Chapter 1

Introduction

The invention of the integrated circuit in 1958 provided the basis for the microelectronics industry which is one of the foundations modern society is based on. The number of transistors per chip has been increased exponentially over time ("Moore's Law") until now. This trend, which had and still has revolutionary impacts on our civilization, is expected to continue in future. The increasing complexity is achieved by shrinking the geometrical dimensions of all devices and interconnects to fit more of them into the same area. While this process was mainly a technological challenge in the past, it is nowadays a physical challenge as well. As the devices are being shrunk, more and more parasitic effects become important, many of them being quantum mechanical phenomena. Unlike technological limitations, the laws of physics cannot be removed, and thus the question arises whether the conventional MOSFET, which is being used for more than 90 % of all integrated circuits, can be used in future at device dimensions of 30 nm and less.

The shrinking of the conventional MOSFET becomes more difficult the smaller its dimensions are. Due to physical limitations, the I_{on}/I_{off} ratio is reduced by the shrinking and since the on-current cannot be reduced arbitrarily it becomes more and more difficult to manufacture MOSFETs which can be *switched off*. Once the channel length approaches approximately 10 nm, direct quantum mechanical tunneling of electrons will be making this impossible.

A possible solution of this problem is to change the device concept by *using* the tunneling instead of trying to avoid it. This is the fundamental idea which forms the basis of the tunnel FET. Like the conventional

MOSFET, it has a gate stack and relies on the field effect. However, the semiconductor below the gate dielectric forms a p-i-n diode, instead of an n-p-n or p-n-p structure. This p-i-n diode is always reverse biased, which results in extremely low off-currents. Unlike the conventional MOSFET, the middle region (the i-part of the diode) does not form the barrier which is to be lowered by the gate voltage in order to switch on the device. Instead, the channel formed by the field effect below the gate dielectric causes an extremely high charge concentration gradient at the p-i or i-p junction (depending on the polarity of the gate voltage), which results in such a strong band bending that electrons can tunnel from the valence band to the conduction band. This tunneling current is the on-current of the tunnel FET. With degenerate p-doping and n-doping, the p-i-n structure represents the highest diffusion barrier that can be achieved with doping. This allows the tunnel FET to be scaled down much easier than the conventional MOSFET. The tunnel FET also has other advantages over the conventional MOSFET.

The device concept was originally proposed in III-V compounds and later on in silicon. Experimental prototypes have been demonstrated in lateral as well as in vertical structures. Albeit having very low leakage currents, they all share the problem of low on-currents which are far too low to match industrial requirements.

Many approximations can be made in order to analyze the conventional MOSFET analytically. Among all of them, the gradual channel approximation is one of the most important simplifications, which ignores the horizontal electric field and reduces the analysis into a one-dimensional mathematical problem, for which useful analytic solutions have been found. In contrast, it is the very nature of a p-i-n diode to show a strong horizontal electric field. Together with the gate field, the analytical problem becomes two-dimensional, and a simplification like the gradual channel approximation would fail. As long as no useful analytic mathematical models exist for the tunnel FET, two-dimensional computer device simulations have to be used.

By applying such simulations, a basic understanding of the tunnel FET has been developed especially in [1], where a simulation based model is developed to describe the current-voltage characteristics and to define the electrical parameters of the tunnel FET. Similar, but less extensive simulations can be found in [16], which also presents experimental data.

It is the aim of this work to experimentally verify the predictions in [1] about ideal tunnel FETs made of silicon and silicon-germanium (SiGe). The

ideal tunnel FET has a rectangular doping profile, meaning very abrupt junctions inside its p-i-n structure. Conventional manufacturing methods, which employ implantation and/or diffusion, are unable to satisfy this requirement sufficiently. Instead, epitaxy can be used in order to define very sharp doping profiles. The available equipment for this work included molecular beam epitaxy (MBE), used in [2,4,5,8,12,15] for example, and a commercially available CVD reactor, for which epitaxy and selective epitaxy (necessary for the proposal made at the end of this work) has been developed in [3]. The vertical transistor concept, which is explained in Chapter 5, has been employed in this work, because it allows to realize very low channel lengths with cheap photo lithography with 1 µm minimum feature size. Thus, this work can be seen as the continuation of [5].

Chapter 2 recapitulates important details about the basics of the conventional MOSFET, the understanding of which is still important for the tunnel FET.

Chapter 3 briefly deals with the scaling limitations of the conventional MOSFET, which represent the actual motivation for the tunnel FET.

Eventually, Chapter 4 summarizes most of the knowledge about the physics of the tunnel FET obtained in [1].

Chapter 6 contains the essence of this work. In order to satisfy the specifications demanded in [1], substantial improvements of the manufacturing process used for the vertical prototypes had to be made. Extensive efforts have been invested in order to improve the surface quality of the vertical channel surfaces, because this was the most efficient way to reduce the minimal thickness of the gate dielectric and thus the thermal budget. In addition, it was possible to build SiGe-prototypes for the first time.

Chapter 7 shows the electrical characteristics of the experimental results as well as SIMS spectra. In addition, the data is compared to the theoretical expectations obtained from the abovementioned device simulations, and explanations for deviations are given accordingly.

Finally, Chapter 8 concludes, that almost all theoretical predictions could be verified experimentally. In addition, a proposal is made for a lateral tunnel FET technology, which could overcome the inherent limits of the vertical concept that have been reached as a result of this work.

At the time of this writing, the tunnel FET is in an early stage of development. Many physical aspects are still unknown. But it

should be kept in mind, that the understanding and optimization of the conventional MOSFET took more than 30 years [75] and is still an ongoing process. The tunnel FET has a large potential to function at dimensions and conditions which render any conventional MOSFET inoperable.

Chapter 2

Conventional MOSFETs

This chapter reviews the characteristics, the scaling rules and the problems of the conventional MOSFET, taking the n-channel MOSFET as an example. Knowing the limits of this device is crucial to understand the motivation behind the tunnel FET, as it is mainly targeted as a replacement for the MOSFET by addressing certain important physical problems. A detailed discussion of this topic can be found in standard text books like [75] and [73].

2.1 MOS Capacitor

Figure 2.1 shows the well known schematic of a conventional n-channel MOSFET. Consider the MOS part (without drain and source) first for the sake of simplicity. The device is made from different materials with different work functions. Bringing these materials together leads to a separation of charge carriers in order to account for the work function differences. This builds up contact potentials between the different materials. For the net charges to disappear, an external voltage source must cancel the sum of all contact potentials:

$$\phi_{MS} = \phi_{\text{bulk material}} - \phi_{\text{gate material}} = \frac{1}{q}(W_M - W_S) \tag{2.1}$$

where W_S and W_M are the work functions of the semiconductor and the gate material (which was metal in the past), respectively. In order to compensate for the additional effective interface charge (per unit area)

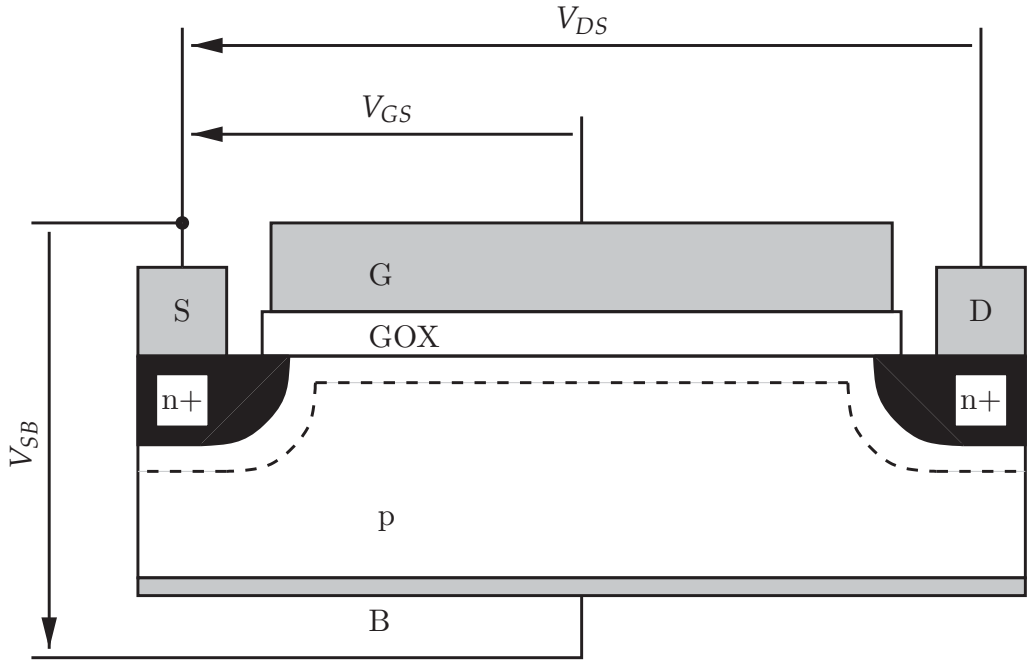

Figure 2.1: *Schematic of a lateral long channel MOSFET.*

Q'_0, a voltage

$$\psi_{OX} = -\frac{Q'_0}{C'_{OX}} \tag{2.2}$$

is required. Here, $C'_{OX} = \epsilon_{OX}/t_{OX}$ is the oxide capacitance per unit area, where t_{OX} is the thickness of the gate dielectric and ϵ_{OX} is its permittivity. The quantity

$$V_{FB} = \phi_{MS} + \psi_{OX} = \phi_{MS} - \frac{Q'_0}{C'_{OX}} \tag{2.3}$$

is called the flat band voltage, because it keeps the semiconductor everywhere neutral and thus the bands flat when applied between the gate and the substrate terminals.

When an external voltage V_{GB} is applied, charges are caused to appear in the semiconductor within a region adjacent to the top surface. The substrate outside this region can be assumed neutral. The charges lead to a voltage drop, defined from the surface to the bulk outside that region, which is called surface potential ψ_S:

$$V_{GB} = \psi_{OX} + \psi_S + \phi_{MS} \tag{2.4}$$

Increasing V_{GB} above V_{FB} results in a total charge on the gate, which is more positive than the value in flat band. This change must be balanced by the charge in the semiconductor under the oxide Q'_C, which is expressed by the bands being bent down. Holes are repelled from the surface, depleting the region beneath, whereas the resulting surface potential attracts electrons. With a sufficiently high V_{GB} the density of electrons will exceed that of holes at the surface, a situation which is called surface inversion. With

$$\begin{aligned} n_{\text{surface}} &= n_0 \exp(\phi_S/\phi_t) && (2.5)\\ n_0 &= n_i \exp(-\phi_F/\phi_t) && (2.6)\\ n_i &= p_0 \exp(-\phi_F/\phi_t) && (2.7)\\ p_0 &\approx N_A && (2.8) \end{aligned}$$

the electron concentration at the surface is

$$n_{\text{surface}} \approx N_A \exp\left(\frac{\psi_S - 2\phi_F}{\phi_t}\right) \tag{2.9}$$

From (2.6) it follows that at $\psi_S = \phi_F$ the surface electron concentration becomes equal to the intrinsic concentration. This is defined as the limit point between depletion and inversion. Note that n_{surface} is nonzero even in depletion. At $\psi_S = 2\phi_F$ the surface is inverted, as then $n_{\text{surface}} = p_0 \approx N_A$.

In order to describe Q'_C, Poisson's equation must be solved, in which y denotes the vertical distance from the surface:

$$\frac{d^2\psi}{dy^2} = \frac{-q}{\epsilon_S}\left(p_0 \exp\left(\frac{-\psi(y)}{\phi_t}\right) - n_0 \exp\left(\frac{\psi(y)}{\phi_t}\right) - N_A\right) \tag{2.10}$$

The solution is

$$\begin{aligned} Q'_C = &-\sqrt{2q\epsilon_S N_A}\left[\phi_t \exp\left(\frac{-\psi_S}{\phi_t}\right) + \psi_S - \phi_t + \right.\\ &\left. \exp\left(\frac{-2\phi_F}{\phi_t}\right)\left(\phi_t \exp\left(\frac{\psi_S}{\phi_t}\right) - \psi_S - \phi_t\right)\right]^{\frac{1}{2}} \end{aligned} \tag{2.11}$$

Certain approximations can be used to simplify this term for the regions of inversion. The charge sheet approximation assumes the inversion layer thickness to be negligible. This implies that practically all of the depletion region is free of electrons, which means that the surface

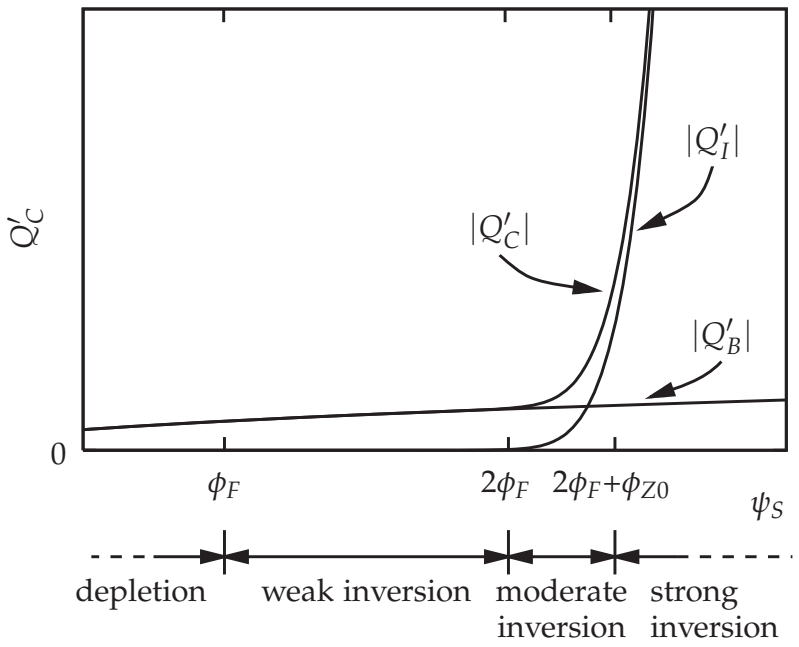

Figure 2.2: *Magnitude of inversion layer charge Q'_I, depletion region charge Q'_B, and their sum Q'_C (all per unit area) versus surface potential*

potential ψ_S drops across the depletion region only and not across the inversion layer. The second approximation, the depletion approximation, assumes the mobile carrier concentration in the depletion region to be negligible in comparison to the acceptor concentration. These assumptions lead to the expressions

$$Q'_B = -\sqrt{2q\epsilon_S N_A}\sqrt{\psi_S} \tag{2.12}$$

for the charge per unit area Q'_B which is due to the uncovered acceptor atoms in the depletion region, and

$$Q'_I = -\sqrt{2q\epsilon_S N_A}\left(\sqrt{\psi_S + \phi_t \exp\left(\frac{\psi_S - 2\phi_F}{\phi_t}\right)} - \sqrt{\psi_S}\right) \tag{2.13}$$

for the inversion layer charge per unit area Q'_I. These two quantities are plotted in Figure 2.2 together with their sum

$$Q'_C = Q'_B + Q'_I \tag{2.14}$$

The inversion region can be divided into the three subregions *weak*, *moderate* and *strong* inversion as shown in Figure 2.2. $\psi_S = \phi_F$ marks the onset of weak inversion, while its upper limit is usually defined

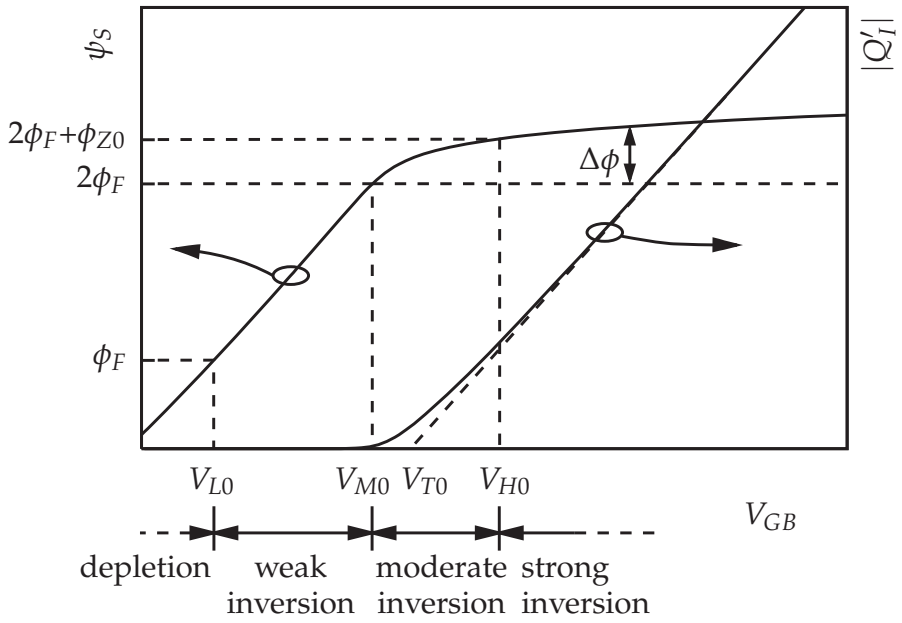

Figure 2.3: *Surface potential and magnitude of inversion layer charge per unit area versus gate-bulk voltage. V_{L0}, V_{M0} and V_{H0} denote the onsets of weak, moderate and strong inversion in terms of gate-bulk voltage.*

at $\psi_S = 2\phi_F$. Often, this point is also defined as the onset of strong inversion, but this cannot be justified. The onset of strong inversion is marked by $\psi_S = 2\phi_F + \phi_{Z0}$, where $\phi_{Z0} = 6\phi_t$ can be used as an average for uniform substrates. The region in between is the region of moderate inversion, where the threshold voltage (see below) can be found.

The transition from weak inversion to strong inversion has direct consequences for the transistor characteristics and is an important feature of the conventional MOSFET[1]. It is this region of moderate inversion, that allows the definition of the "threshold voltage" as one of the most important parameters of the conventional MOSFET, which separates the off- from the on-state. This is *not* possible for the tunnel FET (see Chapter 4).

The gate-bulk voltage and the surface potential are related by

$$V_{GB} = V_{FB} + \psi_S + \gamma\sqrt{\psi_S + \phi_t \exp\left(\frac{\psi_S - 2\phi_F}{\phi_t}\right)} \tag{2.15}$$

where

$$\gamma = \frac{\sqrt{2q\epsilon_S N_A}}{C'_{OX}} \tag{2.16}$$

[1]Unfortunately, it is also a challenge for developing good models.