Report
Fundamentals of Semiconductor Devices

Nikolaus Hammler
Hirose-Lab
The University of Tokyo

December 12, 2009

Contents

1 Task 1 - CV curves of MOS capacitors **2**
1.1 Low frequency case 2
1.2 High frequency case 2
1.3 Deep depletion 3

2 Task 2 - Flatband capacitance of a MOS capacitor **3**

3 Task 3 - Saturation in MOSFETs **5**

4 Task 4 - Comparision of 2-fold and 4-fold valleys **6**

1 Task 1 - CV curves of MOS capacitors

1.1 Low frequency case

Accumulation When a negative voltage on the gate V_g is applied, positive charges go to the oxide layer. This is called *accumulation*. When V_g is changed, also the charge distribution changes in the accumulation layer, therefore the whole MOS capacitor can be considered as a capaciator with the oxide as dielectricum:

$$C \approx C_{OX} \tag{1}$$

Therefore, the capacitance remains constant as it is not dependent on the gate voltage.

Depletion When the voltage is increased, until *depletion* is reached, negative charges go to the oxide layer and recombine with the positive charges. In this layer there are no free charges, therefore it is called depleted. The capacitance can then be described as a series of two capacitors: The oxide one and the capacitor from the depletion region:

$$C = \frac{1}{\frac{1}{C_{OX}} + \frac{1}{C_d}} \tag{2}$$

where C_d is the capacitance of the depletion layer which is in turn dependent on the width of the depletion layer. As the width is dependent on the the gate voltage, the capacitance is not constant.

Inversion When the voltage further increases there are no positive charges left to be recombined. This is called *inversion*. Additional negative charges can not be recombined and are available as free charges. These negative charges are produced by thermal generation which is a slow process. Because of the additional free charges, an n-doped region is created within the p-doped region. The capacitance is essentially the same as in the accumulation case but just inverted. The capacitance can therefore be described again with equation 1.

The low frequency case is described by the line a.

1.2 High frequency case

When high frequencies are used as AC bias, the process of generating negative charges for the inversion mode is too slow to follow the high AC frequency. For this reason the capacitance keeps low as the gate voltage is further increased in the inversion case whereas the curve keeps the same in accumulation and depletion condition. The capacitance can again be regarded as a series of two capacitors, the oxide and the capacitor for the depleted region:

$$C = \frac{1}{\frac{1}{C_{OX}} + \frac{1}{C_d}} \tag{3}$$

The minimum capacitance is reached when C_d in equation 3 reaches its minimum. The capacitor has its minimum when the length of the capacitor is as long as possible. So we seek for the maximum of the depletion width $x_{d_{\max}}$:

$$x_{d_{\max}} = \sqrt{\frac{4 \cdot \epsilon_0 \cdot \epsilon_S \cdot \phi_B}{q \cdot N_a}} \tag{4}$$

where this condition can be assumed when a depletion region is present, i.e. the surface potential at the interface of the oxide and the silicon does not exceed $2 \cdot \phi_B$:

$$2 \cdot \phi_B = \phi_S^{\text{inv}} \tag{5}$$

The high frequency case is described by the line b.

1.3 Deep depletion

The deep depletion corresponds a very fast DC sweep. In this case there is no time for generation of negative charges and therefore no inversion layer can be created. Instead, the width of the depletion layer can even be larger than $x_{d_{\max}}$ (equation 4) which in turn causes the capacitance to go down further. Because no minority charges can be created the capacitance remains constant in this condition.

The deep depletion case is described by the line c.

2 Task 2 - Flatband capacitance of a MOS capacitor

In this task we shall calculate the the flat band capacitance of a MOS capacitor described with:

- Oxide: SiO_2, 5 nm ($5 \cdot 10^{-9}$ m) thick, $\epsilon_{OX} = 3.9$
- Semiconductor: Si, acceptor concentration: $2 \cdot 10^{17}$ cm^{-3} ($2 \cdot 10^{23}$ m^{-3}), $\epsilon_{Si} = 11.9$
- Capacitance area: 10^{-4} cm^2 (10^{-8} m^2)
- Room temperature, i.e. 300 K

The physical constants are defined with:

- Boltzman constant: $k_B = 1.38 \cdot 10^{-23}$ J/K
- Elementary charge: $q = 1.6 \cdot 10^{-19}$ C
- Permittivity in vacuum: $\epsilon_0 = 8.85 \cdot 10^{-14}$ F/cm $= 8.85 \cdot 10^{-12}$ F/m

In the flat band case, the capacitance can be regarded as a connection in series of two capacitors, one from the oxide and one from the semiconductor. The capacitance is then calculated by:

$$C_{FB} = \frac{1}{\frac{1}{C_{OX}} + \frac{1}{C_S}} \tag{6}$$

A single capacitor is defined through the permittivity of the dielectricum, the width and the area:

$$C = \frac{A \cdot \epsilon_0 \cdot \epsilon_r}{d} \tag{7}$$

The first capacitance, C_{OX} can be calculated directly using formula 7:

$$\begin{aligned} C_{OX} &= \frac{A \cdot \epsilon_0 \cdot \epsilon_{OX}}{T_{OX}} \\ &= \frac{10^{-8} \cdot 8.85 \cdot 10^{-12} \cdot 3.9}{5 \cdot 10^{-9}} \\ &= \frac{8.85 \cdot 3.9}{5} \cdot 10^{-11} \\ &= 6.9030 \cdot 10^{-11} \text{ F} \\ &= 69.030 \text{ pF} \end{aligned} \tag{8}$$

The second capacitance, C_S can also be derived from formula 7 but we need to know the length of the dielectricum. This can be found with the Debye-Length:

$$L_D = \sqrt{\frac{\epsilon_0 \cdot \epsilon_{Si} \cdot k_B \cdot T}{q^2 \cdot N_a}} \tag{9}$$

When combining formula 7 and 9 we obtain:

$$\begin{aligned} C_S &= \frac{A \cdot \epsilon_0 \cdot \epsilon_{Si}}{\sqrt{\frac{\epsilon_0 \cdot \epsilon_{Si} \cdot k_B \cdot T}{q^2 \cdot N_a}}} \\ &= \frac{10^{-8} \cdot 8.85 \cdot 10^{-12} \cdot 11.9}{\sqrt{\frac{8.85 \cdot 10^{-12} \cdot 11.9 \cdot 1.38 \cdot 10^{-23} \cdot 300}{(1.6 \cdot 10^{-19})^2 \cdot 2 \cdot 10^{23}}}} \\ &= \frac{8.85 \cdot 11.9}{\sqrt{\frac{8.85 \cdot 11.9 \cdot 1.38 \cdot 300}{1.6^2 \cdot 2}}} \cdot 10^{-10} \\ &= 1.1412 \cdot 10^{-10} \text{ F} \\ &= 114.12 \text{ pF} \end{aligned} \tag{10}$$

Finally we add the results in equation 8 and 10 together using equation 6:

$$\boxed{C = \frac{1}{\frac{1}{69.030} + \frac{1}{114.12}} = 43.0123 \text{ pF}}$$

3 Task 3 - Saturation in MOSFETs

The output characteristics of a MOSFET can be described by the drain current vs. drain voltage diagram. This diagram consists of two areas:

1. The linear region (triode mode). In this region the MOSFET behaves like a resistor and the drain current increases linearily with drain voltage.

2. The saturation region (active mode). In this region the drain corrent does *not* increase when the drain voltage is increased. This mode is used for operating a MOSFET as a switch.

In this task we shall explain why the saturation region exists. I will describe the sitation with an nMOSFET as an example.

The basic idea with a MOSFET is to conduct a channel between source and drain which can be controlled with the gate voltage. If we apply a positive gate voltage, minority charges in the semiconductor drift up to the oxide interface where they recombine with free holes (majority charges). The majority charges are said to be pushed aside and the created zone is called *depletion zone*. If the gate voltage is further increased over a specific threshold voltage (i.e. $V_g > V_{th}$) there are no more holes free for recombination and the additional electrones are available for current transport. The region is said to be *inverted* because inside the p-doped semiconductor (where holes are the majority charges) an n-doped region is created where electrones are the majority charges. With this channel current can flow from source to drain. By variation of the gate voltage the size of the channel can be changed and therefore the current which can flow from source to drain.

Now we set the gate voltage to a constant value (but of course over V_{th}), connect the source to ground and vary the drain voltage in order to inspect the drain current. With low drain voltage the MOSFET is operated in linear mode and behaves more or less like a resistor. The drain current I_D in terms of the drain voltage V_D can be described through equation 11:

$$I_D = \frac{W}{L}\mu C_{OX}(V_G - V_{th} - \frac{V_D}{2})V_D \tag{11}$$

The channel potential is zero around the source and V_D around the drain, therefore the charges are not uniformily distributed, the channel is getting tapered. As the current flowing through the channel is constant (i.e. independent of the position) the electric field is also not uniformily distributed. Now the drain voltage is increased over a specific voltage $V_{D_{sat}}$, see equation 12. In this situation the amount of free charges at the drain

gets smaller and smaller and the gate voltage is too small to further hold the channel between source and drain.

$$V_{D_{sat}} = V_G - V_{th} \tag{12}$$

The location where the channel is broken is called *pinch off* and has the potential $V_{D_{sat}}$. The MOSFET operates in saturation.

Although there is no channel any more between source and drain the MOSFET can still conduct current. The potential drop between the pinch off point and the drain is $V_D - V_{D_{sat}}$ and the electrical field from the source to the pinch-off keeps the same. A strong electrical field exists between pinch-off and drain through which the electrones can drift from the channel to the drain. As a matter of fact the drain current I_D stays constant now although the drain voltage can be further increased if the drain voltage is higher than $V_{D_{sat}}$. The current in saturation mode can be described with equation 13 and is therefore independent of the drain voltage.

$$I_D^{sat} = \frac{W}{2L}\mu C_{OX}(V_G - V_{th})^2 \tag{13}$$

This situation can also be explained with two water bassins which are connected through a channel. The height of the bassins models the voltage (source or drain) and the height of the water between these bassins is controlled through the gate voltage. When one bassin is moved to a lower region the drop of the water lets the water accerelate and the water drops out of the bassin faster. The lower the bassin the faster the water drops out.

But this is only valid when the channel is up, i.e. the gate voltage is strong enough to built up a channel which is big enough. When the bassin gets too low the water just drops in the bassin and indirectly from the bassin out. Now the water does not accerelate any more and the water flows out at constant speed.

4 Task 4 - Comparision of 2-fold and 4-fold valleys

The subband energy of the **2-fold** valleys is lower because in the 2-fold valleys the electrons have a higher effective mass perpendicular to the silicon/oxide interface. This is because the the energy of an electron is indirectly proportional to the effective mass of an electron as can be seen in equation 14:

$$E = \frac{\hbar^2 k^2}{2m^*} \tag{14}$$

This also means that the inversion-layer is thinner. On the other hand, they have a lower conductivity mass in parallel to the interface whereas the oppisite is true for the 4-fold valleys.

The spatial extension of the wave function can be derived with the probability density function of the electrons, because the squared wave function gives the propability that an electron is at at specific location:

$$P(\text{electron at } z) = |\psi(z)|^2$$

The probability of finding an electron at a single position is much higher at lower energy levels, i.e. the probability is more concentrated and therefore the spatial extension is lower. As the 2-fold valleys have the lower energy the conclusion is that the spatial extension of the propability function and therefore of the wave function is smaller in the 2-fold valley case.

The results are very important for the performance of semiconductor devices. In semiconductor devices the drain current is an important measure of performance. The drain current can be derived as:

$$I_D = \frac{W}{L}\mu(V_g - V_{th}) \cdot V_d \tag{15}$$

The parameter μ in equation 15 is called the mobility and is defined as:

$$\mu = \frac{g_T}{m} \tag{16}$$

where g_T is the relaxation time and m is the conductive mass in parallel to the MOS interface. So for having a high drain current we need a high mobility and therefore a low conductive mass. This is true for the 2-fold valleys. This means that it is favorable to enlarge the electron occupancy in the 2-fold valleys which increases the total mobility which is the average of the mobility of all 2-fold and 4-fold valleys.