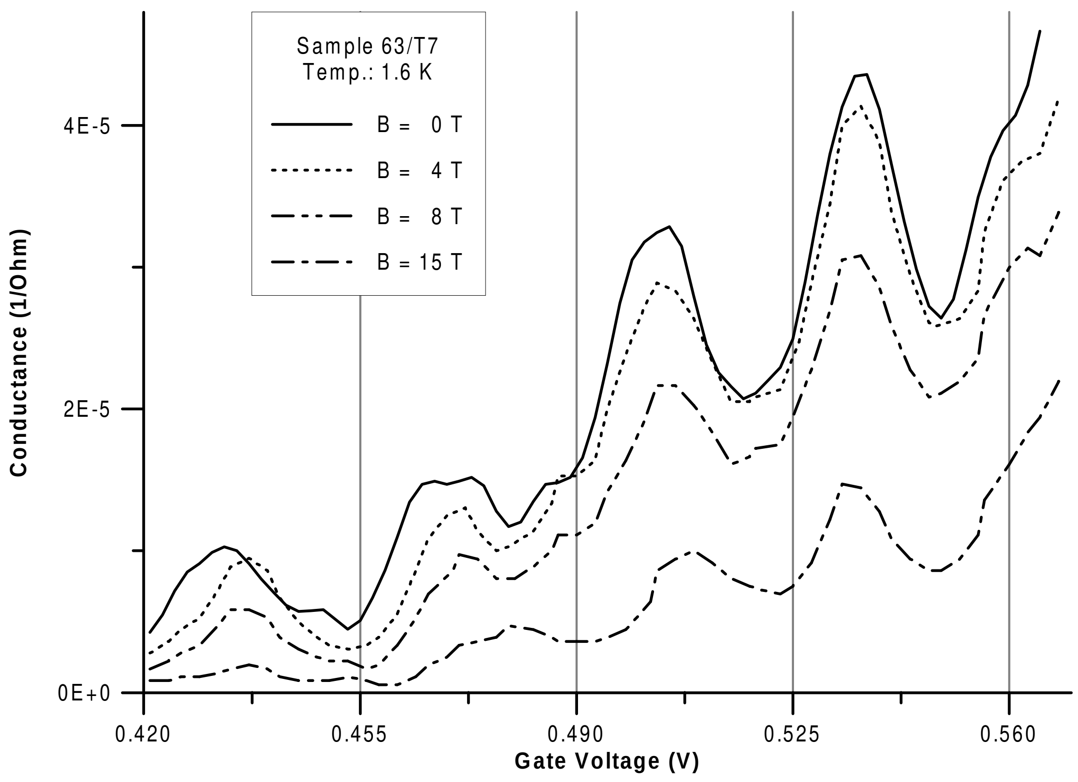


Figure 4.11: Magnetic field dependence of the oscillations. Drain voltage V_D is 0.2 mV. Channel length and width are $L = 30nm$ and $W = 0.8\mu m$, respectively.

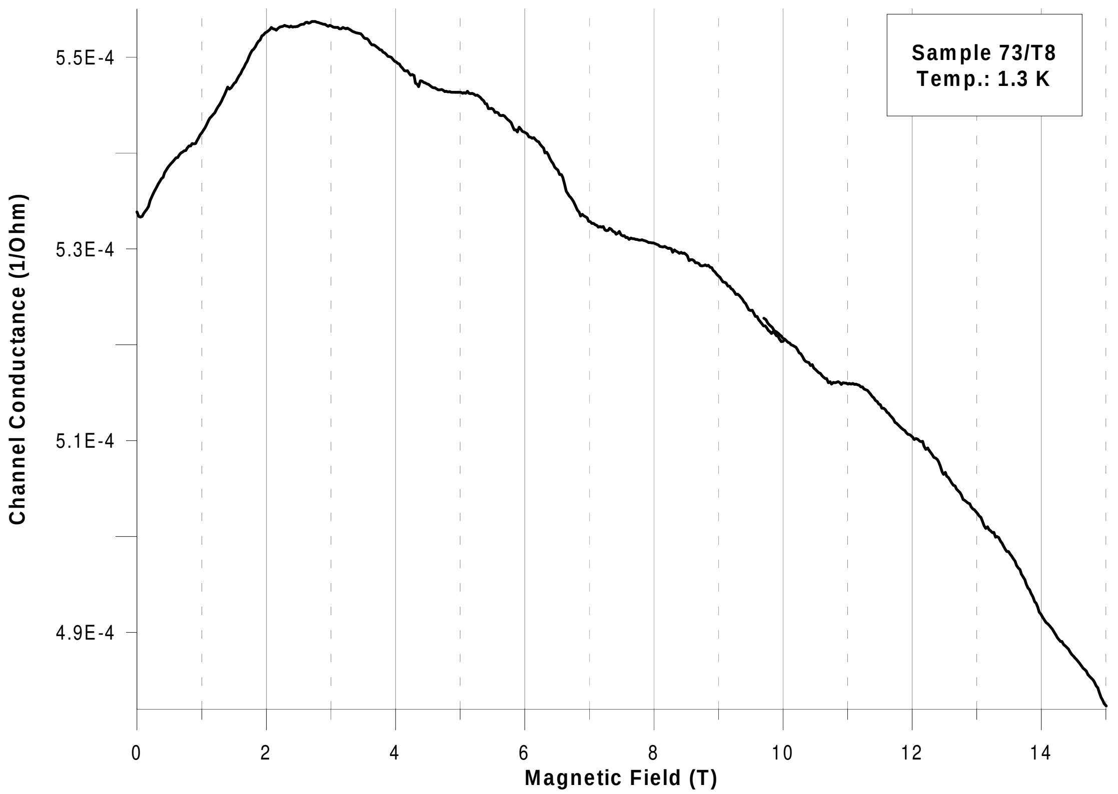


Figure 4.12: Channel conductance as a function of magnetic field at 1.3 K. Drain voltage V_D is 0.2 mV and gate voltage V_G is 1.658 V. Channel length and width are $L = 30nm$ and $W = 1.0\mu m$, respectively.

Similarly, oscillations in the channel conductance are also expected when the magnetic field is held constant while the gate voltage, and thus the inversion layer charge density and the Fermi energy, are varied (Fowler, Fang, Howard, and Stiles 1966; Mieville, Ouisse, Cristoloveanu, Forro, Revil, and Dutoit 1994). In this case, the Fermi surface is expected to move across the Landau levels. As seen in Fig. 4.11, the pattern in the $I_D x V_G$ characteristics is not affected by the magnetic field. Only the overall conductance is seen to decrease with increasing magnetic field. To observe the Fermi surface moving through the Landau levels, the constrains ($k_B T << \hbar\omega_c$) and ($\omega_c \tau << 1$) must be fulfilled.

The results presented here will be discussed further in the scope of the electric transport models presented in chapter 5.

5. Periodic Oscillations: Theory

In this chapter various models for mesoscopic phenomena expected to be important in the electrical transport of MOSFETs at cryogenic temperatures are discussed and compared to our experimental results, previously described in chapter 4.

In section 5.1 hopping conductance and resonant tunneling are analyzed. Section 5.2 deals with weak interference between source and drain junctions. In section 5.3 the new quantum transport model proposed by Dorda (Dorda 1990; Dorda 1992) is studied. Section 5.4 is dedicated to the study of charge density waves. It is shown that, for several reasons, all these models can not be accounted for the periodic oscillations observed. Coulomb blockade, discussed in sections 5.5 and 5.6, can provide a rather plausible explanation for the observed effects, although some questions, as the origin of the quantum dots, remain open.

5.1 Hopping Conductance and Resonant Tunneling

Imperfections as charged centers in the oxide, surface roughness or random fluctuations in the dopant concentration can cause discrete bound states or random surface potential fluctuations, as shown in Fig. 5.1. If the potential fluctuations are mainly long range, one can speak of isolated "lakes" or networks of electrons, and the electrons may be described by Bloch wave functions within a potential minimum. If short range fluctuations dominate, the low energy electrons can be localized in the potential wells until the electron density is such that $k_F l < 1/2\pi$, where l is the mean free path and $k_F = 1/\lambda_F$ is the inverse of the Fermi wavelength (Mott 1974). In either case, the density of states no longer abruptly increases to its constant value for two dimensions. Instead, there is a band tail, as depicted in Fig 5.2. Below the mobility edge E_m the electrons are expected to be localized (Koch, Bollu, and Asenov 1988; Bollu and Koch 1988; Aronzon, Vedeneev, and Rylkov 1997).

In the limit of zero temperature the conductance of such a sample will be limited by tunneling through potential barriers. For finite temperatures phonon-assisted tunneling may become important, as shown schematically in Fig. 5.3 for one dimensional cases. This thermally activated conduction mechanism is called variable range hopping, because the hops that dominate the conduction change when the temperature changes or when the Fermi level is swept. The resistance of a hop is an exponential function of the geometrical separation of the sites and their energy difference, both of which are expected to be random variables. The resulting distribution of hopping resistances is enormously broad (it is log-normal (Kastner, Kwasnick, and Licini 1987)). The activation energy T_A is fixed as long as a given hop limits the current. However, if the temperature is lowered enough, the resistance of that hop becomes high enough that the electron will hop to a more distant site with smaller activation energy. This mechanism is the origin of the variable range hopping itself, and the activation energy, on average, decreases in the way predicted by the $exp[-(T_0/T)^{1/(d+1)}]$ dependence, where d is the dimensionality of the sample (Raikh and Ruzin 1990). At very low temperature only hops with energies within the Fermi levels of the reservoir will be reachable. Then resonant tunneling is the only conduction mechanism.

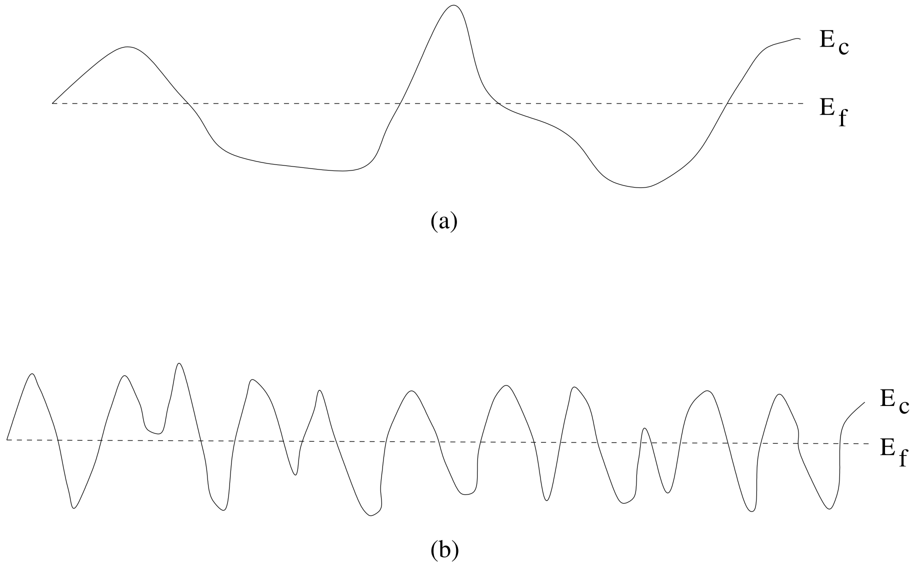


Figure 5.1: Potential fluctuations in the conduction-band edge. In (a) long-range fluctuations are depicted, whereas (b) shows short-range fluctuations. When the Fermi level is at the position shown the electrons might be expected to be localized.

Based on these facts, two different scenarios underlying current transport in the probes could be suggested: variable range hopping or resonant tunneling.

For variable range hopping the decay length $1/\alpha$ of the localized wave functions is a function of the Fermi level (a function of gate voltage). The localized wave functions are expected to decay slowly as the Fermi level gets closer to the conduction band, where V_G increases. Based on the work of Adkins, Pollit and Pepper (Adkins, Pollit, and Pepper 1976), on the experimental results of this work (characteristic temperature of the order of 35 K for the V_G range studied) and on the geometries of our devices, the decay length $1/\alpha$ can be estimated to be of the order of some nanometers, if two-dimensional variable range hopping is to be observed. The most probable hopping length R is given by (Webb, Fowler, Hartstein, and Wainer 1986)

$$R = \frac{1}{(\pi\alpha N_1 k_B T)^{1/3}} \tag{5.1}$$

where N_1 is the two dimensional density of states, typically $N_1 = 1.6x10^{14} eV^{-1} cm^{-2}$ (Webb, Fowler, Hartstein, and Wainer 1986). This leads to an R of the order of 15 nm at 0.35 K, the

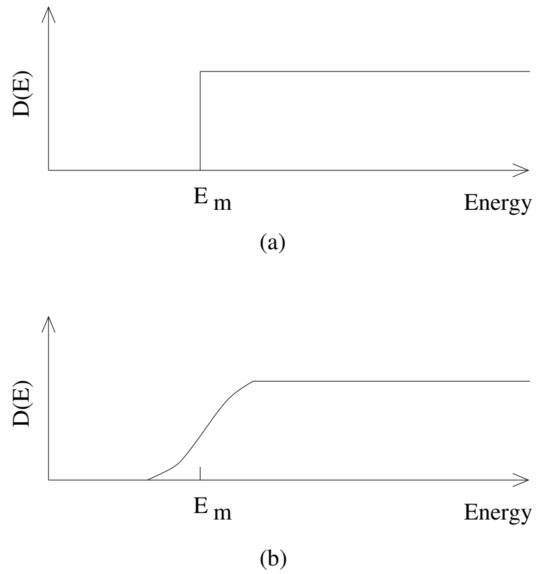


Figure 5.2: The density of states $D(E)$ for (a) an ideal and for (b) a perturbed two-dimensional conduction band. E_m is the mobility edge.

lowest temperature at which measurements where carried out, and of the order of 4 nm at 35 K, the highest temperature at which the oscillations have been observed. Since $R < L < W$ for the whole temperature range, the first condition for two dimensional hopping can be satisfied.

For a sample with, for instance, $L = 50nm$ and $W = 1\mu m$, where the periodic oscillations have been observed, the area is $A = LxW = 5x10^{-11}cm^2$. With the two-dimensional density of states assumed ($1.6x10^{14}eV^{-1}cm^{-2}$), there are $8x10^4$ states per eV in this device area. At 0.35 K one gets approximately 2 states within k_BT. At 4.2 K approximately 29 states. Probably three to five times this number can be expected to contribute to the conduction (Webb, Fowler, Hartstein, and Wainer 1986). With this difference in the number of states distributed within k_BT and the related change in the most probable hopping length R, one would expect a change in the pattern present in the measured current *versus* voltage characteristic of the samples with temperature, that is, the percolation path is expected to be build up by different hops at different temperatures. This was not observed. The oscillation are smeared out with rising temperature, but the pattern does not change.

Variable range hopping is also not expected to produce a pattern that is reproducible between samples and periodic in gate voltage. The gate voltage sweeps the Fermi level and a new percolation path is expected to dominate the electrical transport. There is no reason to expect that equivalent paths do exist at regular steps in gate bias. Furthermore, asymmetrical and rectifying behavior is expected when source and drain are interchanged, depending on the particular structure of the percolation path (Fowler, Timp, Wainer, and Webb 1986). This behavior could also not be observed.

The numerical results used above are simplified but realistic approximations, and the qualitative behavior would remain the same if more precise values of the parameters were available.

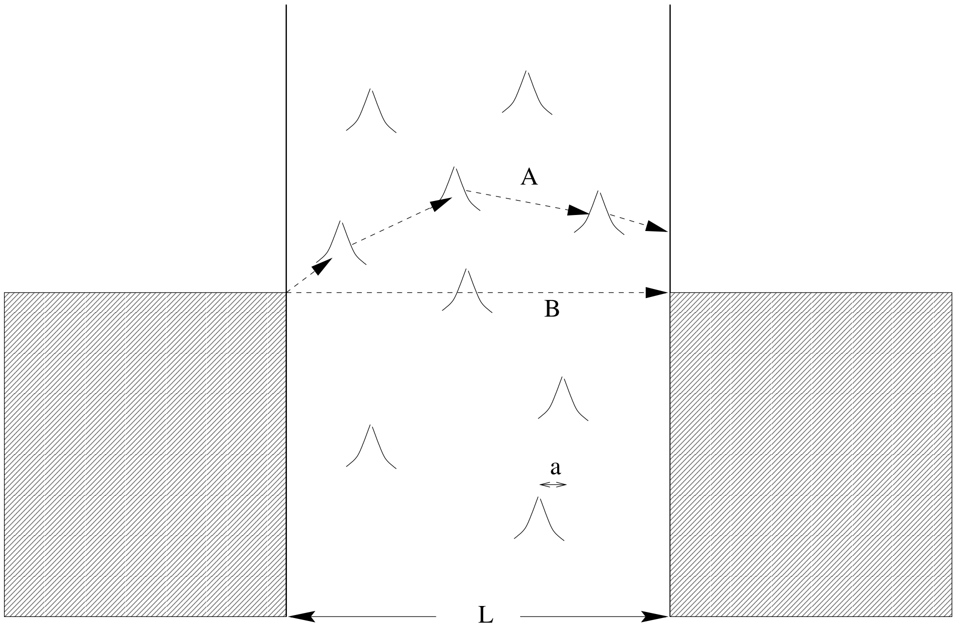


Figure 5.3: Sketch of a disordered conductor with localized states between ideal contacts. The decay length $a = 1/\alpha$ of the localized wave functions is much less than the channel length *L*. Path *A* depicts thermally activated hopping conduction and path *B* resonant tunneling.

Furthermore, although it is difficult to determine the temperature dependence of the model parameters, due to the possible interplay of different phenomena, the expected $-(T_0/T)^{1/(d+1)}$ dependence of conductance on temperature, where d is dimensionality of the system, does not apply to our experimental results, as discussed in chapter 4.

Thus, it is very unlike that the pattern measured is a consequence of hopping conduction.

The competing possible transport mechanism, resonant tunneling, depends also on the sparsity of localized states in the system. If the energy of the incoming electron is resonant with a localized site, as path B in Fig. 5.3 shows, the transmission coefficient is Tr_1/Tr_2, where Tr_1 and Tr_2 are the transmission coefficients into and out of the site, provided that $Tr_1 \leq Tr_2$. This determines that only the sites near the middle of the sample are expected to be effective. As a result, this should be the only current path at $T = 0$, and resonant tunneling is expected to be important only at very low temperatures or for very short samples. In the limit that the decay length equals the channel length the transmission coefficient will be nearly 1 because every state is within $1/\alpha$ of the center of the sample.

As resonant tunneling can occur in parallel with other transport mechanisms, such as conventional drift-diffusion transport, it is difficult to determine the temperature dependence of the model parameters with satisfactory accuracy. As the temperature is swept, other parameters like the threshold voltage and the mobility of the carriers do not keep constant and it is not possible to isolate the individual contributions of the different transport mechanism to changes in the current. This makes a quantitative analysis quite unreliable. However, if electrons are to be localized, the tunneling conductance must be less than e^2/h. As discussed in chapter 4, section 4.1, for some samples the contribution of the mechanism responsible for the negative differential resistance is more than e^2/h. Therefore, the pattern observed cannot be the consequence of tunneling through a single resonant state. Of course, one could have two or more tunneling channels in parallel, all of them at the same resonant level. But this is very unlike, specially if one notes that if the pattern is to be reproducible between samples, as in our case, the distribution and properties of the resonant states have to be the same in these samples. These are expected to be random variables. Furthermore, if a magnetic field is applied normal to the sample in the disordered region, the position, height, and width of every individual peak is expected to change in a random way (Xue and Lee 1988). Peaks that overlap at zero magnetic field should separate as a magnetic field is applied and become discernible. This phenomena was not observed.

For the same reasons, universal conductance fluctuations, as predicted by Lee and Stone (Lee and Stone 1985) for disordered systems, proved unlikely to be the phenomena observed here.

5.2 Weak Interference between Source and Drain

If the channel length is of the order of both the de Broglie wavelength of the electrons at the Fermi surface λ_F and the inelastic scattering length, quantum mechanical phenomena is expected to be observable in the electrical carrier transport (Hartstein 1991). Omura et al. (Omura, Kurihara, Takahasi, Ishiyama, and Izumi 1997; Omura and Izumi 1996; Omura, Horiguchi, Tabe, and Kishi 1993) suggested that in this case weak interference between source and drain should occur. Weak interference predicts that the transmission probability of electron from the source to the drain should depend on the relation between effective channel length L_{eff} and Fermi wavelength λ_F. The transmission probability is expected to be a maximum when L_{eff} is a half integer multiple of λ_F, i.e., when

$$L_{eff} = \frac{n}{2}\lambda_F \tag{5.2}$$

where n is an integer, holds. The Fermi wavelength is given by

$$\lambda_F = \left(\frac{2\pi g_v}{N_{inv}}\right)^{0.5} \tag{5.3}$$

where g_v is the valley degeneracy and N_{inv} is the carrier density, which can be calculated taking into account the inversion layer capacitance. Using 5.2, 5.3 and approximating the inversion layer capacitance as $C_{ox} = \varepsilon_{ox}/t_{ox}$ the following relation for the values of the effective gate bias V_{Geff} at which a peak in the channel conductance is expected is found

$$V_{Geff} = \frac{\pi g_v t_{ox} n^2}{2\varepsilon_{ox} L_{eff}^2} \tag{5.4}$$

This relation clearly predicts a peak separation ΔV_{Geff} that increases linearly with the peak number n, in contrast to the constant peak separation found in the experimental results. The strong dependence on L_{eff} predicted by 5.4 is also not in agreement with the experimental findings, since no relation between channel length and peak separation was observed. It is also worth noting that the group of Omura (Omura, Kurihara, Takahasi, Ishiyama, and Izumi 1997) carried out their measurements at 39 K, a temperature at which many peaks are hard to recognize, particularly for measurements at higher V_G values.

5.3 The Quantum Transport Model proposed by Dorda

A new general quantum transport model based on a reformulation of the Bohr quantum model was proposed by G. Dorda (Dorda 1990; Dorda 1992). In this model the resistance of every specimen is assumed to be quantized in terms of h/e^2, leading to the following quantized formula for two dimensional conductance

$$G = \frac{e^2}{h} \sum_{n_w} (n_d)^{-1}_{n_w} \tag{5.5}$$

where n_w corresponds to the number of current filaments and n_d to the number of coherent sub-domains within a current filament (please refer to Fig. 5.4). It is assumed that only electrons in this collective state can contribute to the electrical transport and that the trend to collectivization predominates in the current filament, prevented mainly by phonons and ion disturbances. In short, the current is assumed to be carried by a net of parallel filaments, the conductance of each depending on the degree of organization (grouping) of the electrons forming it, as schematically shown in Fig. 5.4 for a homogeneous sample.

The number n_d of coherent sub-domains in a filament is calculated as being

$$n_d = \frac{n_l}{n_{\mu e}} \tag{5.6}$$

The quantum number n_l is obtained dividing the sample length L by a reference length $l_e = 18.2nm$, i.e., $n_l = L/l_e$ (Dorda 1990; Dorda 1992). The electrons are thought to be localized at a distance $l > l_e$, whereas at $l < l_e$ they are assumed to be indistinguishable, grouped into the coherent sub-domains.

The quantum number $n_{\mu e}$ is assumed to be a physically incontestable quantum number. The splitting of the magnitude $N\mu$ among N, the number of electrons, and μ, the mobility, is viewed as a physically unfounded procedure. Instead, the relation $N\mu = n_{\mu e}e/h$ must be introduced.

According to the above assumptions, for electrons uniformly distributed over the sample, no electrical transport is possible for electron concentrations below a reference concentration N_0, for which the distance between electrons is greater than l_e, the localization length. This concentration is calculated as being $N_0 = (l_e)^{-2} = 3.0x10^{15}m^{-2}$. This situation is schematically represented on the left side of Fig. 5.5. However, for inhomogeneous samples, the localization predicted at concentrations $N < N_0$ can be countered to some extent by spatial redistribution to create one-dimensional current filaments, as shown on the right side of Fig. 5.5. Equation 5.5

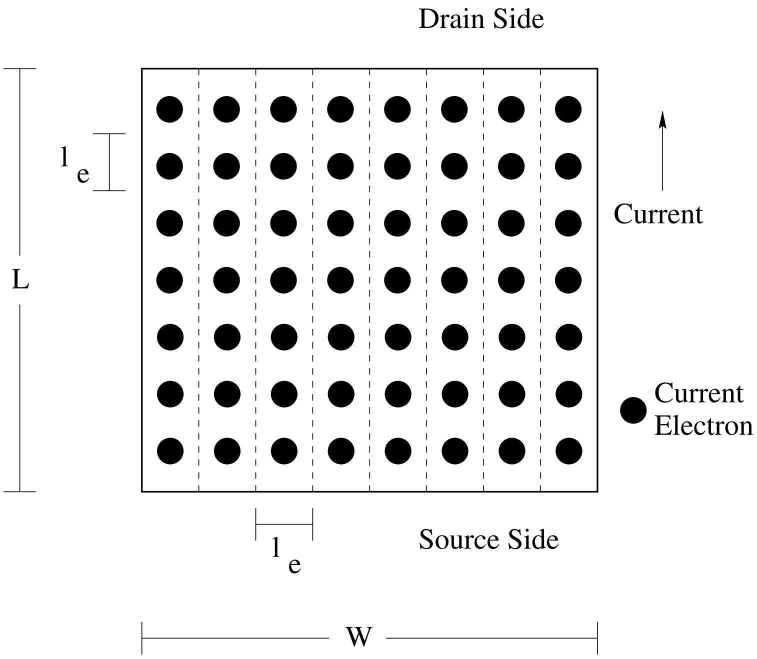


Figure 5.4: Current transport by means of current filaments linking the source and drain contacts. The dashed lines represent the boundaries between the collective modes. The case for a electron density equal to $N_0 = 3.0x10^{15}m^{-2}$ in a homogeneous sample is shown. l_e is the critical width for the electrons being localized or in the collective mode.

shows that a redistribution of the electrons would lead to a finite conductance. N_0 is assumed to be a universal quantity, completely independent of the geometry and material properties of the samples.

The most relevant prediction of the model in the scope of this work is that, according to Dorda, it is expected that electrons can form preferred ordered states at quantized densities $N = nN_0$, with $n = 1,2,3,...$. At these ordered states the perturbation of the current filaments should be altered, affecting the electrical conductance as predicted by eq. 5.5. Since the gate voltage controls the amount of charge in the channel of a MOSFET, this phenomena should be noticeable in the I_D x V_G characteristics, similar to that observed in the devices studied in this work.

The first step in the way to verify if the predictions of the Dorda's model correspond to our experimental findings is to compare the electron densities at which the oscillations are found with N_0. In a first rough approximation, the inversion layer charge density can be estimated using the oxide capacitance C_{ox}. As discussed in chapter 2, this leads to an overestimation of the inversion layer charge density. The dependence of the inversion layer charge density N_{inv} on the gate voltage V_G is then given by $N = C_{ox}(V_G - V_T)$. For the transistors with a gate oxide thickness t_{ox} of 7 nm, C_{ox} is equal to $4.9mF/m^2$. For $t_{ox} = 5nm$, C_{ox} is $6.9mF/m^2$, and finally for $t_{ox} = 8.7nm$, C_{ox} is $4.0mF/m^2$. Using these values the oscillations can be interpreted as occurring at constant intervals of charge inversion layer density of $\Delta N = C_{ox}\Delta V_G = 4.3x10^{15}m^{-2}$, for the samples with $t_{ox} = 5nm$ and $\Delta V_G = 0.1V$. For the samples with $t_{ox} = 7nm$ one obtains $\Delta N = 3.1x10^{15}m^{-2}$

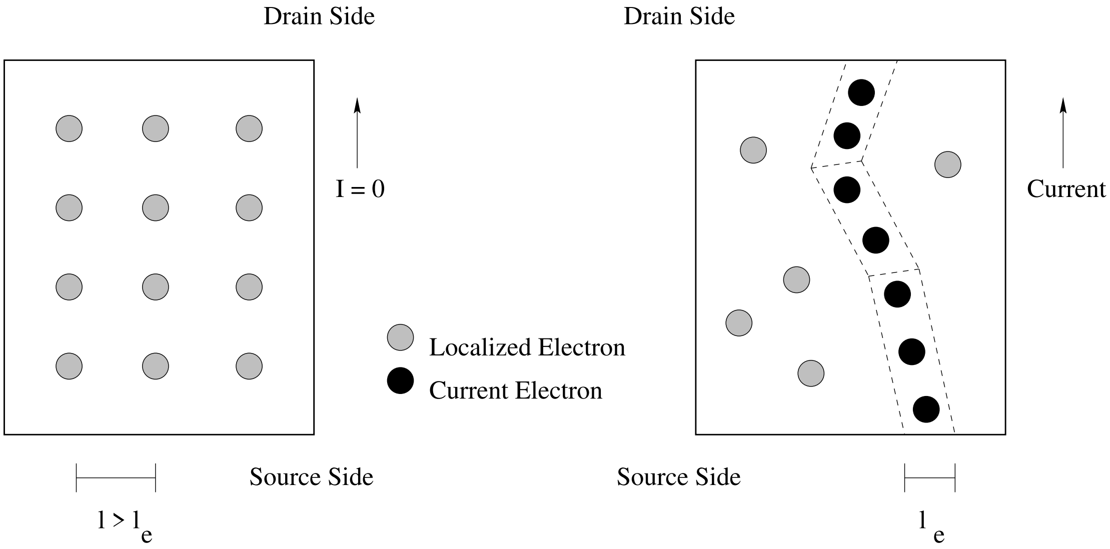


Figure 5.5: Current transport in the case $N < N_0$. For homogeneous samples (left side) the separation between electrons is greater than the critical length l_e and they are localized. On the right side a possible configuration in an inhomogeneous sample is depicted. The electrons group and form a current filament. Dashed lines represent the boundaries between the collective modes.

and finally, with $t_{ox} = 7nm$ one obtains $\Delta N = 2.5x10^{15}m^{-2}$. Although these values are close to N_0, the dependence of the gate voltage steps ΔV_G between peaks on t_{ox} predicted by this model could not be observed. Other groups which used much thicker gate oxides obtained the same gate voltage steps ΔV_G between peaks of $100mV$ and were not able to observe any dependence of ΔV_G on t_{ox}, too (Eisele, Baumgärtner, and Hansch 1995; Poole, Pepper, and Myron 1983).

5.4 Charge Density Waves

At low temperature and in the weak inversion regime, the channel of the devices may be viewed as being built up by narrow conducting segments that electrically link source and drain. Hints to such a channel configuration are given in section 6.2.2. Furthermore, as discussed previously in section 2.3, it is not expected that many subbands for motion transverse to the channel are occupied. It has been pointed out that an one-dimensional electron gas coupled to the underlaying lattice is not stable at low temperatures. For a recent review see (Grüner 1992). The ground state of the coupled electron-phonon system is characterized by a gap in the single-particle excitation spectrum and by a collective mode formed by electron hole pairs involving the wave vector $q = 2k_F$. The charge density associated with the collective mode is given by (Grüner 1988)

$$\rho(x) = \rho_0 + \rho_1 cos(qx + \phi(x)) \tag{5.7}$$

where ρ_0 is the unperturbed electron density. The condensate is called the charge-density wave (CDW). Amplitude variations are energetically much more costly than phase variations, so that

$\phi(x)$ takes into account any deformations caused by external perturbations (Field, Kastner, Meirav, Antoniadis, Smith, and Wind 1990). In contrast to superconductors, the phase excitations of the collective mode are gapless. Hence, Coulombic potentials break the translational symmetry and lead to the pinning of the collective mode. These Coulombic potentials (due to impurities, oxide and interface charges, etc.) can be either of attractive (positively charged) or repulsive (negatively charged) nature (Teranishi and Kubo 1979; Meirav, Kastner, Heiblum, and Wind 1989).

The length L_O between two Coulombic centers will lead to oscillations of the pinning energy, and hence of conductance, with a period of $\Delta q = 2\pi/L_O$. This model clearly predicts oscillations with a period directly determined by the distance between these centers (impurities, oxide and interface charges, etc.). As the distance between centers is expected to be a random variable, characteristics that are reproducible from sample to sample are not expected. In addition, if there is more than one pair of Coulombic centers lying close to the conducting segments, one would get more than one period at once.

The temperature dependence of the conductance is quite similar to that in the Coulomb blockade model, discussed in section 5.5. This is because charge transport requires an extra charge to be added to the region between the pinning centers, leading to a thermally activated transport behavior resembling the Coulomb blockade.

In the CDW model a threshold behavior in the $I_D \, x \, V_D$ characteristics is also expected (Grüner 1992). The collective mode can only be driven into a current-carrying state if the electric field is strong enough to "depinne" the CDW. This will lead to nonlinear current-voltage characteristics with threshold behavior also in the $I_D \, x \, V_D$ traces.

Another point that makes CDW unlikely to be the phenomenon observed here, is the fact that the pinning centers get strongly screened as the inversion layer charge density increases. Hence, the pinning centers are not expected to play a major role in strong inversion.

5.5 Coulomb Blockade in a Single Dot

Due to the discrete charged nature of the electron, the energy required to put one extra electron on a capacitor plate is not a continuous variable. This is not important as long as the capacitance is large and the coulombic energy $e^2/2C$ is quite small compared to the thermal energy $k_B T$. However, as the device dimensions become shrinkingly small we find that the energy $e^2/2C$ required to put even one electron on them can exceed the thermal energy at low temperatures. This leads to the observation of a new phenomena, called Coulomb blockade. Fig. 5.6 shows a schematic cross section of such a device.

In the following discussion the capacitive coupling between the island and the electrodes will be assumed small enough that the coulombic energy is higher than the thermal energy. In addition, the island will be considered big enough that the one-electron energy levels are close together compared to the thermal and Coulombic energies, that is, quantization effects due to the confining potential will be neglected.

As shown in Fig. 5.7 the source and drain reservoirs are assumed to be metallic at thermal equilibrium at temperature T with a Fermi energy E_f. The electrical charge Q_I at the island is an integer multiple N of the elementary charge q. This charge is related to the potential V_I at the

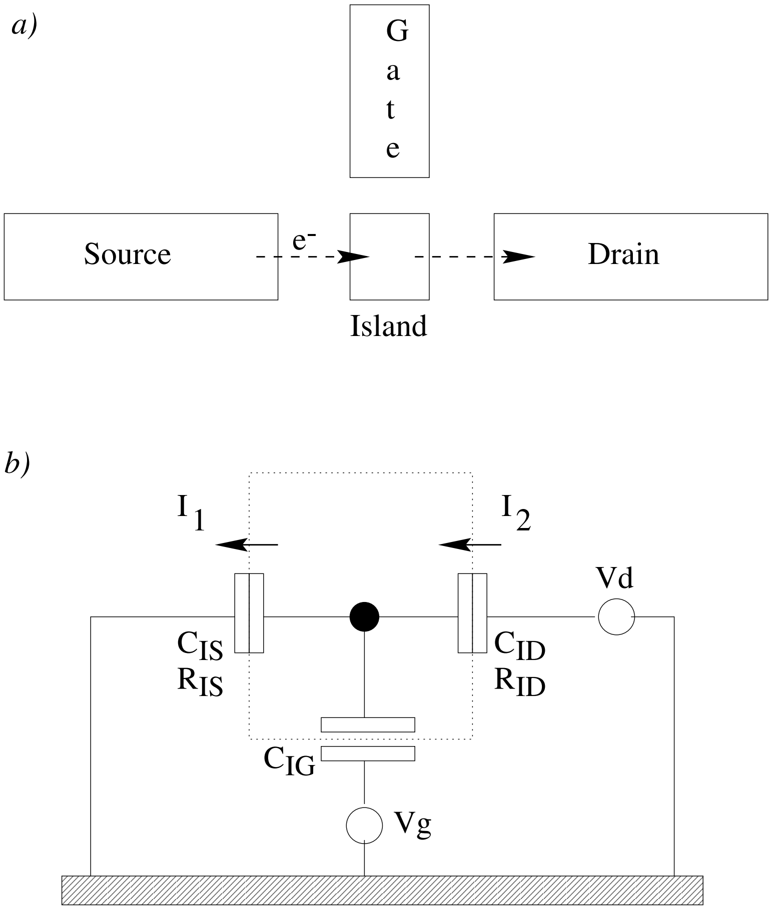


Figure 5.6: A simple coulomb blockade structure: *a)* schematic cross section of an island coupled to three electrodes. The tunneling barrier between the island and drain and source is assumed to be finite. Hence, there is a finite tunneling resistance in parallel to the capacitive coupling. The coupling between the island and the gate is assumed to be of capacitive character only, i.e., with infinite tunneling resistance. *b)* equivalent electrical circuit representation.

island and the potentials V_S at the source, V_D at the drain and to the gate potential V_G by

$$Q_I = C_{IS}(V_I - V_S) + C_{ID}(V_I - V_D) + C_{IG}(V_I - V_G) \tag{5.8}$$

where C_{IS}, C_{ID} and C_{IG} are the capacitances between the island and the source, the island and the drain, and the island and the gate, respectively. If the gate, drain, and source potentials, as well as the charge on the island are known, 5.8 can be rewritten to give the potential on the island as

$$V_I = \frac{Q_I + C_{IS}V_S + C_{ID}V_D + C_{IG}V_G}{C_{IS} + C_{ID} + C_{IG}} \tag{5.9}$$

Starting from 5.9, one can calculate the total electrostatic energy $E_{elst}(N)$ involved in the charging of the island with N electrons as being

$$E_{elst}(N) = \int_0^{-Nq} V_I(q)\,dq = \frac{-Nq(C_{IS}V_S + C_{ID}V_D + C_{IG}V_G) + 0.5(Nq)^2}{C_{IS} + C_{ID} + C_{IG}} \tag{5.10}$$

Now, if $(N-1)$ electrons have already been brought on the island, the charging energy $E_{ch}(N)$ needed to add the $(N)th$ electron is given by

$$E_{ch}(N) = E_{elst}(N) - E_{elst}(N-1) = \frac{-q(C_{IS}V_S + C_{ID}V_D + C_{IG}V_G) + (N-0.5)(q)^2}{C_{IS} + C_{ID} + C_{IG}} \tag{5.11}$$

which shows a linear dependence on the gate, source and drain potentials, as well as on the charge $(N-1)q$ already present at the island.

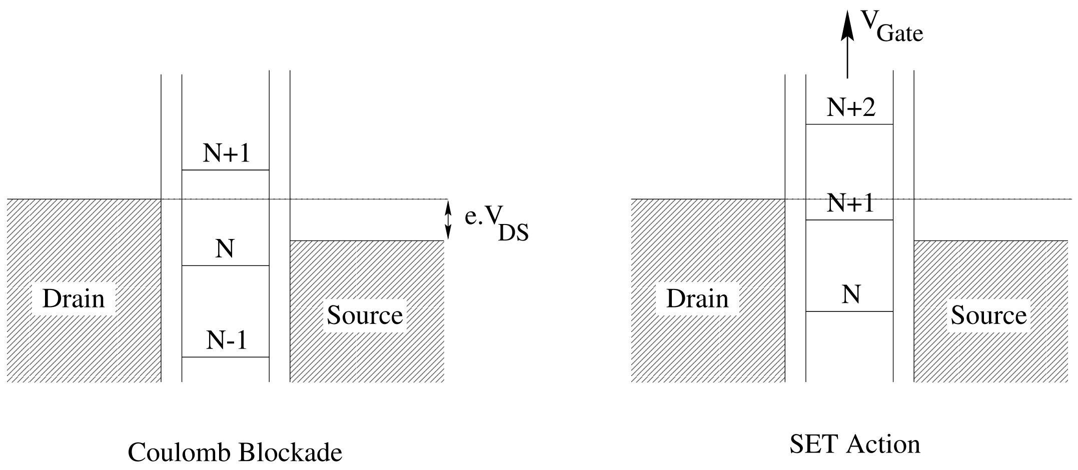


Figure 5.7: Schematic representation showing the Coulomb blockade state and Single Electron Transistor (SET) action. On the left side the energy for charging the island with one more electron, the *(N+1)th* electron, is higher than the Fermi level at the drain side, and no conduction is possible. Increasing gate voltages displace the energy ladder and the energy level for the *(N+1)th* electron becomes to lie between the drain and source Fermi levels and tunneling is possible.

Based on these facts, two different scenarios can occur for the device of Fig. 5.6. The first is called Coulomb blockade and is shown on the left side of Fig. 5.7. In the limit of zero temperature, if N electrons are already on the island and the charging energy, i.e. the energy level for the $(N+1)th$ electron, is above the source and drain potentials, no transfer of electrons into the island is possible. Furthermore, if the energy level of the Nth electron is below the source and drain potentials, no electron can leave the island as well. In this case no electrical transport is possible and the conductance of the channel is zero. Hence, the condition for Coulomb blockade can be mathematically expressed as

$$E_{ch}(N+1) > -qV_D > E_{ch}(N) \qquad and \tag{5.12}$$
$$E_{ch}(N+1) > -qV_S > E_{ch}(N) \tag{5.13}$$

In the second case the energy level for the $(N+1)th$ electron lays between the source and drain potentials. Then, as shown on the right side of Fig. 5.7, one electron can tunnel from the drain side into the island; furthermore, if the Fermi level for electrons at the source side lays below the energy level of the $(N+1)th$ electron on the island, this electron can also leave the island, tunneling into the source. Now electrical transport is possible, and is carried by one single electron at a time. The condition for single electron tunneling can then be written as

$$-qV_D > \Delta E_{ch}(N+1) > -qV_S \qquad and \tag{5.14}$$
$$|V_D - V_S| < q/(C_{IS} + C_{ID} + C_{IG}) \tag{5.15}$$

These two equations guarantee the existence of no more than one energy level of the island that lays between the source and drain potentials.

Due to the capacitive coupling between island and gate, the ladder of energy levels on the island can then be shifted by the gate potential V_G. Hence, for small V_{DS}, depending on V_G, there may be or not one electron energy level laying between the source and drain potentials, and V_G can be used to turn the electrical conduction between drain and source on or off. This is called Single Electron Transistor (SET) action.

In the limit of $V_D = V_S$ and $T = 0$ the separation in gate voltage ΔV_G between conductance peaks can be obtained noting that the potential on the island V_I depends linearly on V_G as given by 5.9 and that the island potential is related to the charging energy according to 5.11. The charging energy $\Delta E_{ch}(N+1)$ falls with raising V_G until it equals the source and drain potentials and an additional electron can tunnel into the island, that is, until the charging energy satisfies

$$\Delta E_{ch}(N+1) = -qV_D = -qV_S \tag{5.16}$$

Substituting 5.9 and 5.11 in 5.16 it is possible to obtain the equivalent condition for the potential energy on the island V_I

$$V_I(N+1) = V_D - \frac{q}{2(C_{IS} + C_{ID} + C_{IG})} = V_S - \frac{q}{2(C_{IS} + C_{ID} + C_{IG})} \tag{5.17}$$

For the $(N)th$ electron, one will find

$$V_I(N) = V_D + \frac{q}{2(C_{IS} + C_{ID} + C_{IG})} = V_S + \frac{q}{2(C_{IS} + C_{ID} + C_{IG})} \tag{5.18}$$

With V_D and V_S kept constant, the change in potential at the island due to its charging with the extra electron is equal to

$$V_I(N+1) - V_I(N) = -\frac{q}{C_{IS} + C_{ID} + C_{IG}} \tag{5.19}$$

It is interesting to note that as long as the above conditions are satisfied, tunneling into the island and from the island is possible. Hence, due to the interplay between V_G, V_I and the charge on the island, the potential of the island will continue to jump between the two values

$$V_D - \frac{q}{2(C_{IS} + C_{ID} + C_{IG})} \leq V_I \leq V_D + \frac{q}{2(C_{IS} + C_{ID} + C_{IG})} \tag{5.20}$$

This situation persists as long as the charging energy for the electron coincides with the source and drain potentials, and the number of electrons on the island keeps jumping by 1.

Consequently, it is easy to calculate the gate voltage steps ΔV_G between conductance peaks. Using 5.9 one finds the value ΔV_G for which V_I is the same for N or $N+1$ electrons on the island as being

$$\Delta V_G = \frac{q}{C_{IG}} \tag{5.21}$$

Equation 5.21 tells us that with sweeping V_G the number of electrons on the island will jump by one at constant steps ΔV_G equal to q/C_{IG}, and that conductance peaks will occur with this constant spacing, as shown in Fig. 5.8.

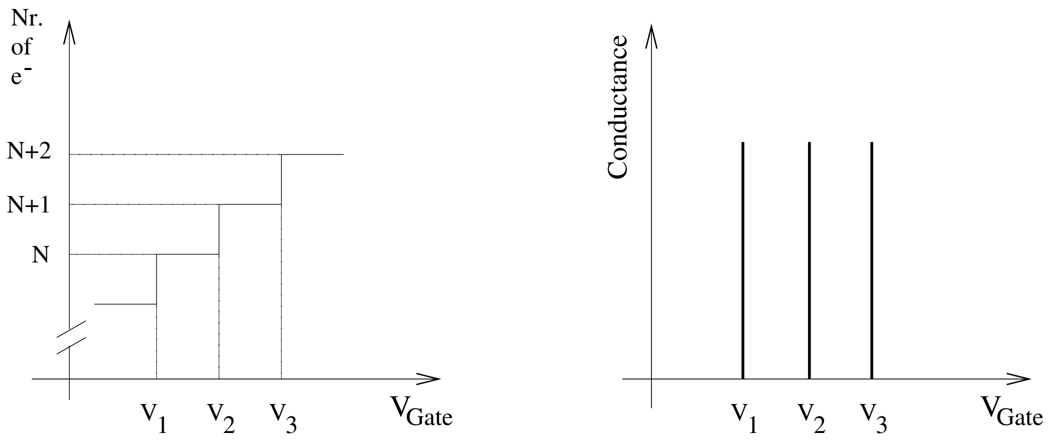


Figure 5.8: The gate voltage controls the number of electrons at the island (right) and causes conductance peaks in the channel conductance (left).

In order to observe the conductance peaks in the $I_D \, x \, V_G$ characteristics some constraints must be imposed on the temperature T, on the drain-source voltage V_{DS}, and on the total capacitance of the island.

With rising temperature, the characteristic of 5.16 is softened and thermal electrons at the source and drain electrodes can reach energy levels at the island that lay well above the Fermi level E_F. Electrons on the island at energy levels below E_F can be thermally excited as well and leave the island. This causes a broadening of the conductance peaks, and they can smear out with rising temperature. To avoid this phenomenon, the coulombic energy of an electron at the island must be much greater than the thermal energy, i.e.

$$\frac{q^2}{2(C_{IS}+C_{ID}+C_{IG})} \gg k_B T \tag{5.22}$$

Additionally, in order to talk about one electron being localized on the island, the lifetime τ of the corresponding quantum mechanical state should not be vanishingly small. The life time is related with the energy difference ΔE between the states with N and $N+1$ electrons through the Heisenbergs uncertainty relation. For small biases, the tunneling rate through a potential barrier is linearly proportional to the applied bias and the barrier can be viewed as an electrical resistance. Hence, for the electrons to be localized, the equivalent tunneling resistance R_T should not be vanishingly small. If the resistance is too small, the wave functions for the electrons on the island overlap with the wave functions for the electrons on the source and drain contacts. The life time of one electron on the island can be estimated by the equivalent electrical time constant $R_T C_{TOT}$, where C_{TOT} is the total capacitance of the island. The energy difference ΔE between the states is given by the coulombic energy $E_C = q^2/2C_{TOT}$. Hence, using Heisenbergs uncertainty relation, the minimum tunneling R_T resistance can be estimated as

$$\tau \gg \frac{h}{\Delta E} \qquad \rightarrow \qquad R_T C_{TOT} \gg h\frac{2C_{TOT}}{q^2} \qquad \rightarrow \qquad R_T \gg \frac{2h}{q^2} \tag{5.23}$$

Equation 5.23 tells us that in order to observe the Coulomb blockade phenomenon the tunneling resistance R_T must be much greater than $2h/q^2$. In previous literature, however, the value $h/q^2 \approx 26k\Omega$ is most commonly used as reference value (Weis 1994; Nakajima and et al. 1994; Takahashi, Namatsu, Kurihara, Iwadate, Nagase, and Murase 1996; Welser, Tiwari, Rishton, Lee, and Lee 1997).

The last constrain imposed is that the drain source bias V_{DS} must be less than the separation between the charging energies for subsequent electrons placed on the island, where $V_{DS} < e/C_{TOT}$. If $V_{DS} \geq e/C_{TOT}$, there will always be at least one single electron energy level on the island that lays between the source and drain potential and Coulomb blockade can then not be achieved.

In conclusion, one can expect to observe Single Electron Transistor action if the following conditions are fulfilled:

$$\frac{q^2}{2C_{TOT}} \gg k_B T \tag{5.24}$$

$$R_T \gg \frac{2h}{q^2} \tag{5.25}$$

$$V_{DS} < \frac{q}{C_{TOT}} \tag{5.26}$$

The next step is to compare the experimental results with the predictions of the Coulomb blockade model. The most reasonable approach to this task seems to be modeling the electrical transport characteristics of the devices as being the result of two conduction mechanisms occurring in parallel: *a)* the conventional drift-diffusion transport as described in section 2.1.2 and *b)* Single Electron Transistor (SET) action. In the limit of zero current carried by the conventional drift-diffusion mechanism one would have a SET.

In order to realistically compute the current contributed by possible SET action the parameters will be estimated based on our experimental results. In this procedure a "best case" study will be done, where values as close as possible to the limiting conditions will be set to the parameters. Doing so, it will be possible to conclude if the phenomena observed in the experimental data may be originated by SET action or not.

The first step is to evaluate the values of the parameters to fulfill 5.24. As the oscillations can be observed at temperatures up to 35 K, the constraint on the total island capacitance C_{TOT} is found to be

$$\frac{q^2}{2C_{TOT}} \gg k_B T \qquad \rightarrow C_{TOT} \ll 38aF \tag{5.27}$$

meaning that the total island capacitance is required to be much less than 38 aF. To get an insight, this value corresponds to the self-capacitance of a metal island of radius $R \approx 30$ nm immersed in a material with a dielectric constant $\varepsilon = 11.9\varepsilon_0$, such as silicon. If capacitive coupling between the island and other electrodes does exists, the radius has to be smaller in order to get a capacitance not greater than 38 aF.

According to 5.21, in order to achieve gate voltage steps $\Delta V_G = 100mV$ between conductance peaks of, the typical ΔV_G found in the experimental results, the capacitive coupling C_{IG} between the island and the gate electrode must then be equal to

$$\Delta V_G = 0.1V = \frac{q}{C_{IG}} \rightarrow \qquad C_{IG} = 1.6aF \tag{5.28}$$

However, it should be noted that this can not be equal to the total capacitance of the island C_{TOT}. According to 5.24 and simulations, for an island with $C_{TOT} = 1.6aF$, SET action should be noticeable even well above room temperature and at drain source biases of the order of 100 mV. To bring theory and experiment into better agreement, capacitive coupling between the island and other electrodes, such as the source and drain electrodes, must be then assumed. The value of this capacitive coupling will be used as a fitting parameter to get the best possible agreement between experiment and the simulations described below.

In order to get the maximum drain source current I_{DS} and satisfying 5.23, the tunneling resistance R_T of the tunneling barriers between source and island and between drain and island will be set equal to $27k\Omega$ each.

These values are set for the parameters in order to simulate the electrical transport characteristics of the samples. The current component due to SET action is simulated using SIMON1.1 (Wasshuber 1997). It is important to note that this simulator implements a classical electron-electron interaction model and imposes no constrains in the tunneling resistance R_T. This can lead to an overestimation of the peak valley current relation, especially for smaller values of R_T. Quantum confinement energies due to the small dimensions of the island are also not modeled. This can become critical for very small islands, where the quantum confinement energy is not negligible when compared with the electrostatic energy. The component due to drift-diffusion transport is evaluated using DESSIS (ISE - Integrated Systems Engineering AG 1994).

Another experiment was carried out to verify if there is a capacitive coupling to a fourth electrode, such as the bulk. This was made by observing the peak frequency and the peak widths in the I_D *x* V_S characteristics, with V_{DS} kept constant. If V_G and V_B are kept constant while V_S and V_D are swept (but keeping V_{DS} constant), the gate voltages at which the conductance peaks occur are shifted. An increasing source bias shifts the gate voltage value V_G at which a peak occurs to a more positive value. This is a consequence of shifting the charging energies to more positive values, described by equation 5.11. Hence, a more positive gate voltage is needed to obtain the same bias between the island and the gate electrode, i. e., to induce the same amount of charge on the island[1]. With increasing V_S, at every conductance peak shown in Fig. 5.9, the number of electrons on the island is reduced by one.

Another effect that can be seen on Fig. 5.9 is the shifting the threshold voltages to more positive gate voltages with increasing source bias V_S. This classical phenomenon is boosted in very short channel devices such as the ones used in this study. With increasing source and drain potentials higher gate biases are needed to get the same amount of inversion at the channel ends.

If there were capacitive coupling to a fourth electrode, the oscillation period observed in the I_D *x* V_S characteristics should differ from that of the I_D *x* V_G characteristics. Namely, the oscillation period would be smaller in the I_D *x* V_S characteristics, because an extra term describing the capacitive coupling between the electrode and the island would appear in 5.20. In our measurements, as described in section 4.3, the oscillation period, the peak widths and heights measured in these traces are nearly the same as the ones obtained in the I_D *x* V_G characteristics, indicating that there is significant no capacitive coupling to a fourth electrode. Furthermore, neither in these experiments, nor in the I_D *x* V_D, nor in the V_G *x* I_D measurements, a change in the traces could be observed when source and drain are interchanged. Hence, it would be concluded that C_{IS} and C_{ID} should have the same value. This would be a very unusual coincidence, and may suggest that a model that accounts for fluctuations periodic in channel carrier number should be favored.

Simulation results for device 366/13-64 T4 are shown in Fig. 5.9. The value of C_{TOT} is used to bring experimental data and measurement into agreement. As one can see, the greater C_{TOT} the broader the conductance peaks are. A capacitive coupling between island and source $C_{IS} = 7.5aF$ and between island and drain $C_{ID} = 7.5aF$, amounting a total capacitance $C_{TOT} = C_{IG} + C_{ID} + C_{IS} = 16.6aF$, leads to the best possible agreement. The parameters for the conventional drift-diffusion simulator were determined using the process simulator DIOS (ISE - Integrated Systems Engineering AG 1994). The I_D *x* V_G characteristics for the same device is shown in Fig. 5.10. The

[1] Do not forget that there is an ohmic link between island and both source and the drain. Hence, in steady state and for vanishingly small V_{DS}, the difference between the potential at the island and the potential at source and at drain can not be larger than $\pm\frac{q}{2(C_{IS}+C_{ID}+C_{IG})}$

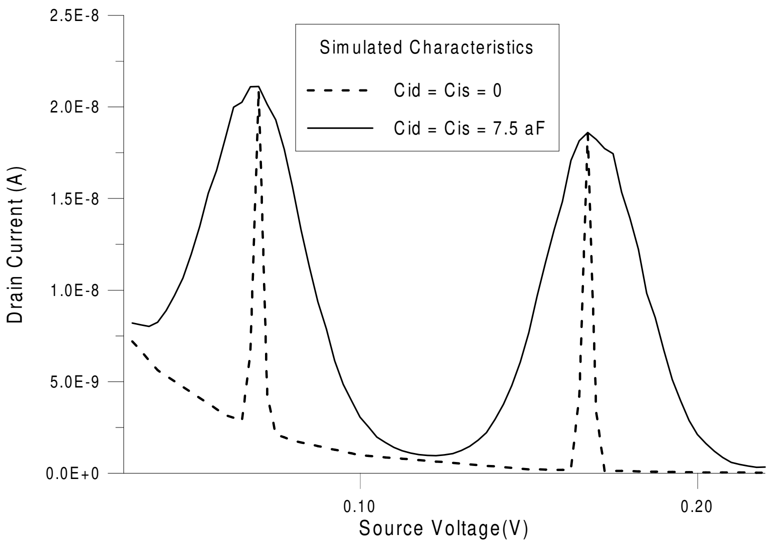


Figure 5.9: Simulation results showing the I_D x V_S characteristics for two different values of $C_{IS} = C_{ID}$ at 4.2 K. V_{DS}, V_G and V_B are kept constant while V_S is swept. The solid line is for $C_{IS} = C_{ID} = 7.5aF$. The dashed line for $C_{IS} = C_{ID} \approx 0aF$. C_{IG} is $1.6aF$ for both curves.

same values for C_{ID}, C_{IS}, and C_{IG} as above are used, and reasonable agreement can be obtained. This means that after this simulations the observed phenomena could be Coulomb blockade, but it is not a proof that it is Coulomb blockade.

As stated before, the quantum confinement energies due to the small dimensions of the island were not accounted for in the above calculations. In order to estimate the splitting between the single particle energy levels, a harmonic potential is assumed for a circular quantum dot. In the isotropic case the two-dimensional parabolic confinement potential is given by

$$V(x,y) = \frac{m^*}{2}\omega^2(x^2+y^2) \tag{5.29}$$

and the resulting eigenvalues are

$$E_{x,y} = (n_x + n_y + 1)\hbar\omega \tag{5.30}$$

This situation is schematically depicted in Fig. 5.11. As can be seen from Eq. 5.30, the separation between the single particle energy levels ΔE is $\hbar\omega$. As the channel length of the shortest devices is $L = 30nm$ and the oscillations are resolved at temperatures up to 35 K, one could estimate the single particle energy splittings assuming a quantum dot with radius $r = 15nm$ and the potential depth needed to accommodate approximately 10 electrons on the dot as being $10meV$. This leads to a splitting $\Delta E = \hbar\omega \approx 2.5meV$. This is comparable to the charging energy

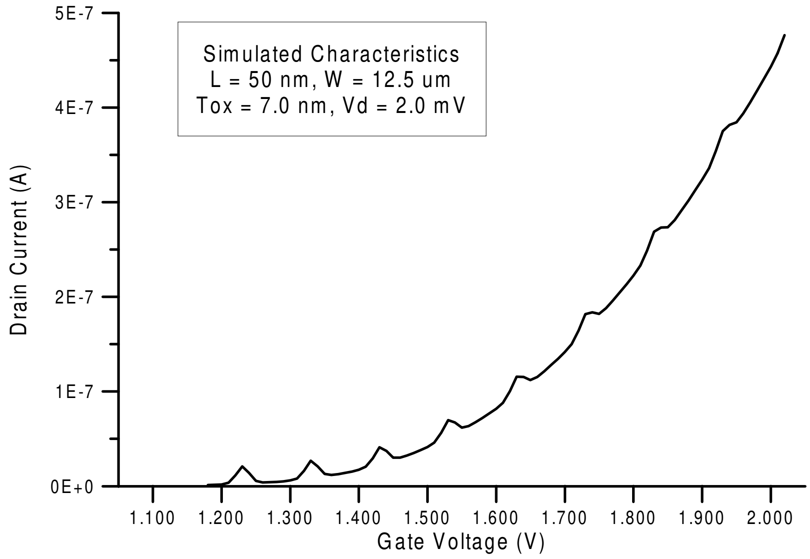


Figure 5.10: Simulation Results for the same device and conditions as for the experimental results in Fig. 4.1. For more details please refer to the text.

of the island $q^2/(2C_{TOT}) = 4.8meV$, for $C_{TOT} = 16.6aF$, as used in the above calculations. Hence, the quantum confinement energies must be accounted for.

The exact Hamiltonian of N electrons in a two-dimensional parabolic confinement potential is quite difficult to treat. But its properties have been studied in the scope of different approximation methods (Janik and Majkusiak 1998; Ishikuro and Hiramoto 1997; Nakazato and Ahmed 1995). In the *constant interaction picture* the interaction term is excluded from the Hamiltonian and treated by means of classical electrostatics, as in the analysis of the SET made above. The N-particle Hamiltonian is then the N-fold sum of the single-particle Hamiltonian plus the electrostatic energy $(Nq)^2/(2C_{TOT})$ of N electrons on the quantum dot. Hence, the total energy of the N-particle states is given by putting the N electrons one by one into the single-particle levels E_i and adding the classical electrostatic energy, where it is described as

$$E_{N,i} = \frac{(Nq)^2}{2C_{TOT}} + \sum_{i=1}^{N} E_i \tag{5.31}$$

When the second term above can be neglected, the islands are said to be metallic. As this does not hold in this case, the signature of the energy spectrum of the quantized states should be observed in the I_D *x* V_G characteristics (Schmidt 1997; McEuen, Foxman, Kinaret, Meirav, and Kastner 1992; Merkt, Huser, and Wagner 1991; Bryant 1987). Thus, the spacing between peaks should not be constant, but dependent on the degeneracy of the single-particle levels E_i.

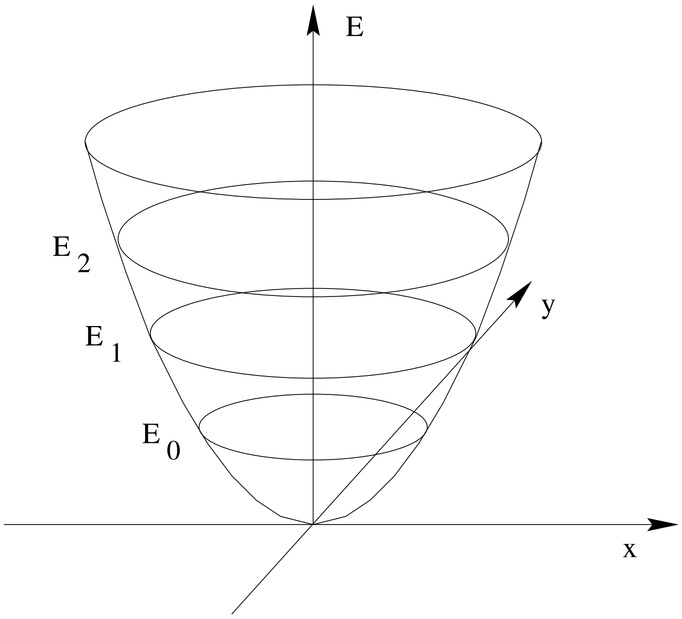


Figure 5.11: Single particle energy levels in a two dimensional harmonic confining potential.

The ground state is two-fold degenerate (spin degeneracy), the first excited state is four-fold degenerate, the second six-fold, and so on.

The application of a magnetic field perpendicular to the surface of the island allows the strength of the lateral confinement to be varied, and the energy of the single electron ground state to increase continuously with the magnetic field. The dependence of single electron energy levels of a harmonic oscillator on magnetic field is given by (Fock 1928)

$$E_{n,l} = (2n+1+|l|)\sqrt{(\hbar\omega_0)^2+(\hbar\omega_c)} \; + \; \hbar\omega_c l/2 \tag{5.32}$$

where $\omega_c = qB/m^*$ is the cyclotron frequency and ω_0 is the natural frequency of the harmonic oscillator. The Landau level separation $\hbar\omega_c$ is a measure of the strength of the confinement related to magnetic field. Setting $m^* = m_0$, the free electron rest mass, underestimates the Landau level separation, because the effective mass of the electrons in the silicon lattice is smaller than m_0. Doing so, a Landau level separation of about $1.75meV$ at a magnetic field of $15T$ is calculated; a value that is greater than the thermal energy k_BT for temperatures up to $20K$. Furthermore, the Landau level separation is then comparable to the charging energy of the island. Hence, it must be concluded that the application of the magnetic field changes the energy spectrum of the island, leading to a shift in the values of V_G at which Coulomb Blockade and Single Electron Transistor Action occur, that is, a change in the I_D x V_G characteristics is expected. In addition, the spin splitting of the single particles levels E_i is expected to become observable. As described in chapter 4, the pattern observed in the I_D x V_G characteristics does not change if magnetic fields up to 15 T are applied.

5.6 Coulomb Blockade in Parallel Dots

In the previous section it was concluded that the observed periodic oscillations are unlikely the result of a Coulomb blockade on a single quantum dot. In this section it will be discussed if the oscillations could be originated by a configuration where more than one quantum dot are present.

For the devices being studied, the channel width is much greater than the channel length ($W >> L$). Hence, if a multiple dot geometry is to be present in the channel of our devices, it is reasonable to expect that these will be in a parallel configuration. Furthermore, regarding a series configuration, one has to note that the current flow through the dots is decreased if they are connected in series. For some samples, the *difference* between the conductance values of a peak and the subsequent valley is more than e^2/h. This can not be accounted for by a series configuration.

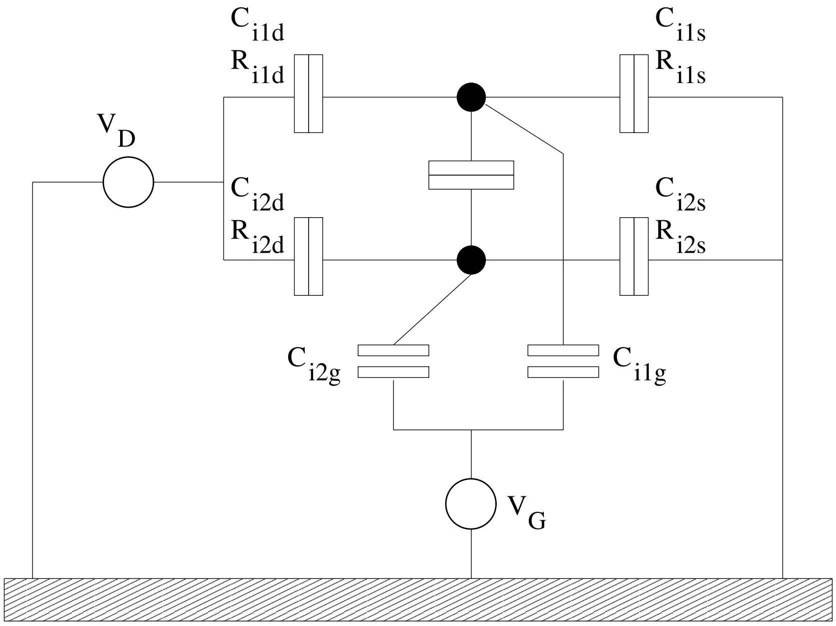


Figure 5.12: Coulomb blockade in a parallel geometry.

The system of capacitances and tunnel barriers of a possible parallel configuration with two islands in the device channel is shown in Fig. 5.12. It is assumed that the number of electrons on the first (N_{I1}) and second island (N_{I2}) are independently quantized. This assumption depends critically upon the tunnel barriers between the islands, and upon the tunnel barriers between the islands and the source and drain regions. The charging energy of such a system can be expressed in terms of the capacitance matrix $C^{-1}_{i,j}$ (Hofmann and Wharam 1996)

$$H_{ch} = \frac{1}{2}\sum_{i,j} Q_i C^{-1}_{i,j} Q_j \tag{5.33}$$

where Q_i is the total charge on island i.

Within the framework of this electrostatic description, the Coulomb blockade of current transport is maintained if the exchange of an electron between the dots and the surrounding reservoirs does not yield a configuration with the same energy. The problem is then reduced to finding the stable configurations (N_{I1}, N_{I2}) for which the total electrostatic energy given by 5.33 is a minimum. The points of degeneracy between stable configurations then yield the points where the blockade is lifted, and the exchange of one electron between the islands and source or drain reservoirs is favored. This can be schematically depicted in a so called stability diagram. In such a diagram, each cell corresponds to a stable configuration (N_{I1}, N_{I2}). At the cell boundaries the states are degenerate and a change in the total number of electrons $(N_{I1} + N_{I2})$ is then expected. At this point, the Coulomb blockade is lifted.

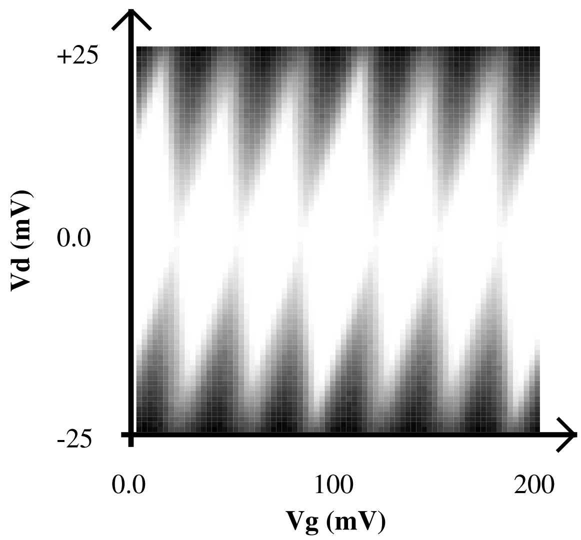


Figure 5.13: Stability diagram. Different diamonds and lines running parallel to the diamond edges are clearly visible. The darkness is proportional to the amount of tunneling going on in the device. In the white diamonds the device is in a stable configuration (Coulomb-blockade), and current flow is negligible. The diamond boarders correspond to the onset of single electron tunneling, where current starts flowing.

Figure 5.13 depicts such a stability diagram, simulated using Simon 2.0 (Wasshuber and Kosina 1998). An island-gate capacitance C_{i1g} of 1.6 aF is assumed for the first island. For the second island, island-gate capacitance C_{i2g} of 3.2 aF is assumed. The capacitive coupling between the islands is assumed to be 10 aF, and $C_{i1d} = C_{i2d} = C_{i1s} = C_{i2s} = 2.0aF$ is assumed (refer to Fig. 5.12).

From this stability diagram it can be seen that there are separate states for every stable configuration (N_{I1}, N_{I2}). Every white diamond like cell represents such a state. In this diagram, the darkness is proportional to the amount of tunneling going on in the device. In the white diamonds the device is in a stable configuration (N_{I1}, N_{I2}), corresponding to Coulomb blockade, and current flow is then negligible. On the diamond boarders the states are degenerate, corresponding to the onset of single electron tunneling, where current begins to flow.

In this simulation, the first diamond, centered around ($V_D = 0$, $V_G = 0$) is assumed to represent the stable state ($N_{I1} = n_1$, $N_{I2} = n_2$). As V_G increases, more electrons are induced on the island. The second diamond represents the stable state ($N_{I1} = n_1$, $N_{I2} = n_2 + 1$). The third stable state ($N_{I1} = n_1$, $N_{I2} = n_2 + 2$), and the next stable state ($N_{I1} = n_1 + 1$, $N_{I2} = n_2 + 2$), and so on. There are separate stable states for every configuration (N_{I1}, N_{I2}). This is true as long as there is a capacitive coupling between the two islands. As a consequence, peak overlaping is only possible if there is no capacitive coupling between the islands. In those cases, there is a separate stability diagram for each island, and the points where Coulomb blockade is lifted may then overlap.

This is illustrated in Fig. 5.14, where the characteristics expected for two non-interacting islands is depicted. An island-gate capacitance C_{i1g} of 1.6 aF is assumed for the first island. This leads to the characteristics already evaluated in the last section (i.e., an oscillation period of 100 mV in V_G). For the second island, island-gate capacitance C_{i2g} of 3.2 aF is assumed, leading to an oscillation period of 50 mV in V_G. For non-interacting islands, one can just arithmetically add the characteristics predicted for each island, and the peaks may overlap, as shown in Fig. 5.14.

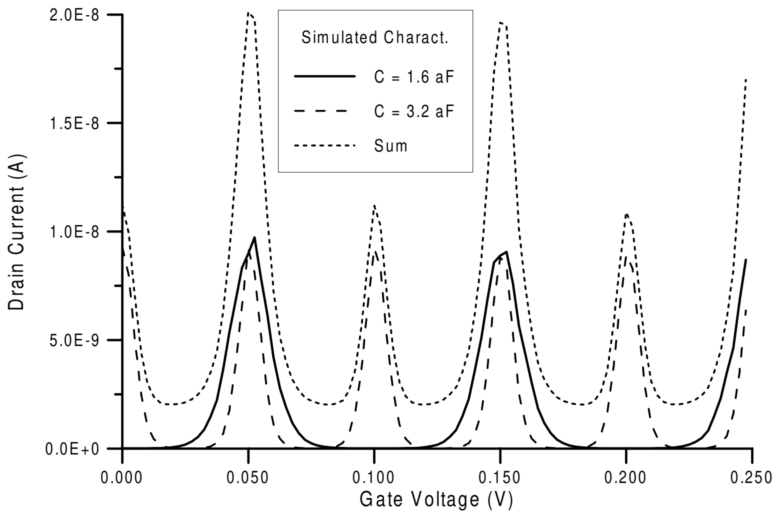


Figure 5.14: Simulated characteristics for two quantum dots that *do not interact*. Two islands are assumed to be in parallel in the channel. One with a gate island capacitance of 1.6 aF, and the other with a gate island capacitance of 3.2 aF. As there is no capacitive coupling between the islands, the resulting characteristics (doted line) is the superposition (sum) of the contribution of each. For better visualization, the resulting characteristics (dotted line) is slightly shifted in the current axis.

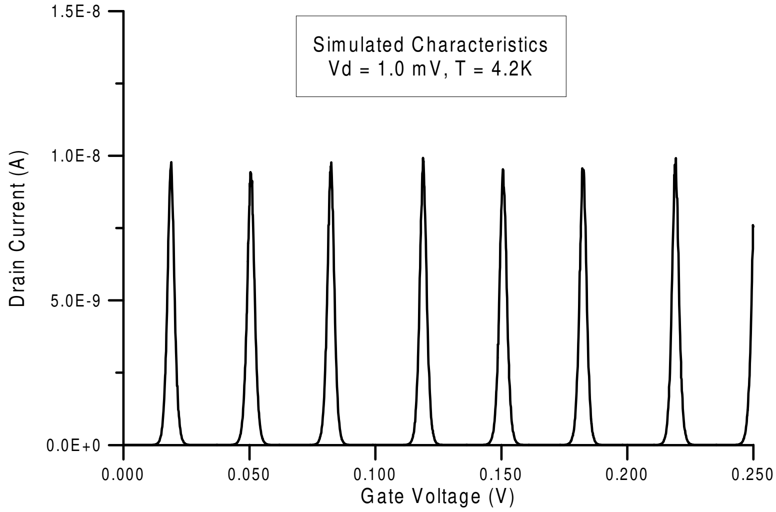


Figure 5.15: Simulated characteristics for two quantum dots that *do interact*. Two islands are assumed to be in parallel in the channel. One with a gate island capacitance of 1.6 aF, and the other with a gate island capacitance of 3.2 aF. The capacitive coupling between the islands is considered to be 10 aF. The geometry is depicted in Fig. 5.12

If the islands are now assumed to be capacitively coupled, as shown in Fig. 5.12, the expected characteristics will change significantly. Due to the interaction between the islands, the points where the Coulomb blockade is lifted will move. As shown in Fig. 5.15, this can lead to an oscillation period of $\approx 33mV$ in V_G, resembling experiment for the oscillations observed near threshold.

The central point in the interpretation of the results in the framework of Coulomb blockade in parallel quantum dots is the existence of confined electron gases (islands) that are electrostatically coupled to each other and to the source and drain reservoirs. As previously described, two parallel islands can account for the pattern measured near threshold if specific values are assigned to the model parameters. The tunnel barriers between islands and the source and drain reservoirs may not be independent of the gate bias. Indeed, if these barriers are due to the inhomogeneous potential landscape in the channel, one should expect the potential barriers to decrease as the level of inversion increases, where V_G increases. The potential barriers of the two islands may also depend on V_G in different ways. One of the barriers may be lowered faster than the others. It could, for instance, be reasonable to expect that the resistance of the barriers for tunneling into the greater island (the one with greater capacitance) will reach the critical value for electron localization first. In this regime one can not speak of an isolated island anymore, since the electron wave functions start being delocalized, and the exchange of an electron between the second dot and the source and/or drain reservoirs costs almost no energy. At this point the second island merges into the channel and only the first island is left. As the capacitive coupling between the first island and the gate electrode was assumed to be $C_{I1G} = 1.6aF$, the oscillation period will change from 33 mV to 100 mV, matching the results from experiment.

If the gate voltage is increased further, the first island will also start merging into the channel. This result is the suspected mechanism responsible for the smearing out of the oscillations at the highest gate voltages.

As discussed above, the current contributed by two quantum dots can be arithmetically added only if there is no interaction between these dots, where the interdot capacitance is vanishingly small. If the interdot capacitance can not be neglected, two separate conductance peaks will become visible. Hence, the conductance peaks of two different quantum dots can overlap only if the corresponding interdot capacitance is vanishingly small, that is, if they are electrostatically decoupled. To match experimental results, a mechanism that can account for a difference in conductance between a peak and the subsequent valley greater than $q^2/(2h)$, found in some samples, is required. The presence of a third island, decoupled from both other islands, is needed. Subsequently, according to the Coulomb blockade model, in some samples the existence of a third island, decoupled from both other islands, is predicted.

Similar to Coulomb blockade in a single dot, discussed in the previous section, the application of a magnetic field perpendicular to the islands allows the strength of the lateral confinement do be varied, causing the energy spectrum of the islands to change. Hence, a change in the measured characteristics would be expected if a magnetic field is applied. As described in chapter 4, this could not be observed.

6. The Random Telegraph Signal

In this chapter experimental results that indicate the possibility of observing the behavior of individual defect sites near the Si/SiO_2 interface are presented. Studying the bias-voltage dependence of the random telegraph signal (RTS) it is possible to determine the location of these defects. It is shown that a trap can be used as a probe into the local surface potential. It is also shown that as the behavior of the RTS does depend on the properties of the channel electrons near the defect, RTS analysis is a valuable way to study effects as Coulomb scattering, electron gas heating and electrical channel formation in very small area devices.

It is also indicated that the fluctuations in the drain current, related to the trap, tend to increase as the device active area decreases. As a result, RTS is of increasing relevance as device dimensions keep shrinking and supply voltages continue being lowered.

6.1 Single-Electron Switching

The continuous switching of a signal between two well defined levels is called Random Telegraph Signal Noise, or simply RTS Noise. Fig. 6.1 shows an example of such a signal in the drain current I_D of a MOSFET. During the measurement the device is held at a fixed bias point and the current is observed as a function of time. The switching in I_D is related to the alternate capture and emission of carriers at an individual defect site.

Small area devices as the ones studied in this work are particularly well suited for the study of the RTS Noise. In large area devices there may be many defect sites affecting the electrical transport at every bias point, making it impossible to study the behavior of a single one. Furthermore, the relative amplitude of the signal tends to decrease with increasing active device area, imposing practical difficulties in measuring the signal with sufficient accuracy.

6.1.1 Capture and Emission Kinetics

In MOSFETs most of the defects are found to reside in the oxide close to the Si/SiO_2 interface, within tunneling distance of the inversion layer. In a nMOSFET, if one electron tunnels into a neutral defect (electron capture), the number of mobile carriers in the inversion layer is reduced. Furthermore, the defect is then charged, and scattering of inversion layer carriers may increase. Therefore, a reduction in the channel current can be expected. After the release of the trapped electron (emission) the current returns to its higher level. Hence, the mean time at the low and at the high current levels are directly related to the electron emission τ_e and capture τ_c times, respectively. τ_c is the mean time needed to capture an electron, that is, the mean time during

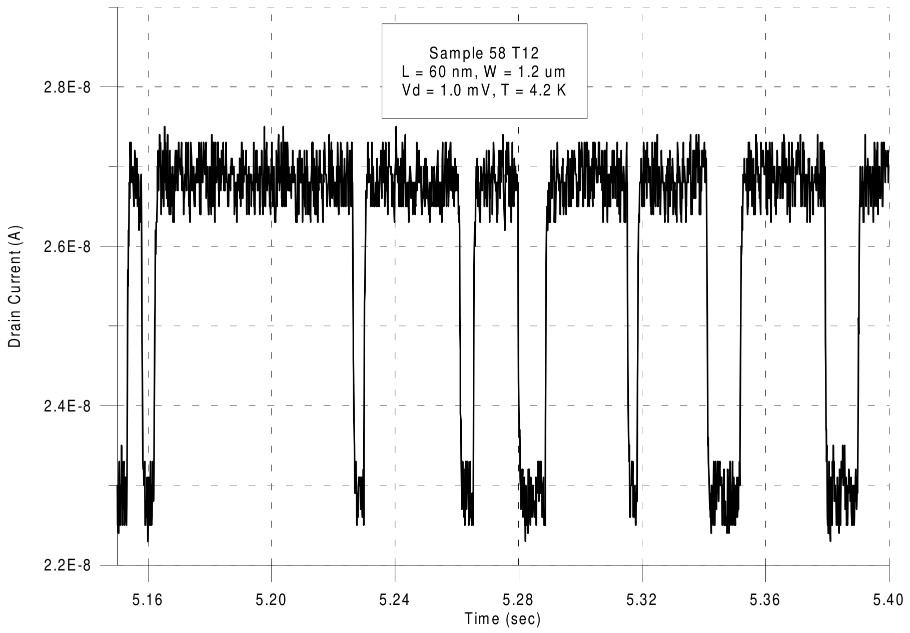


Figure 6.1: A typical Random Telegraph Signal (RTS) in the drain current of a MOSFET. Discrete current changes against time.

which the trap is empty, and τ_e is the mean time needed to emit it, that is., the mean time during which the trap is occupied by the electron.

Observing the gate voltage dependence of the RTS shown in Fig. 6.4, it can be seen that as the gate voltage is increased the mean time in the high-current state (τ_c) is strongly reduced. This can be understood with the help of Fig. 6.2. It shows the band bending in the channel of a MOSFET in which there is only one defect energetically reachable for the electrons in the inversion layer, that is, there is only one defect energy level E_T within k_BT of the surface Fermi level E_F. The dotted lines and the arrows sketch the effect of a positive increment in the gate voltage V_G. With increasing gate voltage the energy separation $E_T - E_F$ becomes smaller, and the probability of the electron being captured increases. This corresponds to a decrease in the high-current state time.

In the linear regime of operation the fractional occupancy of the defect is given by (Kirton and Uren 1989b)

$$\frac{\tau_c}{\tau_e} = g\, e^{\frac{E_T - E_F}{k_B T}} \tag{6.1}$$

where g is the ratio between the degeneracies $\gamma(n)$ and $\gamma(n+1)$ of the $n-$ and $(n+1)-$electron states, respectively (i.e., $g = \gamma(n)/\gamma(n+1)$. In most cases $\gamma(n)$ and $\gamma(n+1)$ are equal ($g = 1$).

It can be seem from Fig. 6.2 that the dependence of $(E_T - E_F)$ on V_G can be used to estimate the distance of the defect from the interface. If the defect is located at the interface, the change in

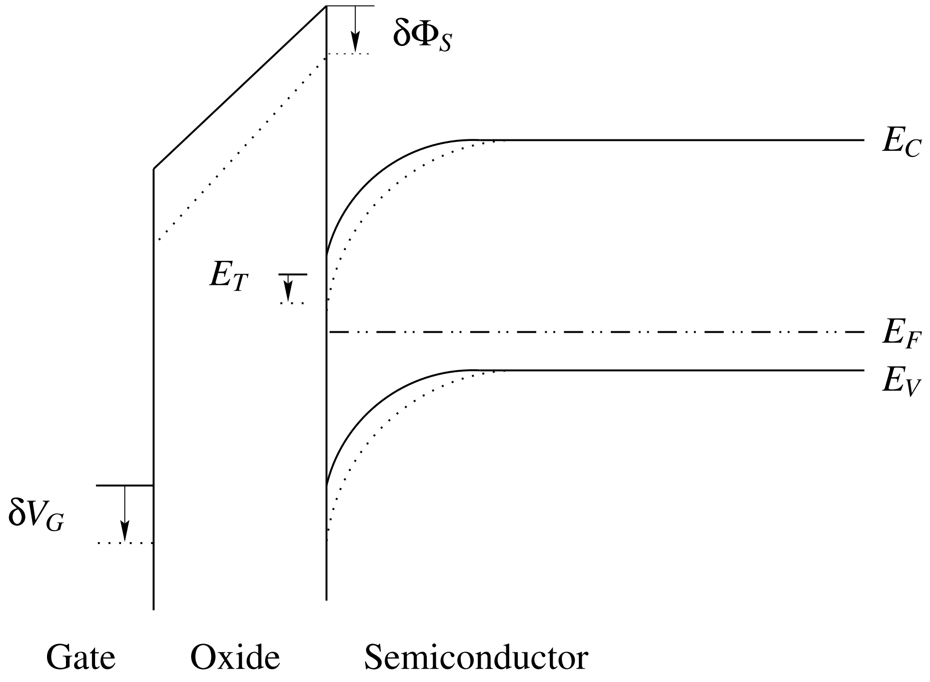


Figure 6.2: Band diagram of a nMOSFET with a localized energy state (trap) in the oxide. The dotted lines show the changes due to an increment δV_G in the gate voltage. $\delta\Phi_S$ is the change in the surface potential. The trap energy level E_T is pushed down by incrementing the gate voltage.

$(E_T - E_F)$ equals the change in the surface potential $\delta\Phi_S$. If the defect is at a distance d into the oxide, the potential drop due to the electrical field in the oxide must be accounted for. Assuming a constant electrical field in the oxide, the distance d of the trap into the oxide is then given by

$$\frac{\delta V_G - \delta\Phi_S}{t_{ox}} = \frac{\delta(E_T - E_F)/q - \delta\Phi_S}{d} \tag{6.2}$$

The surface potential change $\delta\Phi_S$ can be easily estimated from the standard MOSFET theory presented in chapter 2.

Taking the logarithm of both sides of eq. 6.1 and differentiating with respect to V_G, leads to

$$\frac{d}{dV_G}(E_T - E_F) = k_B T \frac{d}{dV_G}\left(ln\frac{\tau_c}{\tau_e}\right) \tag{6.3}$$

Since the ratio $(\delta ln\tau_c/\tau_e)/(\delta V_G)$ can be experimentally determined, equations 6.3 and 6.2 can be used to estimate d.

The above equations describe the fractional occupancy of the defect (the ratio τ_c/τ_e) and its dependence on V_G and d, but give no information about τ_c and τ_e itself. Now the theory necessary to estimate them will be discussed.

The capture and emission processes are assumed to be thermally activated (Kirton and Uren 1989a). First, the physical origin of the non-radiative transition is investigated. This is done by using the simplified configuration coordinate diagram of Fig. 6.3. It assumes coupling with one lattice coordinate Q. Due to the presence of electron-lattice interaction in the defect (bound) state, the defect electronic energy is minimum for a value $Q = Q_d$ differing from $Q = Q_k$, the Q value at which the delocalized electron energy is a minimum. In the perfect lattice configuration the electron-lattice coupling is infinitely small and $Q_k = 0$ (in Fig. 6.3 this is assumed to be the case). Assuming that the system is initially in the state $E_k(Q)$ corresponding to the empty defect with the electron in the conduction band, it can be shown that the non-radiative transition is induced by off-diagonal elements in the Hamiltonian, which induce transitions between vibronic states that differ in electronic energy but have the same total energy (Bourgoin and Lannoo 1983). In other words, the transition probability is a maximum at the cross-over of curves $E_d(Q)$ and $E_k(Q)$ in Fig. 6.3, where there is a strong mixing between the inversion-layer state and the defect state. As the cross-over does not occur at the minimum of $E_k(Q)$, there is a energy barrier for capture E_B (see Fig. 6.3). One electron in the conduction band can absorb thermal energy and overcome the energy barrier E_B. Hence, the capture process is thermally activated. On electron capture the defect bound state is well away from equilibrium and the excess energy is dissipated by multi-phonon emission.

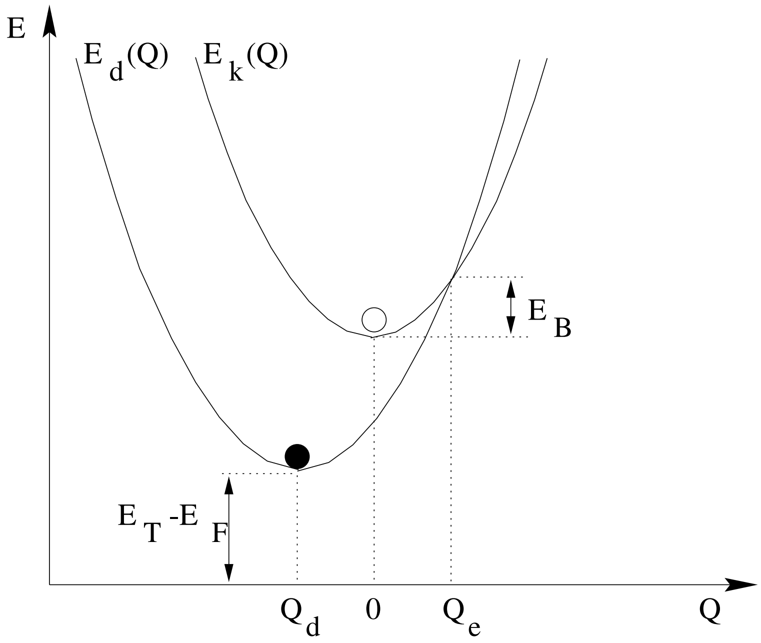


Figure 6.3: Configuration coordinate diagram for an acceptor trap. The energy E is the sum of the elastic and electronic energies. It is plotted against one lattice coordinate Q. • labels the localized defect electronic state. ○ labels the trap empty with the electron free to move in the conduction band.

The capture rate can be written as a function of the particle flux and cross-section (Kirton and Uren 1989a):

$$\frac{1}{\tau_c} = \int_{E_C}^{\infty} n(E)v(E)\sigma(E)dE \tag{6.4}$$

where $n(E)$ is the inversion-layer carrier density, $v(E)$ the carrier velocity and $\sigma(E)$ the cross-section. This equation can be simplified if one assumes a constant carrier density n with an average thermal velocity $\overline{v}$ in the inversion-layer. By using the average capture cross-section σ, the capture rate reads

$$\frac{1}{\tau_c} = n\overline{v}\sigma \tag{6.5}$$

The capture cross-section is found to be of the form (Schulz and Johnson 1978)

$$\sigma = \sigma_0 e^{\frac{-E_B}{k_B T}} \tag{6.6}$$

Substituting eq. 6.6 in eq. 6.5 leads to

$$\tau_c = \frac{e^{\frac{E_B}{k_B T}}}{\sigma_0 n \overline{v}} \tag{6.7}$$

The inversion-layer carrier density n is given by

$$n = N_C e^{\frac{-(E_C - E_F)}{k_B T}} \tag{6.8}$$

where N_C is the effective density of states in the conduction band. Now the electron emission time τ_e can be obtained by combining eq. 6.1, 6.8 and 6.7:

$$\tau_e = \frac{1}{g\sigma_0 \overline{v} N_C} e^{\frac{E_B + (E_C - E_T)}{k_B T}} \tag{6.9}$$

An electron at the conduction band edge E_C can tunnel into a trap lying at a higher energy E_T by phonon absorption. On electron capture the defect bound state is well away from equilibrium and the excess energy is dissipated by multi-phonon emission (Kirton and Uren 1989a). That means that the defect state lies at a higher energy when it is empty as when it is filled. Hence, the filled trap state may lie below E_C and the electron has to absorb thermal energy from the lattice in the emission process.

6.1.2 The Amplitude of RTS

The trapping of an electron into a localized defect state can induce a change in source-drain current in two ways: by a reduction in the number of free carriers in the channel and by a change in the effective mobility (Hung, Ko, Hu, and Cheng 1990; Pacelli, Villa, Lacaita, and Peron 1999; Villa, Lacaita, Perron, and Bez 1998). In a first step the signal amplitude will be estimated as the result of the individual contribution of two terms: one due to the charge number fluctuation, and the other due to mobility change. By following this approach the relative change in source-drain current can be described as

$$\frac{\Delta I_D}{I_D} = \frac{\Delta N_{inv}}{N_{inv}} + \frac{\Delta\mu}{\mu} \tag{6.10}$$

where N_{inv} is the inversion layer charge density (per unit area).

First the contribution due to a change in the number of channel free carriers (number change) will be evaluated. When the trapped charge density N_{trap} fluctuates at fixed V_G, charge conservation in the device can be written as

$$\Delta N_{inv} + \Delta N_{dep} + \Delta N_{trap} + \Delta N_G = 0 \tag{6.11}$$

where N_{dep} is the depletion region charge density and N_G is the charge density at the gate electrode.

The charge density fluctuations can be related to the change in surface potential $\Delta\Phi_S$ observing that

$$\Delta N_{inv} = -C_{inv}\Delta\Phi_S \tag{6.12}$$

$$\Delta N_{dep} = -C_{dep}\Delta\Phi_S \tag{6.13}$$

$$\Delta N_G = -C_{ox}\Delta\Phi_S \tag{6.14}$$

where C_{ox}, C_{dep} and C_{inv} are the capacitances (per unit area) associated with the oxide, depletion region and inversion layer, respectively.

Inserting 6.12 to 6.14 into 6.11 leads to

$$\frac{\Delta N_{trap}}{\Delta\Phi_S} = C_{inv} + C_{dep} + C_{ox} \tag{6.15}$$

This expression can be developed further using equations 6.12 to 6.14 and using the relation $\Delta N_{inv}/\Delta N_{trap} = (\Delta N_{inv}/\Delta\Phi_S)/(\Delta\Phi_S/\Delta N_{trap})$:

$$\frac{\Delta N_{inv}}{\Delta N_{trap}} = \frac{\Delta N_{inv}/\Delta\Phi_S}{C_{dep} + C_{ox} - (\Delta N_{inv}/\Delta\Phi_S)} \tag{6.16}$$

As discussed in chapter 2, section 2.1.1, once strong inversion is achieved the surface potential Φ_S gets virtually pinned to $2\Phi_B$ (Φ_B is the potential in the bulk). Hence, in strong inversion the denominator in 6.16 is dominated by $\Delta N_{inv}/\Delta\Phi_S$, and $\Delta N_{inv}/\Delta N_{trap}$ turns out to be -1. If one electron is trapped the inversion layer looses one electron. In weak inversion charge sharing between the gate, the inversion layer and the depletion region does occur. Part of the trapped charge is neutralized by a change in the depletion region charge density, or by a change on the charge on the gate electrode, and one gets $|\Delta N_{inv}/\Delta N_{trap}| < 1$.

After the charge-sheet model of the MOSFET developed in chapter 2, the inversion layer charge density can be written as

$$N_{inv} = N_0\, e^{\frac{q\alpha\Phi_S}{k_BT}} \tag{6.17}$$

and therefore

$$\frac{\Delta N_{inv}}{\Delta\Phi_S} = \frac{q\alpha\Phi_S}{k_BT} N_{inv} \tag{6.18}$$

where α has a value of 1 in weak inversion and decreases monotonically down to 0.5 in strong inversion (Brews 1978).

Finally, substituting 6.18 into 6.16, the ratio $\Delta N_{inv}/N_{inv}$ can be estimated to be

$$\frac{\Delta N_{inv}}{N_{inv}} = \frac{\alpha(q/k_BT)\Delta N_{trap}}{C_{dep}+C_{ox}-\alpha(q/k_BT)N_{inv}} \tag{6.19}$$

Now, that the equations to evaluate the first term at the right side of 6.10 have been developed, the term due to mobility change has still to be analyzed.

The mobility of the carriers is a function of the average number of scattering events that occur in the path between source and drain. If the scattering events at the trap under observation and those at the other scattering centers are uncorrelated, the resulting total mobility μ can be evaluated using Matthiessen's rule as being

$$\frac{1}{\mu} = \frac{1}{\mu_{ch}} + \frac{1}{\mu_{trap}} \tag{6.20}$$

where μ_{trap} is the mobility term due to the trap under observation only and μ_{ch} is the mobility due to all other scattering events that occur in the channel. The mobility term μ_{trap} can be related to the effective mass $1/m^*$ and scattering time τ_{trap} by

$$\frac{1}{\mu_{trap}} = \frac{m^*}{q\tau_{trap}} = \frac{m^*\overline{v}\sigma}{qWL} \tag{6.21}$$

where σ is the scattering cross-section and $\overline{v}$ the average thermal velocity. The factor $1/WL$ is the density of scatterers, for one trap and a channel of length L and width W.

The above description treats the effects of number and mobility fluctuations as being distributed over the whole channel. All values are taken as mean values integrated over the channel area. This uniform number and mobility fluctuation model is the standard procedure used to study RTS in MOS devices.

6.2 Discussion

In this section the RTS measured in nanometer scale MOS devices is described and the adequacy of the uniform number and mobility fluctuation model for this small devices is discussed.

The random telegraph signal was measured using a HP4156A semiconductor parameter analyzer. Since the capture and emission times are exponentially distributed, in accordance with a Poisson distribution, a large number of transitions must be recorded (Kirton, Uren, Collins, Schulz, Karmann, and Scheffer 1989). To obtain an estimate of the up and down times with an error of 10% or less, records containing 200 or more transitions are stored. Additionally, the sampling time is held short. Otherwise transitions would be lost. During measurements, the sampling rate used was, always when possible, of at least 100 times the capture or emission time. This makes data acquisition a very time consuming task and presents a practical limitation on the number of samples that can be analyzed. Since the minimum sampling time of the measurement apparatus used is $60\mu sec.$, for capture or emission times shorter than $6msec.$ it was not possible to satisfy the last constraint. Hence, for those samples the accuracy is worse than 10%. The statistical and experimental errors are always indicated in the figures that present the data.

6.2.1 Capture and Emission Times as a Function of Bias Point

This section describes first the gate voltage V_G dependence of the measured RTS. The data shown in figures 6.4 and 6.5 are representative for the random telegraph signal noise found in devices at room and low temperatures, respectively.

Figure 6.4 shows that as the gate voltage is increased while V_D is kept constant, the capture time is strongly reduced, while the emission time happens to remain largely unaffected. This is in accordance with equations 6.7 and 6.9. A positive increase in the gate voltage does induce more free carriers in the inversion layer, where n in Eq. 6.7 does increase. Hence, a decrease in τ_c becomes expected. Although the change in capture time on increasing gate voltage is predominantly a result from changes in n, the decrease in τ_c can be boosted further by other effects. An increasing V_G means increasing carrier velocity at the Fermi level, and one may expect an increasing average thermal velocity $\bar{v}$, depending on temperature and drain bias. In addition, as discussed in chapter 2, section 2.3, as the transversal electric field strength increases, the inversion layer charge-density peak moves closer to the interface, thus increasing wave-function overlap. The changing oxide field strength also lowers the tunneling barrier. The net result is an increase in σ_0 in Eq. 6.7, contributing to the decrease in τ_c. This factors may also affect the emission time and could be the origin of the small variations in τ_e that can be seen in Fig. 6.4. However, since these variations are not accentuated, it is not possible to say if they are a real effect or if they are an artifact of measurement accuracy.

The data in Fig. 6.5 shows some deviation from the behavior just discussed. The major difference is that for this trap, active at low temperature, the emission time does increase with increasing gate bias. Figure 6.2 shows that a change in the gate voltage of δV_G causes a change in surface potential of $\delta\Phi_S$, consistent with a more negative charge in the inversion layer. Thus, the electron states in the oxide will be shifted down in energy, as shown schematically for the trap state E_T. Hence, the energy needed to emit an electron from the trap state into the inversion layer does increase with gate bias. This is the origin of the increase in τ_e observed with increasing gate bias in Fig. 6.5.

Now one could ask why this is not observed for the room temperature data presented in Fig. 6.4. One hint can be found looking to the difference in the magnitude of the time scales (*y-axis*) in figures 6.4 and 6.5. As one can see, the capture and emission times for the room temperature data are on a scale two orders of magnitude greater. Equation 6.7 tells us that in order to get a longer emission time at higher temperature, the energy barrier for electron emission has to be much higher. This is the reason why τ_e is quite insensitive to changes in the gate bias for the trap observed at room temperature. For this trap at room temperature, the changes in energy for the electron states in the oxide are not significant compared to the energy barrier for electron emission.

The drain bias dependence of the RTS in our probes appears to be much more exciting, and deserves more careful analysis. In the following the dependence of the mean capture and

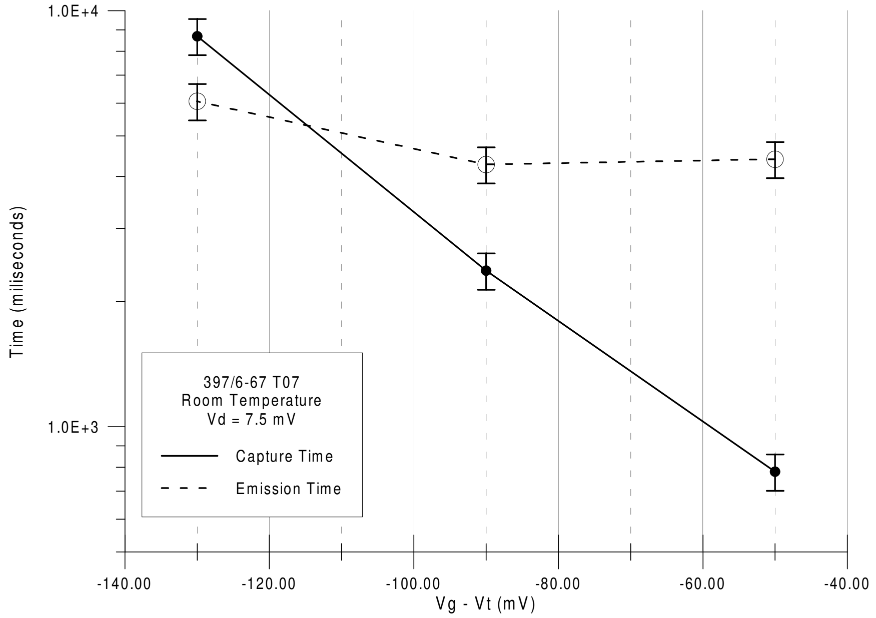


Figure 6.4: Mean lifetime for capture (solid symbols) and emission (open symbols) as a function of gate bias at room temperature. Threshold voltage V_T is about 50 mV. Design channel width and length are $W = 1.0\mu m$ and $L = 30nm$, respectively. Gate oxide thickness t_{ox} is 4.7 nm.

emission times on V_D is discussed. The drain bias dependence of the RTS amplitude is treated in the next section.

Since MOSFETs are macroscopically symmetrical devices, the source and drain biases can be interchanged so that a single trap can be used as a probe near the drain in one orientation and near the source in the other. These orientations are arbitrarily called forward and reverse in the following discussion.

In analyzing the mean capture time as given by Eq. 6.7, if the electron gas is not in thermal equilibrium with the lattice, the first question that arises is whether T should be replaced by T_e, the average electron temperature, or not. As discussed in section 6.1.1 and shown in Fig. 6.3, in order to reach the cross-over of curves $E_d(Q)$ and $E_k(Q)$, a defect must absorb energy from the lattice. The capture time is expected to be shorter than the optical phonon relaxation time, and optical phonons emitted by hot electrons would provide the energy necessary for reaching the cross-over (Shi, Miéville, and Dutoit 1993). However, since the energies of optical phonons in silicon are higher than 50 meV (Johnson 1959; Strauch, Mayer, and Dorner 1990), this effect is not expected to be relevant at the drain biases used in this work. Hence, the equilibrium lattice temperature T is used in Eq. 6.7.

The second question is regarding which value for the average thermal velocity $\overline{v}$ of the electrons should be used. If the electron gas is in thermal equilibrium with the lattice, $\overline{v}$ obviously follows the lattice temperature. But this assumption will fail if the electrons are heated by the

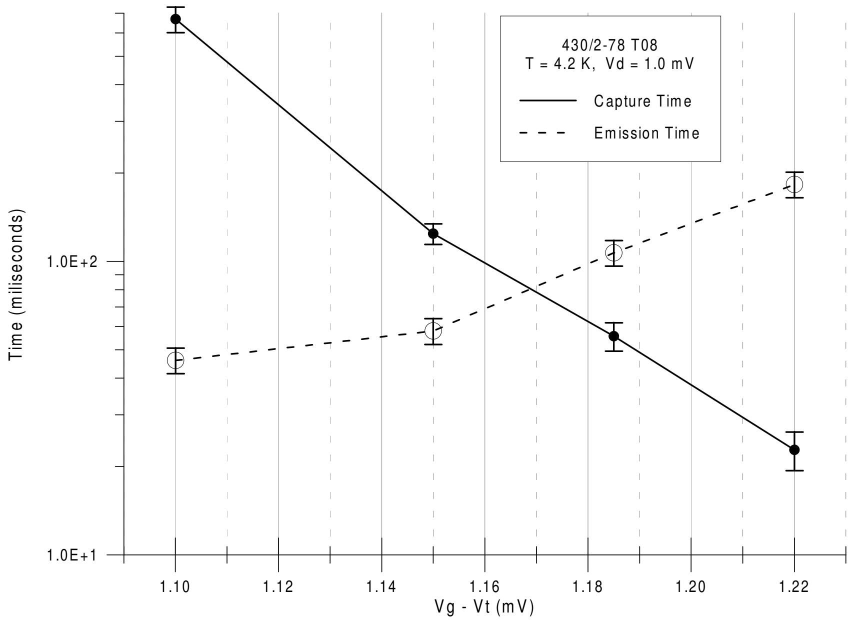


Figure 6.5: Mean lifetime for capture (solid symbols) and emission (open symbols) as a function of gate bias at 4.2 K. Threshold voltage V_T is approximately 1.1 V. Design channel width and length are $W = 0.8\mu m$ and $L = 80nm$, respectively. Gate oxide thickness t_{ox} is 4.0 nm.

electric field, which can be the case even for small drain biases in sub-100nm MOSFETs, specially at low temperatures. Under these conditions, $\overline{v}$ depends on the drain bias V_D, and τ_c is expected to decrease with increasing V_D.

In order to develop a quantitative approach for carrier heating due to the longitudinal electric field, the evaluation of the relative importance of elastic and inelastic scattering in the inversion layer is needed.

The design channel lengths of the devices under observation in this chapter are 80 nm, 50 nm or 30 nm. For the devices with a channel length of 80 nm there was no low energy ion implantation to adjust the threshold voltage, leading to a relatively low dopant concentration at the channel surface. Hence, most of the electrons are expected to travel from source to drain without suffering inelastic scattering at low temperature, as long as the drain bias is kept small. Elastic scattering does randomize momentum but does not change the kinetic energy. This is then related to the electrostatic potential along the channel. In a first order approximation, it can be assumed that the kinetic energy of inversion layer electrons is proportional to the longitudinal electrical field multiplied by the relative position from the source, and to the equilibrium lattice temperature T_l. In other words, the electrons are assumed to be in thermal equilibrium with the lattice at the source and then accelerated towards the drain by the longitudinal electrical field, and that elastic scattering does prevail. This leads to the following expression for kinetic energy:

$$E_k(l) = k_B T_l + qV_D \frac{l}{L} \tag{6.22}$$

where (l/L) is the fraction of the drain-source potential which appears between the trap and the source. If the trap is at the source side of the channel $(l/L) \approx 0$. At the drain side $(l/L) \approx 1$.

It follows that the electron velocity is then simply given by

$$\overline{v} = \sqrt{\frac{2E_k}{m^*}} = \sqrt{\frac{2(k_B T_l + qV_D \frac{l}{L})}{m^*}} \tag{6.23}$$

At higher temperatures and higher drain biases, the electron velocity in the channel can be better calculated as being (Taniguchi and Mamaguchi 1991)

$$\overline{v} = \sqrt{v_{th}^2 \frac{T_e}{T_l} + v_d^2} \tag{6.24}$$

where T_e and T_l are the electron and lattice temperatures, respectively, and v_d is the electron drift velocity as a function of the longitudinal electrical field E_l, that is, $v_d = \mu(E_l)E_l$.

The fact that the electrons are gradually accelerated between source and drain in this short devices makes it possible to determine the effects of electron heating due to the longitudinal electric field. The decrease in τ_c due to an increase in V_D will depend on the distance of the trap from the source. This effect was observed experimentally by studying the V_D dependence of τ_c in the forward and reverse modes, and shows that RTS measurements thus constitute a valuable way to evaluate electron heating in deep-submicron MOSFETs.

At 4.2 K the equilibrium thermal energy is $k_B T \approx 0.36 meV$ and electron heating can occur at drain biases as small as some tenths of $1 meV$. If the trap is near the drain, the term due to electron heating in the right side of 6.23 will become dominant at drain biases greater than $1 mV$, and one would expect the mean thermal velocity of the electrons to increase with the square root of V_D. According to Eq. 6.5, one would then expect the capture time to show a $1/\sqrt{V_D}$ dependence. The data shown in Fig. 6.6 displays a much stronger dependence on V_D for forward operation, whereas τ_c is quite insensitive to changes in V_D when source and drain are interchanged. The exponential dependence of the capture time on V_D seen in forward operation can not be brought about only by changes in $\overline{v}$. As the drain bias is small compared to $(V_G - V_T)$, the inversion layer electron concentration n is not substantially affected by V_D. Hence, the efficiency of the capture process, thus σ, must be governed by a local, non-equilibrium temperature. Since it can not be implied that the whole lattice is at temperature T_e, it has to be verified if electron heating is not affecting the activation energy for capture, E_B in Eq. 6.6.

Electron capture is known to proceed via all states which can supply the extra electron needed (van Vliet 1994; Bonani, Ghione, Pinto, and Smith 1998). According to Eq. 6.4, the actual rate depends on the probability that a given state is occupied, its overlap with the trapped state and the activation energy out of the state. The effect of electron heating is schematically shown in the configuration-coordinate diagram of Fig. 6.7. The broken lines show the variation

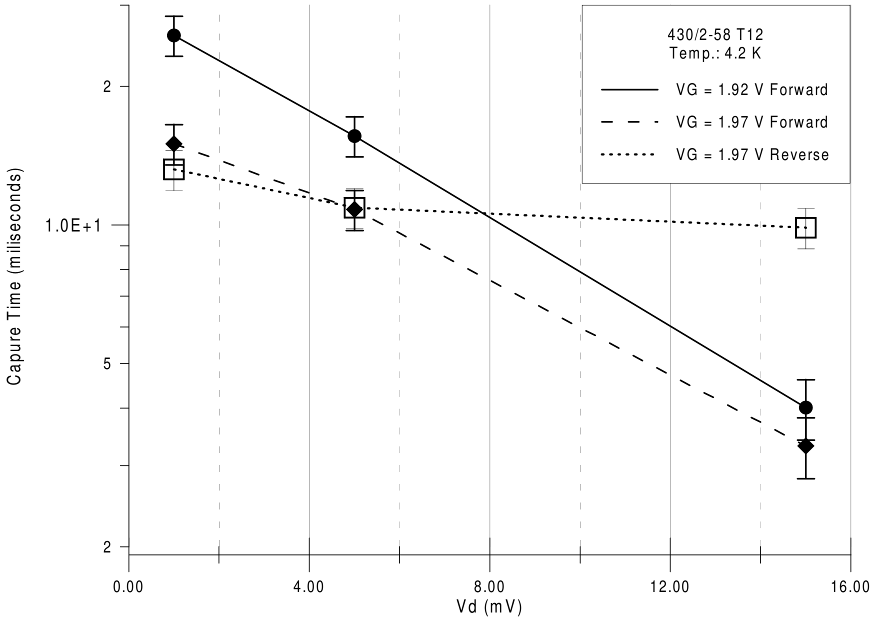


Figure 6.6: Mean lifetime for capture as a function of drain bias at 4.2 K for forward (solid symbols) and reverse (open symbols) operation. The threshold voltage V_T of the device is about 1.0 V. Design channel width and length are $W = 1.2\mu m$ and $L = 80nm$, respectively. Gate oxide thickness t_{ox} is 4.0 nm.

of total energy of the empty defect plus inversion layer electron as the electronic energy is increased. The full curve with the filled circle represents the total energy of the defect with its trapped electron. The activation energy for capture is represented by the energy difference between the minimum of a given broken curve and its intersection with the full curve. It can be seen from the figure that the activation energy decreases as the electron temperature increases. Since the capture time as given by 6.5 will exponentially depend on the activation energy E_B, this factor then explains the dependence observed in Fig. 6.6.

For the reversally operated device one has $l/L \approx 0$, and no significant carrier heating in the inversion layer near the trap is expected.

At room temperature it is not possible to locate the traps between source and drain using the above technique, because, once again, there is a competing effect that makes the analysis difficult. Namely, if V_D is of the order of $(V_G - V_T)$, the drain bias can strongly influence the surface potential at the trap. This is particularly true for our extremely short devices, and this effect is expected to be more pronounced if the trap is located near the drain side. An increasing V_D will cause the amount of inversion at the trap to decrease, where n in Eq. 6.7 will decrease. Hence, one would expect τ_c to increase with V_D, which is in contradiction to the decrease in τ_c expected from an increasing $\overline{v}$. This effect has particular importance when the device is biased near threshold. When the device is biased well above threshold and V_D is kept comparatively small, the decrease in n is small compared to the increase in $\overline{v}$, as it was the case for the devices studied at low

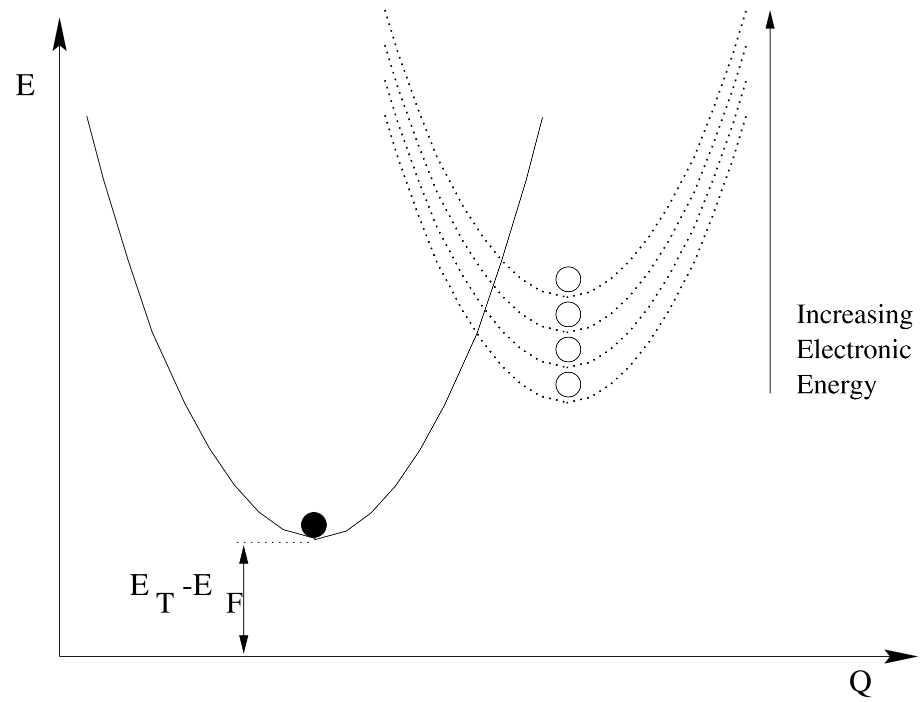


Figure 6.7: Configuration coordinate diagram showing the effect of carrier heating. The energy E is the sum of the elastic and electronic energies. It is plotted against one lattice coordinate Q. $\bullet$ labels the localized defect electronic state. The broken curves labeled with $\circ$ represent the variation of E for the empty trap and the free electron as the electronic energy increases. The mean capture time depends on the occupation probability of these states, the activation energy and the overlap with the final state. The activation energy for capture is given by the energy difference between the minimum of a given broken curve and its intersection with the full curve. It appears to decrease with increasing electronic energy.

temperature. Unfortunately, all traps found in the devices studied at room temperature are active near threshold.

Furthermore, at room temperature and higher drain biases inelastic scattering starts playing a major role and Eq. 6.23 becomes inappropriate. To estimate $\overline{v}$ near threshold using Eq. 6.24 is a very time consuming task and good accuracy is not expected. Nevertheless, there are traps for which it is possible to get some information concerning their location. The data shown in Fig. 6.8 reveals a quite asymmetrical behavior of the capture time as source and drain are interchanged. Whereas τ_c strongly increases with increasing V_D under forward operation, the changes in τ_c are much less significant if the device is operated reversally. The significant increase in τ_c can be only brought about by a decrease in the level of channel inversion below the trap. Hence, one has to conclude that the trap is located near the drain. If the device is operated near threshold, increments in the drain bias strongly reduce the level of inversion at the drain side of the channel. The source side is much less affected.

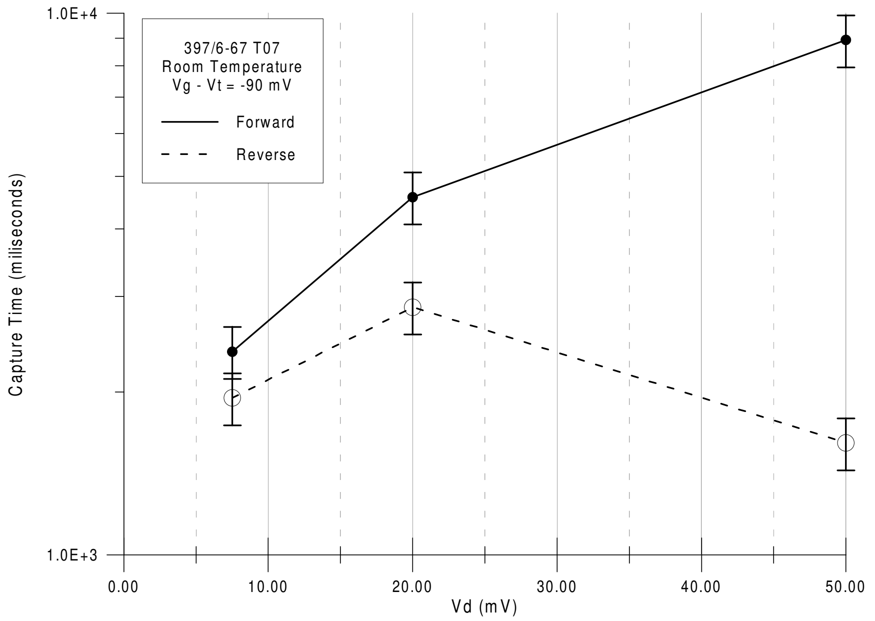


Figure 6.8: Mean lifetime for capture as a function of drain bias at room temperature under forward (solid symbols) and reverse (open symbols) operation for the same device as in Fig. 6.4.

6.2.2 Amplitude as a Function of Bias Point

The first step in estimating the amplitude of the random telegraph signal due to number fluctuation is to determine the number of free carriers in the channel. This was done by using the two-dimensional electrical simulator DESSIS (ISE - Integrated Systems Engineering AG 1994). The simulation results showed that for devices with $W = 1\mu m$, $L = 50nm$, $t_{ox} = 5nm$ and substrate doping $N_{Sub} = 8\ x\ 10^{17}\ cm^{-3}$, the inversion layer charge density is of approximately $8\ x\ 10^{10}\ cm^{-2}$ at threshold. Hence, there are about 50 free carriers in the active area of the device at threshold. If one of these is trapped a considerable fraction of the image charge is not located in the channel, but rather on the gate electrode and in the substrate, as can be inferred from eq. 6.19 and confirmed in numerical simulations done by (Martin, Li, Worley, and White 1997) and by (Mueller and Schulz 1992). The image charge in the channel induced by one trap is expected to be as small as 0.2 at threshold, for devices with thin gate oxide (Mueller and Schulz 1992). The estimate $\Delta N = 1$ is thus a worst case.

The amplitude of the signal due to mobility fluctuations can be estimated making the simplifying assumption that the trap is not a source of scattering when it is empty. This means that $1/\mu_{trap} = 0$ when the trap is empty. Hence, $\mu + \Delta\mu$, the mobility at the high current state, is the mobility limited by all other scattering centers in the channel (i.e., $\mu + \Delta\mu = \mu_{ch}$). The mobility μ in the low current state is then given by 6.20.

One clarifying step can be calculating the scattering cross-section of the trap needed to cause a mobility fluctuation $\Delta\mu$. This can be done by solving 6.20 for $1/\mu_{trap}$ and inserting the result in 6.21. This leads to the following expression for the scattering cross-section:

$$\sigma = \frac{qWL}{m^*\overline{v}}\left(\frac{1}{\mu} - \frac{1}{\mu + \Delta\mu}\right) \tag{6.25}$$

In order to cause a change in mobility of, for instance, $\Delta\mu/\mu = 2\%$ in a transistor with $W = 1\mu m$, $L = 50nm$ and $\mu = 300cm^2V^{-1}s^{-1}$ (a typical mobility value for the devices with $N_{Sub} = 8\ x\ 10^{17}\ cm^{-3}$) at room temperature, the trap must have a scattering cross-section of $160nm$. This dimension is larger than the channel length and at least one order of magnitude larger than realistic scattering cross-sections. These are expected to be of the order of the screening length($\approx 10nm$ at $300K$). As the RTS amplitude is overestimated by the assumption that the trap is not a source of scattering when it is empty, even greater scattering cross-sections would be needed.

Therefore, total RTS amplitudes as the ones measured can not be explained by the combined effect of number and mobility fluctuations in a uniform channel model. In addition, this mechanism can also not explain the wide range of amplitudes found in this work. It therefore appears that there is a real effect that is not accounted for by this model.

It has already been suggested by some authors that the range of amplitudes may be accounted for by some scattering centers being more strategically located than others (Siegert, Vitanov, and Eisele 1994; Martin, Li, Worley, and White 1997; Mueller and Schulz 1992; Ralls, Skocpol, Jackel, Howard, Fetter, Epworth, and Tennant 1984; Wang, Chang, Chiang, Wang, Zous, and Huang 1998). The starting point is in general fluctuations in the surface potential due to a spatially random distribution of fixed charges close to the Si/SiO_2 interface. Each fixed charge is assumed to cause a perturbation similar to the perturbation caused by the traps responsible for the RTS noise, and the perturbation cause by the fixed charges is assumed to be time independent.

The anomalous behavior in the dependence on gate voltage shown by the amplitude of the RTS depicted in Fig. 6.9 can not be appropriately explained by the uniform number and mobility fluctuation model. Based on the experimental data, this behavior is suggested to be originated by surface potential fluctuations in the channel inversion layer. An inhomogeneous potential distribution can be caused by a random individual dopant distribution in the channel (Asenov 1998; Stolk, Widdershoven, and Klaassen 1998). This inhomogeneity can allow for an early turn on in some parts of the channel. This is depicted in Fig. 6.10, where it is assumed that the channel can be subdivided into segments with different threshold voltages. The probability for a channel potential configuration which result in direct paths (filaments with lower V_T) between source and drain increases with the increase in channel width and decrease in channel length. The devices used in this work all have $W >> L$.

In an inhomogeneous channel, one can expect amplitude and capture time to depend on the relative position of the trap. If the trap is located above one segment that does carry current at the given gate voltage, the capture time is expected to be short and the charged defect will scatter the electrons in the segment. For a trap over a segment with a higher V_T, which is therefore much less inverted for the device operated near threshold, the capture time will be longer and the number of electrons that are effectively scattered by the charge trap will be smaller.

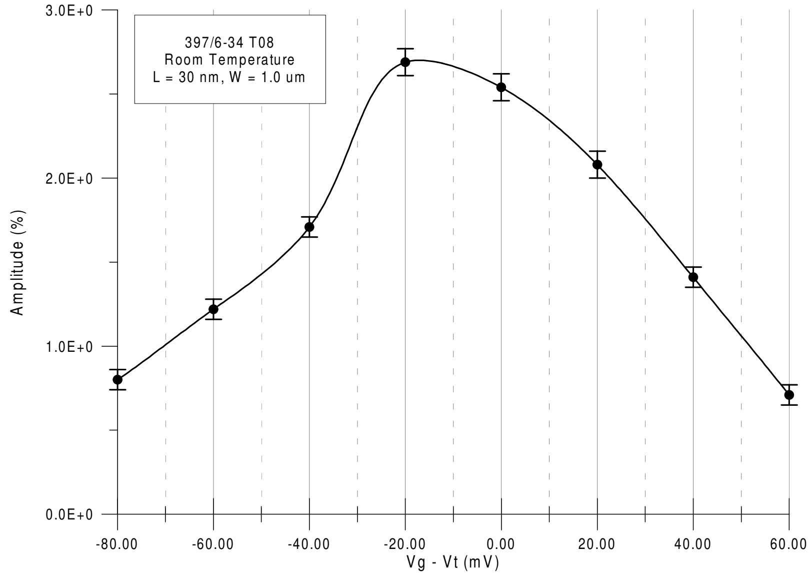
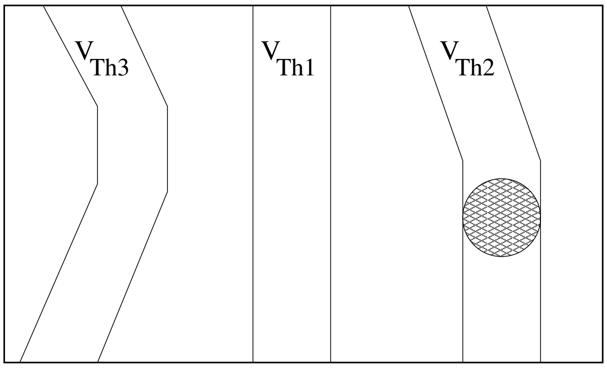


Figure 6.9: Anomalous behavior in the gate voltage V_G dependence of the RTS amplitude at room temperature.

A simplified speculative configuration that can explain the experimental data of Fig. 6.9 is presented in Fig. 6.10. Only the segments with the three lowest threshold voltages, $V_{Th1} < V_{Th2} < V_{Th3}$, are explicitly shown. The channel will then be gradually built up by this segments (percolation paths). The trap is located over the filament with a threshold voltage equal to V_{Th2}. First segment 1 starts conducting. As the gate voltage increases, segment 2 turns on, and the trap can become active. Worst case, the conductivity perturbation is assumed to be a round disk of zero conductivity that can block the entire filament. At this point, little current is carried by this segment and the amplitude of the RTS will be low. In weak inversion, the inversion layer charge density grows exponentially with increases in V_G. In strong inversion it shows a linear dependence on V_G. In moderate inversion it goes over from the exponential to the linear dependence. Hence, depending on the difference between V_{Th1} and V_{Th2}, the fraction of the total source-drain current carried by the second segment can increase, as V_G is increased. Consequently, the RTS amplitude can also increase. As the gate voltage increases further, segment 3 will also begin carrying current, and the fractional measure of the current flowing through the second segment will start decreasing. Furthermore, with an increasing level of inversion in the second segment, screening will become more effective. Thus, the RTS amplitude is then expected to start decreasing.

Although anomalous gate voltage dependence in RTS was previously observed (Kirton, Uren, Collins, Schulz, Karmann, and Scheffer 1989), this is the first time that the described mechanism has, to this author's knowledge, been suggested. Placing single traps into strategic positions, that is, in the correct current filament of a nonuniform channel, can lead to the behavior depicted in

Drain Side



Source Side

Figure 6.10: A simplified speculative channel configuration that can explain anomalous RTS amplitude dependence on gate voltage. The channel is viewed as being built up by several segments with different threshold voltages. Only the segments with the three lowest gate voltages, $V_{Th1} < V_{Th2} < V_{Th3}$, are explicitly shown. There is an interface trap in the second segment. When the trap is charged and scatters channel carriers it is represented by the disk with reduced conductivity.

Fig. 6.9.

This mechanism can also be the origin of the unexpected high RTS amplitudes measured in some probes. The signal amplitude strongly depends on the position of the individual trap relative to the percolation pattern of MOSFETs operated near threshold. If the trap is located in the segment that start conducting first, an unexpected high RTS amplitude will be measured.

Continuing the analysis of the RTS amplitude dependence on bias point, its dependence on V_D is studied. Figures 6.11 and 6.12 show that the amplitude of the signal does significantly decrease with increasing drain bias, at low temperature. The changes in ΔI for the device of Fig. 6.11 are more accentuated when it is forward biases, indicating that the trap ins near the drain under this condition. As a first step to understand the dependence of the RTS amplitude on drain bias, one could try to explain the experimental data using the classical approach, that is, studying the effect of electron heating on the parameters that determine the scattering cross-section at low temperature.

At low temperature the scattering cross-section of a charged center depends on the kinetic energy of the electrons and on the screening of its potential by the two-dimensional electron gas. For low drain bias at cryogenic temperatures, an increase in V_D results in an increase in the kinetic energy of the electrons. This increase in the kinetic energy of electrons tends to reduce scattering but at the same time reduce the screening effect. Therefore, the actual kinetic energy dependence is determined by the relative importance of these two effects. In order to investigate if this assumption can explain the experimental data, the screening length l_S and the critical radius r_C of the scattering potential must be estimated. The critical radius r_C can be assumed to be equal to the radius at which the interaction energy is larger than the kinetic energy of the electron (Simoen, Dierickx, Claeys, and Declerck 1992), where, it is assumed that the electron

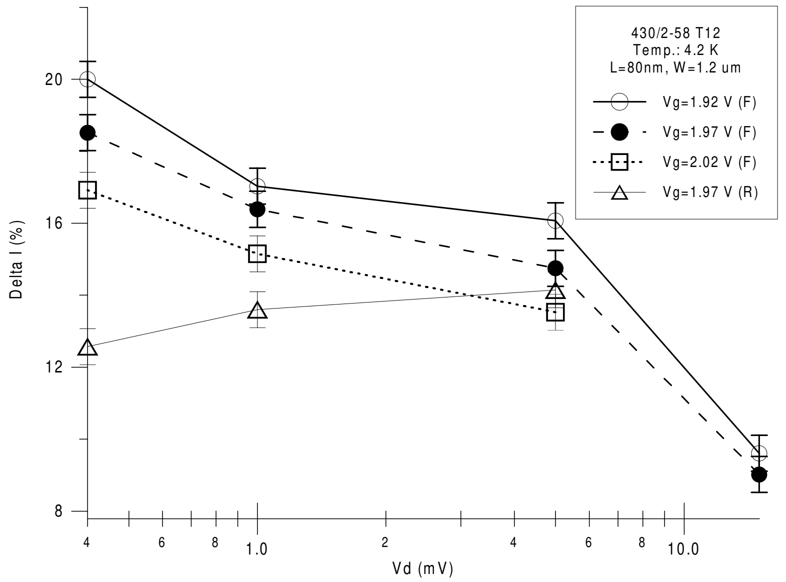


Figure 6.11: The dependence of the RTS amplitude on drain bias V_D under forward (F) and reverse (R) operation at different gate voltages V_G. Device channel width and length are $W = 1.2\mu m$ and $L = 80nm$, respectively. The gate oxide thickness is $t_{ox} = 4.0nm$. Measurements were carried out at $4.2K$.

is scattered once the interaction energy gets larger than the kinetic energy.

The presence of the trapped electron at the interface changes the local surface potential and channel electrons are scattered by the Coulombic potential

$$V(r) = \frac{q}{4\pi\varepsilon_{sc} r} \tag{6.26}$$

where r is the distance between the trapped and the channel electron and ε_{sc} the permittivity of silicon. Equating this Coulomb potential to the the kinetic energy of the electron E_k (in eV), one gets the following expression for the critical radius:

$$r_C = \frac{q^2}{4\pi\varepsilon_{sc} E_k} \tag{6.27}$$

As discussed in chapter 2, in strong inversion screening of the Coulomb potential by the channel carriers does occur. The potential fluctuation then typically extends over some screening lengths l_S, which for a two-dimensional electron gas can be approximated as (Singh 1993)

$$l_S = \sqrt{2}\frac{\varepsilon_{sc} E_k}{q N_{inv}} \tag{6.28}$$

The kinetic energy in the channel can be estimated with the same assumption already done above, i.e., assuming that mainly elastic scattering does occur between drain and source. This is

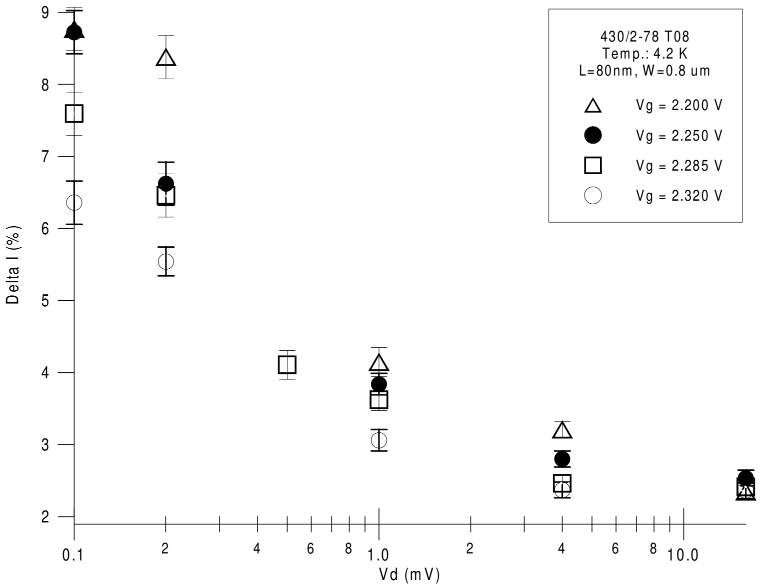


Figure 6.12: Measured RTS amplitude as a function of drain bias V_D for different gate voltages V_G at a temperature of $4.2K$. Device channel width and length are $W = 0.8\mu m$ and $L = 80nm$, respectively. The gate oxide thickness is $t_{ox} = 4.0nm$.

the expected case at low temperatures up to moderate drain bias, and 6.22 can then be used for the kinetic energy of the electrons in the inversion layer.

In Fig. 6.13 both r_C and l_S are plotted as a function of temperature, i.e., average kinetic energy E_k. An average kinetic energy of $1meV$ corresponds to an electron temperature $T_e \approx 12K$. The trap length is defined as the smallest of r_C and l_S. While r_C increases with $1/E_k$, l_S decreases with E_k and the curves will cross at a sufficiently low E_k. It follows that for the range of kinetic energies used in this work the trap length is determined by l_S in strong inversion and by r_C in weak inversion. The inversion layer charge density can be estimated using the oxide capacitance C_{ox}. For the device of Fig. 6.11 operated at the bias point where RTS is observed ($V_G - V_T \approx 1.0V$), the inversion layer charge density is calculated to be $N_{inv} \approx 5.4x10^{12}$. For the device of Fig. 6.12, $V_G - V_T \approx 1.3V$ and $N_{inv} \approx 7.0x10^{12}$. Hence, for both devices the trap is active when they are biased in strong inversion. Thus, after the results shown in Fig. 6.13, the scattering at Coulombic centers is expected to increase as the electron gas is heated by the longitudinal electric field.

The mobility as estimated by $\mu_{eff} = (LI_D)/(WV_DqN_{inv})$ did remain essentially constant at the high current state over the whole range of V_D under analysis ($0.1mV < V_D < 16.0mV$). Only for the device of Fig. 6.12 operated at the lowest drain biases ($V_D \leq 1mV$) a slight decrease in mobility could be seen as V_D was increased. It must than be concluded that the changes in ΔI_D are related to carrier scattering at the charged trap. Since screening is expected to become less effective as the thermal energy increases, scattering at the charged trap should become stronger. Hence, ΔI should be expected to increase as the electron gas is heated by the longitudinal electric

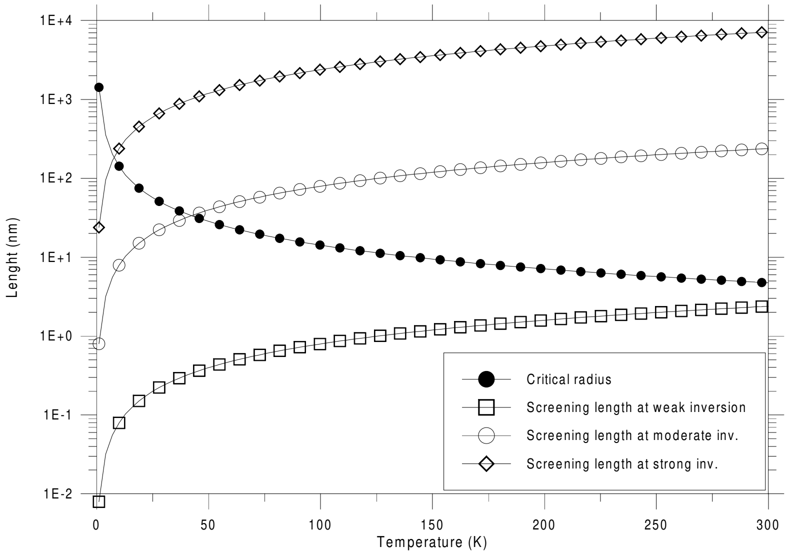


Figure 6.13: Screening length and critical radius of the scattering potential (trap) as a function of T (average kinetic energy). The screening length is calculated for three different values of the inversion layer charge density, corresponding to weak ($3.3x10^8 cm^{-2}$), moderate ($10^{10} cm^{-2}$) and strong ($10^{12} cm^{-2}$) inversion.

field, which is in contrast to these experimental results. It must then be concluded that the classical approach does not hold for very short devices operated at cryogenic temperatures.

According to the Monte Carlo simulations of the low-temperature behavior of inversion layer electrons performed by (Borzdov and Petrovich 1997), a decrease in scattering may occur if the electron gas is slightly warmed up by the longitudinal electric field. Other authors have also suggested that mobility could increase with temperature up to $T \approx 50K$ or so, due to changes in Coulomb scattering (Singh 1993). This would lead to decreases in the RTS amplitude as the drain bias is increased and the electron gas becomes consequently heated, resembling the experimental results obtained in this work. Unfortunately, there is no analytical form to treat the problem and one has to conclude that Coulomb scattering calculations can be performed properly only by complex numerical simulations.

7. Relevance for Nanoscale Electronics

In the previous chapters it was demonstrated that the behavior of sub-100 nm MOS devices can strongly deviate from that of their counterparts of greater dimensions. According to the silicon technology roadmap (SIA - Semiconductor Industry Association 1997), MOS devices with feature sizes below 100 nm should enter mass production around the year 2006. This will affect the performance of integrated circuits and systems in two quite different ways: *i)* the deviations can become a major factor limiting performance or *ii)* new phenomena found in these devices can be used to develop novel circuit concepts that can enhance performance further.

Concerning the first point, it has to be considered that the supply voltages are being continuously lowered, in order to reduce the power consumption of the individual devices (essential to allow high packing densities) and assure device reliability (e.g., scaling of the electric fields). A supply voltage $V_{DD} \approx 1.0V$ and threshold voltage $V_T \approx 0.2$ are demanded for logic circuitry. The steady decreasing supply voltages will make circuits more sensitive to fluctuations in device electrical characteristics. Furthermore, the decreasing device dimensions will boost fluctuations. As discussed in chapter 6, the trapping and detrapping of electrons at individual defect sites can cause large channel conductance fluctuations. In section 6.2.2, it was shown that the influence of the traps on channel conductance depends on their location along the channel and on the range of gate biases in which the trap is active. It was also discussed that the relative amplitude of the fluctuations is expected to increase as the active device area decreases. These fluctuations are directly related to the device noise levels, and noise may become an issue for the reliability of circuits based on devices with smaller dimensions, particularly if the voltage swing between the on and off states continues to decrease.

In section 6.2.2, RTS behavior suggesting a highly inhomogeneous channel potential at device turn on was also found. This channel potential fluctuations lead to devices with an increased threshold voltage V_T scatter. Hence, scatters in V_T can start playing a major role if the supply voltage continues to decrease.

The designers of future high performance integrated circuits with billions of devices have to keep all these issues in mind while defining the circuits and system architectures and, since the devices will keep shrinking to dimensions even shorter as the ones studied in this work, it is likely to become relevant for the correct modeling of MOSFET characteristics, that the new phenomena discussed in chapters 4 and 5 is accounted for in future device and circuit simulators. It can thus be expected that at the smallest dimensions single electron effects will limit normal transistor action.

Concerning innovative circuits and systems approaches, the constant improvements in device processing techniques, making devices with always smaller feature sizes available, may open the

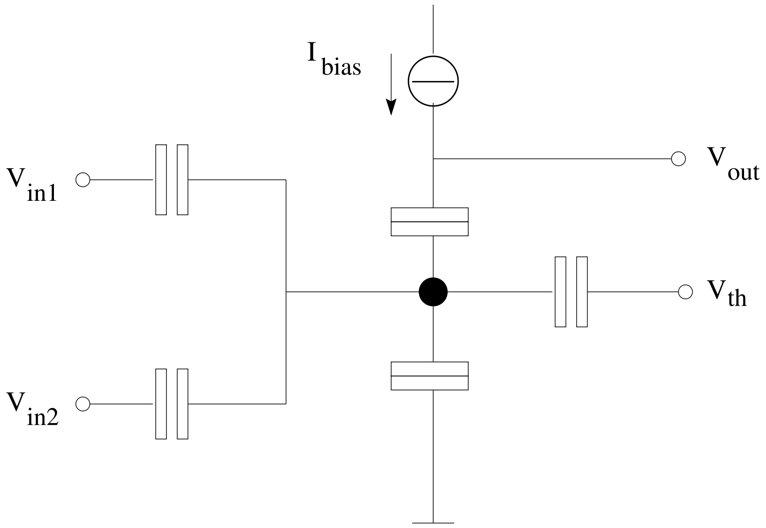


Figure 7.1: A neuron device built using only one SET.

door for new device and circuit concepts.

As described in sections 5.5 and 5.6, if there is a small isolated island or dot between the source and drain reservoirs and this island is capacitively coupled to the gate contact of the MOS structure, one has a Single Electron Transistor (SET). SET is at present believed to be useful mainly for memory, electrometer and metrology applications (ESPRIT Long Term Research: Microeletronics Advanced Research Initiative (MEL-ARI) 1999), although other possible applications, as in analog neural networks are also being investigated. An example that shows why a SET can be used to implement an extremely compact neuron device is depicted in Fig. 7.1 (Goossens 1998). The inputs define a capacitive charge addition circuit, and the output voltage of the current biased SET is a periodic function of the input voltages.

SET logic has also been proposed (Fujiwara, Takahashi, Yamazaki, Namatsu, Nagase, Kurihara, and Murase 1999). The most attractive approach is to implement circuits with reduced complexity based on the higher functionality offered by these devices. This leads to a reduction in the implementation cost per function. It is extremely desirable that such functional devices can be integrated on a single chip with MOSFETs, justifying the recent research efforts in the field of MOS-based SET technologies. One example of a logic gate that takes advantage of the higher functionality of these devices is shown in Fig. 7.2 (Goser and Pacha 1998). It is a programmable gate, where the Boolean function can be determined by means of the control input. Using this approach a CMOS-like logic family could be implemented.

One final example, showing a circuit approach where both MOSFETs and SETs are integrated to compose a single-electron memory is depicted in Fig. 7.3 (Yano, Ishii, Sano, Mine, Murai, Hashimoto, Kobayahi, Kure, and Seki 1999). Integrating SETs and MOSFETs on the same circuit makes it possible to take advantage of the best features of both worlds. Furthermore, changing the whole circuit concept from the conventional MOS technology to a new SET-based approach may be very cost intensive, as well as technologically as in the human aspect.

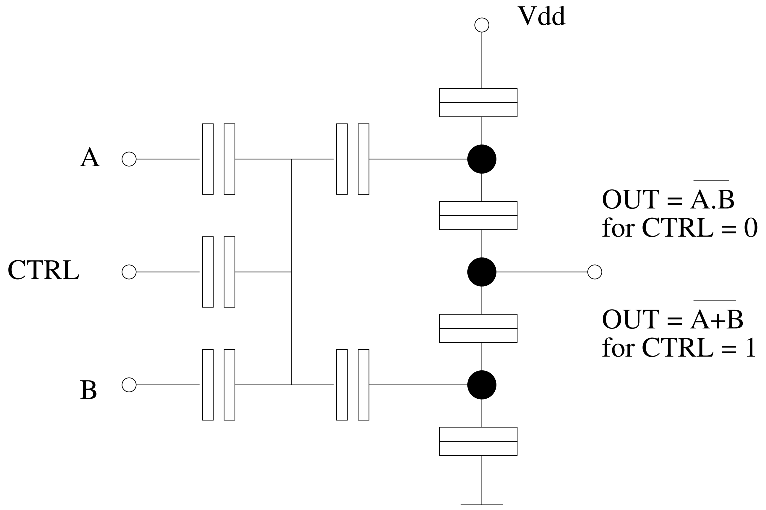


Figure 7.2: A programmable NAND/NOR logic gate composed of capacitively coupled SETs.

Therefore, a hybrid circuit approach seems to be economically advantageous and may favor the transition.

As discussed in chapter 2, a finite gate voltage V_T is required to turn the channel of a MOSFET conductive. In addition, if the channel area is small enough, the trapping of an electron into a localized state (such as an interface state or a quantum dot), can strongly affect the channel current. This was demonstrated in chapter 6, where it is shown that the trapping of a single electron into an interface state can induce significant changes in the channel current. The group at Hitachi Ltd. (Yano, Ishii, Sano, Mine, Murai, Hashimoto, Kobayahi, Kure, and Seki 1999) takes advantage of this two basic facts to propose the single-electron memory element of Fig. 7.3(a).

Memory operation can be achieved by controlled transfer of a single electron from and into the storage dot. Because of the barrier region formed between the storage dot and the reservoir (the source), and because of the small capacitance of the dot, Coulomb blockade can be achieved. By increasing the gate voltage to a given value (well above V_T), an electron is transfered into the dot at this critical V_G value. This point is labeled as ”write” in Fig. 7.3(b). As can be seen on this figure, the trapping of a single electron into the dot shifts the threshold voltage V_T of the MOSFET. The resulting hysteresis can be used for memory. By sensing the current difference between the two states, the stored information can then be retrieved. This point is labeled as ”read condition” in the figure.

Although the alternative design approaches presented above are very promising, there are major challenges and difficulties for SET-based circuits and systems. The most serious technological bottleneck seems to be overcoming the problem of background charge fluctuations, which can lead to severe reliability problems. A single background charge, such as a charged ion in the insulator, might strongly affect the electrical behavior of those devices (Likharev 1999). One alternative could be to approach this problem from both the circuit and system front. Neuronal networks, for instance, are known to be robust to variations in hardware, and if large systems

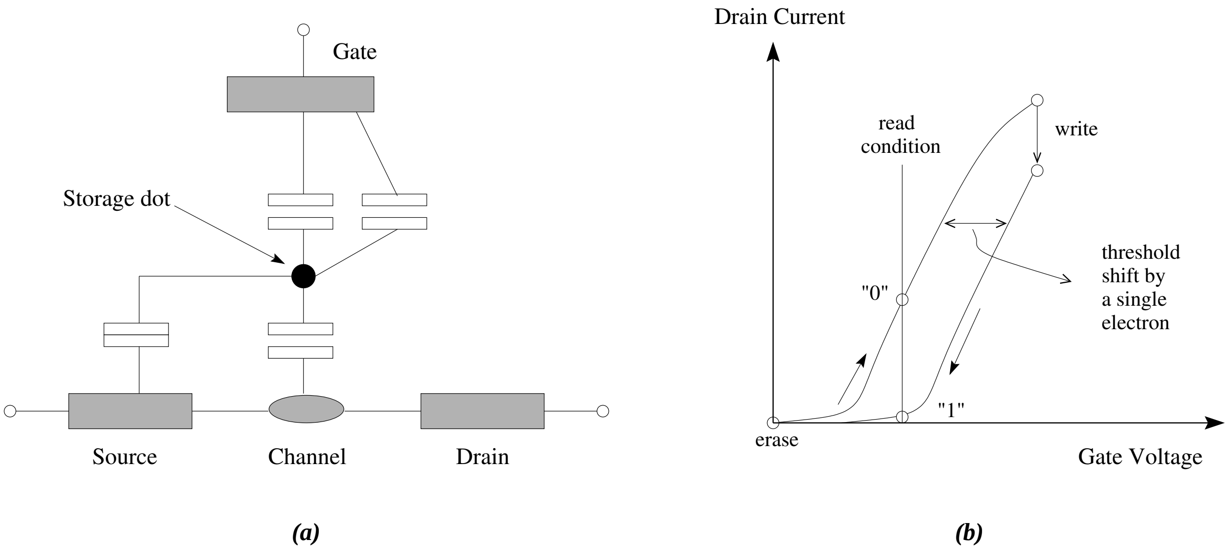


Figure 7.3: (a) Schematic representation of a single-electron memory element. (b) The schematic drain current *versus* gate voltage behavior of the proposed memory element. After (Yano, Ishii, Sano, Mine, Murai, Hashimoto, Kobayahi, Kure, and Seki 1999).

able to overcome this problem will be developed, neuronal networks are likely to be among them (Goossens 1998). For the circuit presented in Fig. 7.2, co-tunneling is also a challenge.

In section 5.6 it was shown that the Negative Differential Resistance (NDR) characteristics found in the devices studied in this work can match that of SETs, although it is not mature for circuits and system applications. First of all, it is present only at cryogenic temperatures and for small source-drain bias. Furthermore, it is not present in all devices, and the gate bias values at which it does occur seems not to be determined by device processing parameters. Indeed, device-to-device scattering has been the major concern against serious consideration of single-electronics for mass production (Averin and Likharev 1991). Nevertheless, many groups are working to achieve good NDR characteristics in MOS systems. To discuss all efforts being undertaken at the moment is not the purpose of this work. One of the promising approaches is the formation of tunneling junctions in thin silicon-on-insulator layers. Narrow insulating regions can be produced between the source and drain reservoirs, defining an isolated island capacitively coupled to the gate electrode (Ali and Ahmed 1994), needed if the device is to operate as a SET. Other implementations have also been studied, as in the work of the group of Mizuta, where a single-electron memory cell for use as a high-speed RAM in silicon technology is presented (Mizuta, Katayama, Müller, Williams, Nakazato, and Ahmed 1998).

In conclusion, experiments indicate that advanced silicon nanofabrication processes could be useful in implementing single electronic circuits in the near future.

8. Conclusion

This thesis covers mesoscopic phenomena in nanometer scale MOSFETs. The devices, studied at low ($0.3K < T < 35K$) and room temperature, have channel lengths between 30 nm and 120 nm and channel widths between 0.6 μm and 100 μm. They are processed on (100) silicon in a conventional planar bulk technology.

Mesoscopic phenomena in the form of negative differential resistance and single electron switching events in the channel of the devices are demonstrated.

First, reproducible unexpected periodic transconductance oscillations in the drain current *versus* gate voltage characteristics of nMOSFETs are studied. The oscillations, present from subthreshold up to strong inversion, are reproducible from sample to sample and with temperature cycling. No dependency of the oscillation period on gate oxide thickness, channel length or width could be observed. The period of the oscillations does not change in magnetic fields up to 15 T. The periodic pattern smears out with rising temperature or increasing drain voltage, but can be found at temperatures up to 35 K and drain source electrical fields up to $2.4 kV/cm$, corresponding to a drain bias of $12 mV$ for devices with a channel length of $50 nm$.

Various electric transport models for small size MOS systems are studied, such as hopping conduction, resonant tunneling, universal conductance fluctuations, weak interference between source and drain junctions, charge density waves, carrier density quantization and Coulomb blockade. In order to verify if these different models fit experimental results, numerical calculations and theoretical investigations are carried out.

The periodic transconductance oscillations found are a surprising phenomenon. Though they are fairly reproducible and no dependency on device geometry could be observed, it is easier to find them in the smallest devices. For several reasons, Coulomb blockade seems to be a rather plausible explanation for the observed characteristics, although some questions, as the origin of the quantum dots, remain open.

Second, the Random Telegraph Signal (RTS) was characterized at room and low temperatures. The characterization of the Random Telegraph Signal indicated the possibility of observing the behavior of individual localized defect states (charge traps) close to the Si/SiO_2 interface. It is shown that a trap can be used as a probe into the local surface potential. The behavior of the RTS does depend upon the properties of the trap and channel electrons. Hence, studying the bias-voltage dependence of the Random Telegraph Signal, it is possible to acquire valuable information concerning the properties of the trap itself and the properties of the channel electrons near it.

Regarding the trap properties, it is possible to determine the traps geometrical position along the channel and its energetical location relative to the Fermi level at the Si/SiO_2 interface. For the channel electron properties, it is shown that RTS is a valuable tool to study the channel potential landscape, which becomes of particular interest if the trap is active near threshold. Due to the stochastic nature of the channel dopant distribution and local fluctuations in channel geometrical parameters, an inhomogeneous potential distribution can be present in the channel. This inhomogeneity can allow for an early turn on in some parts of the channel. As a result, the bias dependence of the RTS will show an anomalous behavior. Studying this anomalous behavior provides important knowledge concerning the mechanisms that influence electrical channel formation in ultra short devices. Furthermore, it is shown that RTS analysis is a valuable way to study effects as Coulomb scattering and electron gas heating in very small area devices.

Finally, this work attempts to put forth that RTS will become of increasing relevance as device dimensions keep shrinking and supply voltages continue to be lowered, suggesting that RTS noise may become an issue for advanced MOS-technologies, particularly for low power and analog-digital mixed-mode applications.

A. Kurzfassung

Der Hauptgrund für den Erfolg der Halbleiterindustrie ist deren fortwährende Fähigkeit, Elektronikprodukte mit sinkenden Kosten pro Funktion bei gleichzeitig steigender Leistung zu liefern. Dies begründet sich durch die stetige Reduktion der Bauelementgröße verbunden mit einer kontinuierlich wachsenden Packungsdichte. So kommen im Jahr 2012 voraussichtlich Produkte auf den Markt, die auf Bauelemente mit einer *minimum feature size* von 35 nm basieren und 10^8 Transistoren pro cm^2 aufweisen (SIA - Semiconductor Industry Association 1997).

Ab einer bestimmten Bauelementgröße versagt jedoch die Skalierung der physikalischen Prozesse und neue Phänomene, die bei größeren Strukturen nicht auftreten, bestimmen das Bauelementverhalten. Dies wirkt sich, falls die Bauelemente nicht angemessen charakterisiert und modelliert werden, hinderlich auf die weitere Entwicklung der Halbleiterindustrie aus. Daher ist das Hauptziel dieser Arbeit, die elektrische Charakterisierung und Modellierung von Sub-100 nm MOS-Feldeffekttransistoren voranzubringen. Dabei ist das Ziel ein physikalisches Modell, das auf mathematischem Wege die physikalischen Eigenschaften der untersuchten Bauelemente beschreibt.

Die experimentell gewonnenen Daten werden mit den Vorhersagen von verschiedene Modellen verglichen. Einige dieser Modelle sind in Simulatoren eingebettet. Auf diese Weise werden numerische Werte für die Parameter gewonnen und es wird überprüft, ob das Modell den gestellten Anforderungen entspricht. Gegebenenfalls werden die Modelle optimiert und neue Messungen und Simulationen durchgeführt.

Im Rahmen dieser Arbeit wurden MOS-Transistoren, deren Kanallänge und Kanalweite zwischen 30 nm und 120 nm bzw. 0,6 μm und 100 μm liegen, sowohl bei niedriger $(0,3K < T < 35K)$ als auch bei Raumtemperatur untersucht. Hergestellt wurden diese Schaltungselemente auf (100) Silizium unter Verwendung einer konventionellen Planar-Bulk-Technologie, unter Einsatz eines Stufen-Rückätzverfahren, entwickelt am *Lehrstuhl für Bauelemente der Elektrotechnik der Universität Dortmund*. Es werden mesoskopische Phänomene in Form von negativen differentiellen Widerständen in der Eingangskennlinie und Ein-Elektron-Schaltvorgänge im Kanalgebiet festgestellt. Die reproduzierbaren unerwarteten Gegenwirkleitwertoszillationen von n-MOSFETs in der Drainstrom *gegen* Gatespannung Kennlinie werden analysiert. Diese Oszillationen treten sowohl im Anlaufbereich als auch bei starker Inversion auf und sind sowohl von Probe zu Probe als auch innerhalb von Temperaturzyklen reproduzierbar. Eine Abhängigkeit der Oszillationsperiode von der Gateoxiddicke, der Kanallänge oder der Kanalweite läßt sich dabei nicht beobachten. Die Periode ändert sich auch nicht bei anliegenden magnetischen Feldern von bis zu 15 T.

Die auftretenden periodischen Oszillationen sind ein überraschendes Phänomen. Obwohl diese Oszillationen weitgehend reproduzierbar sind und keine Abhängigkeit von den Baulementgeometrien beobachtet werden können, ist es einfacher, sie an den kleinsten Bauelementen meßtechnisch zu erfassen. Das periodische Muster verschmiert mit steigender Temperatur oder zunehmender Drainspannung und tritt bei Temperaturen über 35 K und elektrischen Drain-Source-Feldern von über $2,4\ kV/cm$ nicht mehr auf. Dies entspricht einer Drainvorspannung von $12mV$ bei einer Kanallänge von 50 nm.

Zur Erklärung dieses Phänomens werden zahlreiche Transportmodelle für MOS-Systeme kleiner Abmessung analysiert, darunter *Hopping-Transport*, Resonanz-Tunnelung (*Resonant Tunneling*), allgemeine Transportfluktuation (*Universal Conductance Fluctuations*), schwache Interferenz zwischen der Source und Drain Verbindung, Ladungsdichtewellen (*Charge Density Waves*), Ladungsträgerdichtequantisierung und Coulomb Blockade.

Zur Überprüfung der unterschiedlichen Modelle wurden numerische Berechnungen und theoretische Untersuchungen durchgeführt, und mit der experimentellen Ergebnisse verglichen. Obwohl es Aspekte gibt, die das Modell der Coulomb Blockade in Frage stellen, liegen Anzeichen für die Coulomb Blockade als eine plausible Erklärung für die beobachteten Kennlinienverläufe vor.

Da dies das erste Mal ist, daß von diesen Phänomen im Zusammenhang mit einem herkömmlichen Bulk-MOS-System berichtet wird, tritt die Vermutung auf, daß das Verhalten von Ultrakurzkanal Bauelementen noch unzureichend modelliert ist.

Im nächstem Teil wird ein weiteres Ein-Elektron Schaltungsphänomen behandelt. Es wird gezeigt, wie sich das Verhalten von einzelnen Defektstellen im Oxid untersuchen läßt. Das Einfangen und Emittieren von Ladungsträgern durch einzelne Defektstellen bewirkt diskrete Veränderungen im Drainstrom. Diese sind unter dem Namen *Random Telegraph Signal* (RTS) bekannt.

Eine Defektstelle kann als Sonde zur Bestimmung des lokalen Flächenpotentials innerhalb des Transistors verwendet werden. Das Verhalten des RTS hängt von den Defekteigenschaften und den Kanalelektronen ab. Folglich können durch Untersuchungen des RTS bezüglich seiner Abhängigkeit von der Vorspannung wertvolle Informationen über die Eigenschaften der Defekstelle und auch der benachbarten Kanalelektronen gewonnen werden. In Bezug auf die Defekteigenschaften ist es möglich, die geometrische Position im Kanal und die energetische Lage relativ zum Ferminiveau zu bestimmen.

Hinsichtlich der Eigenschaften der Kanalelektronen wird gezeigt, daß RTS ein wertvolles Werkzeug ist, um die Potentialverteilung im Kanal zu analysieren. Dies ist von besonderem Interesse wenn die Defektstelle nah der Schwellenspannung aktiv ist. Aufgrund der stochastischen Natur der Kanaldotierung und der lokalen Schwankungen der geometrischen Kanalparameter kann eine inhomogene Potentialverteilung im Kanal vorhanden sein. Diese Inhomogenität kann für ein frühzeitiges Einschalten des Transistors in manchen Kanalgebieten in Betracht gezogen werden. Folglich zeigt die Vorspannungsabhängigkeit des RTS ein anomales Verhalten. Eine Untersuchung dieses anomalen Verhaltens liefert wichtige Information über den Mechanismus, der die Kanalbildung in ultrakurzen Bauelementen beeinflußt.

Zusätzlich wird gezeigt, daß die RTS-Analyse ein gutes Werkzeug zu Untersuchung der Elektronengaserhitzung in sehr kleinen Bauelementen ist. Da die Bauelementgrößen weiter sinken und die Versorgungsspannung stetig reduziert wird, wächst die Bedeutung des RTS. Folglich wird RTS-Rauschen eine wichtige Rolle für moderne MOS-Technologien spielen, insbesondere für *low power* und *analog-digital mixed-mode* Anwendungen.

Das Ziel der Forschung, die im Rahmen dieser Arbeit dargestellt wird, ist ein besseres Verständnis der Eigenschaften von ultrakurzen MOS-Bauelementen. Dies ist zwingend notwendig für optimierte und zuverlässige Systeme, einhergehend mit erhöhter Produktqualität.

References

Adkins, C. J., S. Pollit, and M. Pepper (1976). The Anderson transition in silicon inversion layers. *J. de Physique 37*, C4–343–C4–347.

Ali, D. and H. Ahmed (1994). Coulomb blockade in a silicon tunnel junction device. *Appl. Phys. Lett. 64*, 2119–2120.

Ando, T., A. B. Fowler, and F. Stern (1982). Electronic properties of two-dimensional systems. *Rev. of Modern Physics 54*, 437–649.

Aronzon, B. A., A. S. Vedeneev, and V. V. Rylkov (1997). Mesoscopic effects in the hopping conductivity region of macroscopic quasi-two-dimensional systems. *Semicond. 31*, 551–555.

Asenov, A. (1998). Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 μm MOSFETs: a 3-D atomistic simulation study. *IEEE Trans. on Electron Dev. 45*, 2505–2513.

Averin, D. and K. Likharev (1991). *Quantum effects in small disordered systems*. London, U.K.: Elsevier.

Bagwell, P. F., S. L. Park, A. Yen, D. A. Antoniadis, H. I. Smith, T. P. Orlando, and M. A. Kastner (1992). Magnetotransport in multiple narrow silicon inversion channels opened electrostatically into a two-dimensional electron gas. *Phys. Rev. B 45*, 9214–9221.

Berggren, K. F., T. J. Thornton, D. J. Newson, and M. Pepper (1986). Magnetic depopulation of 1D subbabds in a narrow 2D electron gas in a GaAs:AlGaAs heterojunction. *Phys. Rev. Lett. 57*, 1769–1772.

Bollu, M. and F. Koch (1988, Mai). Low temperature conductance of Si-MOS devices after hot carrier stress. In *Proc. of the Meeting on the Phys. and Chem. of the Si/SiO2 Interface*, Atlanta, pp. 345–349.

Bonani, F., G. Ghione, M. R. Pinto, and R. K. Smith (1998). An efficient approach to noise analysis through multidimensional physics-based models. *IEEE Trans. on Electron Dev. 45*, 261–269.

Borzdov, V. M. and T. A. Petrovich (1997). Monte Carlo simulation of the low-temperature mobility of two-dimensional electrons in a silicon inversion layer. *Semicond. 31*, 72–75.

Bourgoin, J. and M. Lannoo (1983). *Point defects in Semiconductors II: Experimental aspects*. Berlin: Springer-Verlag.

Brews, J. R. (1978). A charge-sheet model of the MOSFET. *Solid-State Electr. 21*, 345–355.

Brews, J. R. (1979). Subthreshold behavior of uniformly and nonuniformly doped long-channel MOSFET. *IEEE Trans. on Electr. Dev. 26*, 1282–1291.

Bryant, G. W. (1987). Electronic structure of ultrasmall quantum-well boxes. *Phys. Rev. Lett. 59*, 1140–1143.

Chen, Y.-G., S.-Y. Ma, B. Kuo, Z. Yu, and R. W. Dutton (1995). An analytical drain current model considering both electron and lattice temperatures simultaneously for deep submicron ultrathin SOI NMOS devices with self-heating. *IEEE Trans. on Electron Dev. 42*, 489–496.

Cheng and Sullivan (1973). On the role of scattering by surface roughuness in silicon inversion layers. *Surf. Sci. 34*, 717–731.

Colinge, J. P., X. Baie, V. Bayot, and E. Grivei (1996). A silicon-on-insulator quantum wire. *Solid-State Electronics 39*, 49–51.

Dorda, G. (1990). A new general quantum transport model. *Superlattices and Microstructures 17*, 103–113.

Dorda, G. (1992). Universal behaviour of the electronic transport. *Physica Scripta T45*, 297–299.

Eisele, I., H. Baumgärtner, and W. Hansch (1995). Silicon nanostructure devices. *J. Of Crystal Growth 157*, 248–254.

Eisele, I., H. Baumgrtner, and W. Hansch (1995). *Low Dimensional Structures prepared by Epitaxial Growth or Regrowth on Paterned Substrates*, Chapter Self-assembling growth of silicon nanostructures with micro shadow masks, pp. 161–72. The Netherlands: Kluwer Academic Publishers.

ESPRIT Long Term Research: Microeletronics Advanced Research Initiative (MEL-ARI) (1999). *The Nanoelectronics Road-Map*. Brussels, Belgium: ESPRIT Long Term Research: Microeletronics Advanced Research Initiative (MEL-ARI).

Fang, F. F. and W. E. Howard (1966). Negative field-effect mobility on (100) Si surfaces. *Phys. Rev. Lett. 18*, 797–799.

Field, S. B., M. A. Kastner, U. Meirav, J. H. F. S.-T. D. A. Antoniadis, H. I. Smith, and S. J. Wind (1990). Conductance oscillations periodic in the density of one-dimensional electron gases. *Phys. Rev. B 42*, 3523–3536.

Fock, V. (1928). Bemerkungen zur Quantellung des harmonischen Oszillators im Magnetfeld. *Z. Phys. 47*, 446–448.

Fowler, A. B., F. F. Fang, W. E. Howard, and P. J. Stiles (1966). Magneto-oscillatory conductance in silicon surfaces. *Phys. Rev. Lett. 16*, 901–903.

Fowler, A. B., G. Timp, J. Wainer, and R. Webb (1986). Observation of resonant tunneling in silicon inversion layers. *Phys. Rev. Lett. 57*, 138–141.

Fujiwara, A., Y. Takahashi, K. Yamazaki, H. Namatsu, M. Nagase, K. Kurihara, and K. Murase (1999). Double-island single-electron devices: a useful unit device fo single-electron logic LSI's. *IEEE Trans. on Electron Dev. 46*, 954–959.

Goossens, M. (1998). *Analog neural networks in single-electron tunneling technology*. Ph. D. thesis, Technische Universiteit Delft, Delft, The Netherlands.

Goser, K. F. and C. Pacha (1998, September). System and circuit aspects of nanoelectronics. In *Proceedings of the European Solid-State Cicuits Conference*, pp. 18–29.

Grüner, G. (1988). The dynamics of charge-density waves. *Rev.Mod.Phys. 60*, 1129–1181.

Grüner, G. (1992). *Density waves in solids*. Reading, Massachusetts: Addison-Wesley.

Gutierrez, E. A. (1995). The drain threshold voltage Vtd in submicrometer MOS transistors at 4.2 K. *IEEE Electron Dev. Lett. 16*, 85–87.

Hanamura, H., M. Aoki, T. Masuhara, O. Minato, Y. Sakai, and T. Hayashida (1986). Operation of bulk CMOS devices at very low temperaures. *IEEE J. of Solid-State Circ. 21*, 484–489.

Hansch, W., V. R. Rao, C. Fink, F. Kaesen, and I. Eisele (1998). Electric field tailoring in MBE-grown vertical sub-100 nm MOSFETs. *Thin solid films 321*, 206–214.

Hänsch, W., T. Vogelsang, R. Kircher, and M. Orlowski (1989). Carrier transport near the $Si - SiO_2$ interface of a MOSFET. *Solid State Electronics 32*, 839–849.

Hareland, S. A., S. Jallepalli, G. Chindalore, W.-K. Shih, A. F. T. Jr., and C. M. Maziar (1997). A simple model for quantum mechanical effects in hole inversion layers in silicon pmos devices. *IEEE Trans. on Electron Dev. 44*, 1172–1173.

Hartstein, A. (1991). Quantum interference in ultrashort channel length silicon metal-oxide-semiconductor field-effect transistors. *Appl. Phys. Lett. 59*, 2028–2030.

Hofmann, F. and D. A. Wharam (1996). Investigation of the Coulomb blockade in a parallel quantum dot geometry. *Adv. in Sol. St. Phys. 35*, 197–213.

Horstmann, J. T., U. Hilleringmann, and K. Goser (1996). Characterisation of sub-100nm-MOS-Transistors processed by optical lithography and a sidewall-etchback technique. *Microel. Eng. 30*, 431–434.

Horstmann, J. T., U. Hilleringmann, and K. F. Goser (1997, September). Detailed matching analysis of sub-50-nm MOS-transistors. In H. Grünbacher (Ed.), *Proceedings of the* 27^{th} *european solid-state device research conference*, pp. 240–243.

Hung, K. K., P. K. Ko, C. Hu, and Y. C. Cheng (1990). A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors. *IEEE Trans. on Electron Devices 37*, 654–665.

Ip, B. K. and J. R. Brews (1998). Quantum effects upon drain current in a biased MOSFET. *IEEE Trans. on Electron Devices 45*, 2213–2221.

ISE - Integrated Systems Engineering AG (1994). *Modeling of semiconductor technology, devices, and systems*. Zurich, Switzerland: ISE - Integrated Systems Engineering AG.

Ishikuro, H. and T. Hiramoto (1997). Quantum mechanical effects in the silicon quantum dot in a single-electron transistor. *Appl. Phys. Lett. 71*, 3691–3693.

Jallepalli, S., J. B. W.-K. Shih, M. R. Pinto, C. M. Maziar, and A. F. T. Jr. (1997). Electron and hole quantization and their impact on deep submicron silicon p- and n-MOSFET characteristics. *IEEE Trans. on Electr. Dev. 44*, 297–302.

Janik, T. and B. Majkusiak (1998). Analysis of the MOS transistor based on the self-consistent solution of the Schrödinger and Poisson equations and on the local mobility model. *IEEE Trans. on Electron Devices 45*, 1263–1271.

Johnson, F. A. (1959). Lattice absortion bands in silicon. *Proc. Phys. Soc. LXXIII*, 265–272.

Kastner, M. A., R. F. Kwasnick, and J. C. Licini (1987). Conductance fluctuations near the localized-to-extended transition in narrow Si metal-oxide-semiconductor field-effect transistors. *Phys. Rev. B 36*, 8015–8031.

Kirton, M. J. and M. J. Uren (1989a). In C. R. Helms and B. E. Deal (Eds.), *The physics and the chemestry of the* SiO_2 *and the* $Si : SiO_2$ *interface*, pp. 341–349.

Kirton, M. J. and M. J. Uren (1989b). Noise in solid-state microstructures: A new perspective on individul defects, interface states and low-frequency (1/*f*) *noise.* Advances in Physics 38*, 367–468.*

Kirton, M. J., M. J. Uren, S. Collins, M. Schulz, A. Karmann, and K. Scheffer (1989). Individual defects at the Si : SiO_2 *interface.* Semicond. Sci. Technol. 4*, 1116–1126.*

Koch, F., M. Bollu, and A. Asenov (1988). MOSFET's under electrical stress-degration, subthreshold conduction, and noise in a submicron structure. Proc. of the Fifth Int. School on Physics and Technology of Submicron Structures 5*, 253–262.*

Kunze, U., T. Drebinger, B. Klehn, and J. Lindolf (1994)). Observation of 1D electron states at the boundary between an MOS and a Schottky contact by electron tunneling. Surf. Sci. 305*, 633–636.*

Lee, K., J.-S. Choi, S.-P. Sim, and C.-K. Kim (1991). Physical understanding of low-field carrier mobility in silicon MOSFET inversion layer. IEEE Trans. on Electr. Dev. 38*, 1905–1911.*

Lee, P. and A. Stone (1985). Universal conductance fluctuations in metal. Phys. Rev. Lett. 55*, 1622–1625.*

Likharev, K. (1999). Single-electron devices and their applications. Proceedings of the IEEE 87*, 606–632.*

Martin, S. T., G. P. Li, E. Worley, and J. White (1997). The gate bias and geometry dependence of random telegraph signal amplitudes. IEEE Electron Dev. Let. 18*, 444–446.*

McEuen, P. L., E. B. Foxman, J. Kinaret, U. Meirav, and M. A. Kastner (1992). Self-consistent addition spectrum of a coulomb island in the quantum hall regime. Phys. Rev. B 43*, 11419–11422.*

Meirav, U., M. A. Kastner, M. Heiblum, and S. J. Wind (1989). One-dimensional electron gas in GaAs: Periodic conductance oscillations as a function of density. Phys. Rev. B 40*, 5871–5874.*

Merkt, U., J. Huser, and M. Wagner (1991). Energy spectra of two electrons in a harmonic quantum dot. Phys. Rev. B 43*, 7320–7323.*

Mieville, J. P., T. Ouisse, S. Cristoloveanu, L. Forro, N. Revil, and M. Dutoit (1994). Observation of nonstationary transport in deep submicron n-channel metal-oxide-semiconductor with Shubnikov-de Hass oscillations. J. Appl. Phys. 75*, 4226–4232.*

Mizuta, H., K. Katayama, H. Müller, D. Williams, K. Nakazato, and H. Ahmed (1998). High-speed single-electron memory: cell design and architecture. In Proceedings of the 2nd. Int. Workshop on Physics and Modeling of Devices based on Low-Dimensional Structures*, pp. 18–29.*

Moglestue, C. (1986). Self-consistent calculation of electron and hole inversion charges at silicon-silicon dioxide interfaces. J. Appl. Phys. 59*, 3175–3183.*

Mott, N. F. (1974). Conduction in non-crystaline systems X. mobility and percolation edges. Philos. Mag. 29*, 613–639.*

Mueller, H. H. and M. Schulz (1992). Conductance modulation of submicrometer metal-oxide-semiconductor field-effect transistors by single-electron trapping. J. Appl. Phys. 71*, R23–R41.*

Nakajima, Y. and et al. (1994). Fabrication of a silicon quantum wire surrounded by silicon dioxide and ist transport properties. Appl. Phys. Lett. 65*, 2833–35.*

Nakazato, K. and H. Ahmed (1995). The multiple-tunnel junction and its application to single-electron memory and logic circuits. Jpn. J. Appl. Phys. 34*, 700–706.*

Neugebauer, T., G. Landwehr, and K. Hess (1978). Negative differential resistance in (100) n-channel silicon inversion layers. Solid-State Electr. 21*, 143–146.*

Omura, Y., S. Horiguchi, M. Tabe, and K. Kishi (1993). Quantum mechanical effects on the threshold voltage of ultrathin-SOI nMOSFETs. IEEE Electron Dev. Lett. 14*, 569–571.*

Omura, Y. and K. Izumi (1996). Quantum mechanical influences on short-channel effects in ultra-thin MOSFET/SIMOX devices. IEEE Electron Dev. Lett. 17*, 300–302.*

Omura, Y., K. Kurihara, T. Takahasi, T. Ishiyama, and K. Izumi (1997). 50-nm channel nMOSFET/SIMOX with an ultrathin 2- or 6-nm thick silicon layer and their significant features of operations. IEEE Electron Dev. Lett. 18*, 190–193.*

Ouisse, T., S. Cristoloveanu, and D. K. Maude (1993). Experimental investigation of silicon-on-insulator metal-oxide-semiconductor field-effect transistors at high magnetic field and low temperature. J. Appl. Phys. 74*, 408–415.*

Pacelli, A., S. Villa, A. L. Lacaita, and L. M. Peron (1999). Quantum effects on the extraction of MOS oxide traps by 1/f noise measurments. IEEE Trans. on Electron Devices 46*, 1029–1035.*

Peters, M. G., S. G. den Hartog, J. I. Dijkhuis, and L. W. Molenkamp (1998). Single electron tunneling and suppression of short-channel effects in submicron silicon transistors. J.Appl.Phys. 84*, 5052–5056.*

Poole, D. A., M. Pepper, and H. W. Myron (1983). Loss of dimensionality, localisation and conductance oscillations in n-type GaAs FETs. Physica 117B&118B*, 697–699.*

Raikh, M. E. and I. M. Ruzin (1990). Size effect in the longitudinal hopping conduction of a narrow tow-dimensional channel. Phys. Rev. B 42*, 11203–11207.*

Ralls, K. S., W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth, and D. M. Tennant (1984). Phys. Rev. Lett. 52*, 228–231.*

Schmidt, T. (1997). Single-electron transport in semiconductor nanostructures: Artificial atoms, molecules and local density of state fluctuations. *Ph. D. thesis, Max-Planck-Institut für Festkörperforschung, Stuttgart.*

Schoenberg, D. (1984). Magnetic oscillation in metals. *Cambridge: Cambridge University Press.*

Schulz, M. and N. M. Johnson (1978). Evidence of multiphonon emission from interface states in MOS structures. Solid State Com. 25*, 481–484.*

Shi, Z. M., J.-P. Miéville, and M. Dutoit (1993). Effect of electron heating on electron capture cross section in very small metal-oxide-semiconductor transistors. Appl. Phys. Lett. 62*, 2233–2235.*

SIA - Semiconductor Industry Association (1997). The National Technology Road-Map for Semiconductors. *San Jose, CA: SIA - Semiconductor Industry Association.*

Siegert, G. R., P. Vitanov, and I. Eisele (1994). Observation of discrete energy levels of interface traps in sub-μm MOSFETs. Solid-State Electr. 137*, 1799–808.*

Siggia, E. D. and P. C. Kwok (1970). Properties of electrons in semiconductor inversion layers with many occupied subbands. I: screening and impurity scattering. Phys. Rev. B 2*, 1024–1036.*

Simoen, E., B. Dierickx, C. Claeys, and G. Declerck (1992). Explaining the amplitude of RTS noise submicrometer MOSFETs. IEEE Trans. on Electron Dev. 39*, 422–429.*

Singh, J. (1993). Physics of semiconductors and their heterostructures. *New York: McGraw-Hill.*

Singh, J. (1994). Semiconductor devices: an introduction. *New York: McGraw-Hill.*

Spinelli, A. S., A. Benvenutti, and A. Pacelli (1998). Self-consistent 2-D model for quantum effects in n-MOS transistors. IEEE Trans. on Electron Dev. 45*, 1342–1349.*

Stanford Research Systems (1996). SR830 DSP Lock-in: User's Manual. *Sunnyvale, California: Stanford Research Systems.*

Stern, F. (1972). Self-consistent results for n-type si inversion layers. Phys. Rev. B 5*, 4891–4899.*

Stern, F. and W. E. Howard (1967). Properties of semiconductor surface inversion layers in the electric quantum limit. Phys. Rev. 163*, 816–835.*

Stolk, P. A., F. P. Widdershoven, and D. B. M. Klaassen (1998). Modeling statistical dopant fluctuations in MOS transistors. IEEE Trans. on Electron Dev. 45*, 1960–1971.*

Strauch, D., A. P. Mayer, and B. Dorner (1990). Phonon eingenvectors in si determined by inelastic neutron scattering. Z. Phys. B- Condesed Matter. 78*, 405–410.*

Takagi, S.-I., A. Toriumi, M. Iwase, and H. Tango (1994). On the universality of inversion layer mobility in Si MOSFETs: Part i-effects of substrate impurity concentration. IEEE Trans. on Electron Dev. 41*, 2357–2362.*

Takahashi, Y., H. Namatsu, K. Kurihara, K. Iwadate, M. Nagase, and K. Murase (1996). Size dependence of the characteristics of the Si single electron transistor on SIMOX substrates. IEEE Trans. on Electron Dev. 43*, 1213–1217.*

Taniguchi, K. and C. Mamaguchi (1991). Granular Nanoelectronics. *New York: Plenum.*

Teranishi, N. and R. Kubo (1979). On impurity pinning of one-dimensional charge density waves. J. Phys. Soc. of Japan 47*, 720–728.*

Troger, C., H. Kosina, and S. Selberherr (1997). Modeling nonparabolicity effects in silicon inversion layers. In Proccedings SISPAD 97 Conf.*, Boston, pp. 232–326.*

Tsividis, Y. P. (1987). Operation and modeling of the MOS transistor. *New York: McGraw-Hill.*

van Vliet, C. M. (1994). Macroscopic and microscopic methods for noise in devices. IEEE Trans. on Electron Dev. 41*, 1902–1915.*

Villa, S., A. Lacaita, L. M. Perron, and R. Bez (1998). A physically-based model of the effective mobility in heavily-doped n-MOSFETs. IEEE Trans. on Electron Dev. 45*, 110–115.*

Wang, T., T.-E. Chang, L.-P. Chiang, C.-H. Wang, N.-K. Zous, and C. Huang (1998). Invetigation of oxide charge trapping and detrapping in a MOSFET using a GIDL current technique. IEEE Trans. on Electr. Dev. 45*, 1511–1517.*

Wasshuber, C. (1997). About single-electron devices and circuits. *Ph. D. thesis, Technische Universität Wien, Wien.*

Wasshuber, C. and H. Kosina (1998). Simulation of a single-electron tunnel transistor with inclusion of inelastic macroscopic quantum tunneling of charge. VLSI Design 6*, 35–38.*

Webb, R. A., A. B. Fowler, A. Hartstein, and J. J. Wainer (1986). Hopping conduction in quasi-one-dimensional systems. Surf. Sci. 170*, 14–27.*

Weis, J. (1994). Einzelelektron-Tunneltransistor: Transportspektroskopie der elektronischen Grund- und Anregungszustände in einem $GaAs/Al_xGa_{1-x}As$-Quantentopf. *Ph. D. thesis, Max-Planck-Institut für Festkörperforschung, Stuttgart.*

Welser, J. J., S. Tiwari, S. Rishton, K. Y. Lee, and Y. Lee (1997). Room temperature operation of a quantum-dot flash memory. IEEE Electron Dev. Lett. 18*, 278–80.*

Wolfram, S. (1996). Mathematica 3 standard add-on packages. *Champaign, Illinois: Wolfram Media.*

Wolfram, S. (1997). Das Mathematica-Buch: Mathematica Verison 3. *Bonn, Germany: Addison-Wesley-Longman.*

Wu, S. H. and R. L. Anderson (1974). MOSFETs in the O K approximation: Static characteristics of MOSFETs in the O K approximation. Solid-State Electronics 17*, 1125–37.*

Xue, W. and P. A. Lee (1988). Two-dimensional resonant tunneling. Phys. Rev. B 38*, 3913–3917.*

Yano, K., T. Ishii, T. Sano, T. Mine, F. Murai, T. Hashimoto, T. Kobayahi, T. Kure, and K. Seki (1999). Single-electron memory for giga-to-tera bit storage. Porceedings of the IEEE 87*, 633–651.*

Yokogawa-Hewlet-Packard. Ltd. (1986). HP4145B semiconductor parameter analyzer: operation and service manual. *Tokyo, Japan: Yokogawa-Hewlet-Packard. Ltd.*

Yokogawa-Hewlet-Packard. Ltd. (1995). HP4156 precision semiconductor parameter analyzer: User's dictionary reference. *Tokyo, Japan: Yokogawa-Hewlet-Packard. Ltd.*

Ytterdal, T., M. Hurt, M. Shur, H. Park, R. Tsai, and W. Peatman (1996). High-temperature characteristics of 2-D MESFETs. IEEE Electron Dev. Lett. 17*, 214–216.*

Publications

Wirth, G, R Wartchow, J Petterson and S Bampi (1994). *A Framework for Electrical Test Engineering Automation*, In: Proceedings of the 4th IFIP International Working Conference on Electronic Design Automation Frameworks, Gramado, Brazil, Nov 28-30, 165-174.

Müller, J, G Wirth, U Hilleringmann and K Goser (1996). *Analyses of sub 1/4-micron MOS-Transistors by visible light emission*, In: Proceedings of the 26th European Solid State Device Research Conference ESSDERC 96, Bologna, Italy, Sep 9-11, G Baccarani and M Rudan (Ed.), 947-950.

Wirth, G, U Hilleringmann, J Horstmann and K Goser (1997). *Periodic Transconductance Oscillations in Sub-100 nm MOSFETs*, In: Proceedings of the 27th European Solid State Device Research Conference ESSDERC 97, Stuttgart, Germany, Sep 22-24, H Grünbacher (Ed.), 380-383.

Wirth, G, U Hilleringmann, J Horstmann and K Goser (1997). *Transconductance Fluctuations in Ultrashort Channel MOSFETs*, In: Proceedings of the Second Workshop on Innovative Circuits and Systems for Nano Electronics, Delft, The Netherlands, Sep 29 - 30, A H M van Roermund and K Goser and D Schmitt-Landsiedel (Ed.), 1-4.

Wirth, G, U Hilleringmann and K Goser (1997). *Periodic Transconductance Oscillations in Sub-100 nm MOSFETs*, In: PHASDOM - Phantoms Strategic Domain Meetings (Abstracts), Aachen, Germany, March 10-13, D3.18c.

Wirth, G, U Hilleringmann, J Horstmann and K Goser (1998). *Periodische Leitwertschwankungen in sub-100 nm MOS-Transistoren*, In: ITG-Fachbericht 147: Mikroelektronik fr die Informationstechnik, VDE Verlag, Hannover, Germany, P Pirsch (Ed.), 177-181.

Wirth, G, U Hilleringmann, J Horstmann and K Goser (1998). *Mesoscopic Transport Phenomena in Ultrashort Channel MOSFETs*, In: Proceedings of the Third Workshop on Innovative Circuits and Systems for Nano Electronics, Munich, Germany, Oct 5 - 6, A H M van Roermund and K Goser and D Schmitt-Landsiedel (Ed.), B2/1-B2/6.

Wirth, G, U Hilleringmann, J Horstmann and K Goser (1999). *Mesoscopic Transport Phenomena in Ultrashort Channel MOSFETs*, Solid State Electronics 43, 1245-1250.

Wirth, G, U Hilleringmann, J Horstmann and K Goser (1999). *Negative Differential Resistance in Ultrashort Bulk MOSFETs*. To appear in: IECON 99 the 25th annual conference of the IEEE Idustrial Electronics Society, San Jose (CA), USA, Nov 29 - Dec 3.

Curriculum Vitae

	Gilson Inacio Wirth
	Born 1966 in Nova Hartz, Brazil
1985-1989	B.Sc. in electrical engineering at the Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
1990-1992	Electrical engineer at Embramic Ltda., Porto Alegre, Brazil
1992-1994	M.Sc. in computer science at the Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
1995-1999	Ph.D. in electrical engineering at the University of Dortmund, Dortmund, Germany