The MOSFET Small-Signal Model

Markus Hellenbrand
markus.hellenbrand@eit.lth.se

November 5, 2019

1 Introduction

This is an attempt at capturing all aspects of the small-signal model(s) that can be used to describe the high-frequency performance of the transistors in the *Nanoelectronics* group. The intention (of at least this first version) is not to provide detailed derivations of all properties, but to provide a basic understanding of the small-signal model. For this purpose, this document will begin with a basic description of the small-signal model, which captures the most important aspects of high-frequency MOSFET operation. This basic model, with small variations, often shows up in literature as the core of any small-signal model. Example measurements on a lateral nanowire MOSFET will demonstrate success and shortcomings of this model. Based on these observations, more detailed additional effects such as band-to-band tunnelling, impact ionisation, and the effect of border traps in the gate oxide will be added to the core model later on (work in progress). The effects of the different model components are illustrated by example measurements as well.

2 Calibration, Measurement, De-embedding

At high frequencies, it is not possible anymore to measure voltages and currents as in a DC measurement. Since at high frequencies (up to many gigahertz), the cables used in a measurement setup are in the order - or much larger - than the wavelength of the measured signals, they have to be treated as waves. This means that dedicated RF cables and probes have to be used in the measurement in order to avoid power losses in the signal. The response of the transistor is measured in the form of scattering parameters (s-parameters), which describe the transmission and reflection of the incident power waves at the terminals of the transistor. The theoretical analysis of the s-parameters is treated in chapter 4.

The measurement of these s-parameters is carried out with a Vector Network Analyser (VNA). All stages of the measurement setup, from the port of the VNA through the cables and the probes down to the sample, have to be impedance-matched in order to avoid reflections the intermediate interfaces. Different frequency ranges require different cable specifications and different measurement probes. Typically, common to all probes is the ground-signal-ground (GSG) configuration as illustrated in Fig. 1(a) with the schematic of a so-called infinity probe; the frequency range of cables is linked to their diameter, see Table 1 for an overview.

Table 1: Diameters, frequencies, and mating of RF cables.

Outer dielectric diameter	Frequency up to	Mates with
SMA	24 GHz	2.92 mm & 3.5 mm
3.5 mm	34 GHz	2.92 mm & SMA
2.92 mm	40 GHz	3.5 mm & SMA
2.4 mm	50 GHz	1.85 mm
1.85 mm	70 GHz	2.4 mm
1.0 mm	110 GHz	1.0 mm

In order to be able to measure the intrinsic response of a transistor, first, the reference plane of the measurement has to be moved from the port of the VNA to the tip of the measurement probes. This is achieved with a calibration

Figure 1: Schematics of (a) an infinity type RF probe, (b) the calibration structures used for an LRRM calibration (from left to right: load, short, through), and (c) open (left) and short (right) de-embedding structures to remove the effect of the transistor contact pads.

of the measurement setup with the help of dedicated calibration structures. Fig. 1(b) provides schematics of the structures, which are used for a load-reflect-reflect-match (LRRM) calibration. The reflection and transmission coefficients of these structures are well known and by measuring the structures and comparing the results with the known coefficients, the measurement reference plane can be moved accordingly. (The calculations are done by a software.)

Since the physical dimensions of transistors are much smaller than the probe tips, large metal pads have to be added on the transistor samples in order to be able to establish electrical contact between the transistor and the measurement setup. These metal pads add large capacitances and inductances to the RF response of the total structure so that the reference plane has to be move one more time to remove the effect of the pads. This process is called *de-embedding* and in the case of an *open-short* de-embedding it is carried out with the help of structures as the e.g. ones in Fig. 1(c). These de-embedding structures have to be exactly the same as the metal contact pads, except that the transistor is not present in the structure. As can be seen in Fig. 1(c), one structure represents an open circuit in the centre, and the other a short circuit. The measured response of the pads can be subtracted from the measured response of the complete structure (with transistor) so that the resulting response contains only the effect of the intrinsic transistor.

3 Small-Signal Parameters

As mentioned previously, the measured quantity in a high frequency characterisation are s-parameters, which describe the reflection and transmission of power waves incident to the transistor. For this characterisation method, the transistor is represented as a *two-port network*, where the gate corresponds to the input (port 1), the drain to the output (port 2), and the source is the common ground terminal. This representation is visualised in Fig. 2, where the a_j and b_j are the power flowing into and out of the ports, respectively, as

$$a_j = \frac{v_j + Z_0 i_j}{2\sqrt{Z_0}} \qquad \text{(1a)} \qquad\qquad b_j = \frac{v_j - Z_0 i_j}{2\sqrt{Z_0}}. \qquad \text{(1b)}$$

Here, v_j and i_j are the voltage and the current, respectively, at port j, and Z_0 is the characteristic impedance

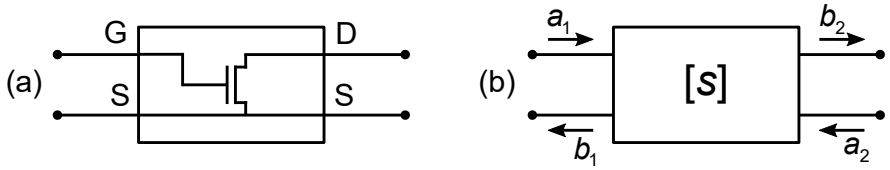

Figure 2: Two-port representation of a transistor. (a) Illustration of the port assignment. G, D, and S stand for gate, drain, and source, respectively. (b) Power input a and output b.

of the system. With (1a) and (1b), the complete two-port response can be described with the so-called $\boldsymbol{S}$-matrix as

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (2) \qquad \text{with} \qquad s_{kl} = \left.\frac{b_k}{a_l}\right|_{a_{\neg l}=0}. \quad (3)$$

For modelling, instead of s-parameters, admittance parameters (y-parameters) are much more intuitive, since they can be derived for a transistor small-signal model in a straightforward manner. Formally, the set of four y-parameters is completely equivalent to the set of s-parameters described by (2) and (3). The formal relation between the s-parameters and the y-parameters is [1]

$$y'_{11} = \frac{(1-s_{11})(1+s_{22}) + s_{12}s_{21}}{(1+s_{11})(1+s_{22}) - s_{12}s_{21}} \quad (4a)$$

$$y'_{12} = \frac{-2s_{12}}{(1+s_{11})(1+s_{22}) - s_{12}s_{21}} \quad (4b)$$

$$y'_{21} = \frac{-2s_{21}}{(1+s_{11})(1+s_{22}) - s_{12}s_{21}} \quad (4c)$$

$$y'_{22} = \frac{(1+s_{11})(1-s_{22}) + s_{12}s_{21}}{(1+s_{11})(1+s_{22}) - s_{12}s_{21}}. \quad (4d)$$

Note that the y'_{kl} in (4a) to (4d) are normalised with the characteristic impedance Z_0 of the system (usually 50 Ω) so that the unnormalised $y = y'/Z_0$.

While (4a) to (4d) are not intuitive yet, the definition of the y-parameters, which is "independent" of the s-parameters, is, since it relates the y-parameters to the physical properties of the transistor. Generally, the admittance y is the reciprocal of the impedance z, so that $y = 1/z = i/v$, where i is a current and v a voltage. In the small-signal model, i and v correspond to small AC values at the ports of the transistor around the DC large-signal bias point. The underlying and important assumption here is that the response of a transistor to a small AC signal can be approximated as linear around the DC bias point:

$$i_{\text{total}}(V_{\text{DC}} + \delta v_{\text{ac}}) \approx i_{\text{total}}(V_{\text{DC}}) + \frac{\partial i_{\text{total}}}{\partial v_{\text{ac}}}\delta v_{\text{ac}} \approx I_{\text{DC}} + i_{\text{ac}}. \quad (5)$$

Here, i_{total} is the total device current, uppercase letters with index DC denote (large-signal) DC values, and lowercase letters with index ac denote small AC variations. An example for the total source-to-drain current i_{DS} of a transistor as function of the gate and drain bias v_{GS} and v_{DS}, respectively, would thus become

$$i_{\text{DS}}(v_{\text{GS}}, v_{\text{DS}}) \approx I_{\text{DS}}(V_{\text{GS}}, V_{\text{DS}}) + \frac{\partial i_{\text{DS}}}{\partial v_{\text{gs}}}\delta v_{\text{gs}} + \frac{\partial i_{\text{DS}}}{\partial v_{\text{ds}}}\delta v_{\text{ds}} = I_{\text{DS}}(V_{\text{GS}}, V_{\text{DS}}) + g_{\text{m}}\delta v_{\text{gs}} + g_{\text{ds}}\delta v_{\text{ds}}. \quad (6)$$

Here, lowercase letters with uppercase indices denote quantities consisting of a (large-signal) DC and a (small-signal) AC part, g_{m} is the transconductance, and g_{ds} the output conductance. The other notations are as before.

In (6) we see two specific examples of admittances: the transconductance g_{m} and the output conductance g_{ds}. More generally, the admittances y_{kl} are complex values and their general definition, analogous to (3, is

$$y_{kl} = \left.\frac{i_k}{v_l}\right|_{v_{\neg l}=0} \quad (7)$$

with the currents i_k and voltages v_l at the respective ports k and l.

4 The Basic Small-Signal Model

With the general definition of (7), we can now approach the small-signal model for a MOSFET and derive the concrete y-parameters. A basic small-signal model, which captures the essential RF response of a MOSFET in the on-state, is provided in Fig. 3. The physical meaning of the different components will be explained below. First, for practicality, the equations for the y-parameters are provided close to the figure.

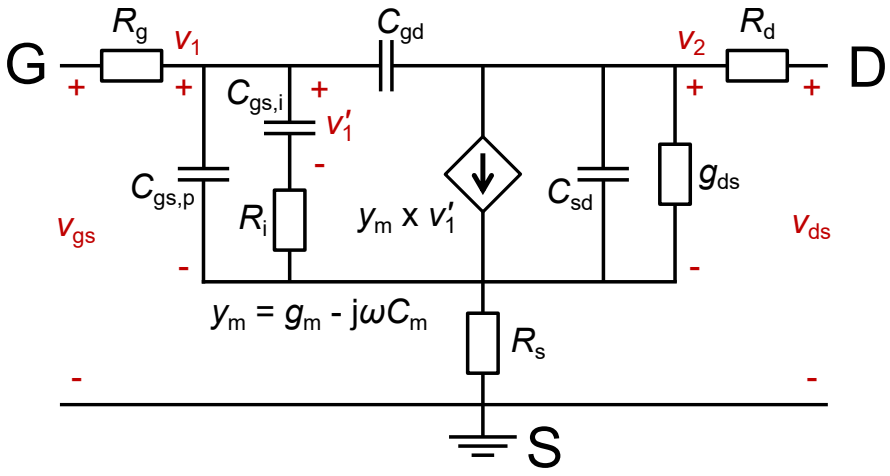

Figure 3: The most basic small-signal model that captures the essential intrinsic (and parasitic) high-frequency properties of a MOSFET.

4.1 The Intrinsic Model

The expressions for the y-parameters become much easier, if the gate, source, and drain resistances R_g, R_s, and R_d, are removed first. The procedure for this, as well as a way to determine their values, will be explained in 4.2. After removing the resistances, with (7), and with angular frequency ω and the imaginary unit j, the *intrinsic* y-parameters for Fig. 3 (with v_1 and v_2 marked) become

$$y_{11} = \left.\frac{i_1}{v_1}\right|_{v_2=0} = \mathrm{j}\omega\left(C_{\mathrm{gs,p}} + C_{\mathrm{gd}}\right) + \frac{\mathrm{j}\omega C_{\mathrm{gs,i}} + \omega^2 R_\mathrm{i} C_{\mathrm{gs,i}}^2}{1 + \omega^2 R_\mathrm{i}^2 C_{\mathrm{gs,i}}^2} \tag{8a}$$

$$y_{12} = \left.\frac{i_1}{v_2}\right|_{v_1=0} = -\mathrm{j}\omega C_{\mathrm{gd}} \tag{8b}$$

$$y_{21} = \left.\frac{i_2}{v_1}\right|_{v_2=0} = \frac{g_\mathrm{m} - \mathrm{j}\omega C_\mathrm{m}}{1 + \mathrm{j}\omega R_\mathrm{i} C_{\mathrm{gs,i}}} - \mathrm{j}\omega C_{\mathrm{gd}} \tag{8c}$$

$$y_{22} = \left.\frac{i_2}{v_2}\right|_{v_1=0} = g_{\mathrm{ds}} + \mathrm{j}\omega\left(C_{\mathrm{sd}} + C_{\mathrm{gd}}\right). \tag{8d}$$

The physical meaning of the different components is as follows:

- R_s, R_d, and R_g are the source, drain, and gate resistance respectively. They contain the contact and access resistance of the different terminals.
- $C_{\mathrm{gs,p}}$ is the parasitic gate-to-source capacitance. It originates in the physical extension of the gate metal, which forms parasitic capacitances with the source of the MOSFET.
- $C_{\mathrm{gs,i}} = -\partial Q_\mathrm{s}/\partial v_\mathrm{g}$ is the intrinsic gate-to-source capacitance, which models the movement of charge on the source side of the channel when the gate voltage is varying.
- R_i is the intrinsic channel resistance. It is not a resistance in the ohmic sense and it is not the same as the channel resistance as observed in DC measurements, either. Instead, its origin is the finite time that it takes for charge carriers to move through the channel. A short derivation is provided below this list.
- C_{gd} is the gate-to-drain capacitance. In this simplified model it contains both the intrinsic $C_{\mathrm{gd,i}} = -\partial Q_\mathrm{d}/\partial v_\mathrm{d}$ and the parasitic $C_{\mathrm{gd,p}}$. Both $C_{\mathrm{gd,i}}$ and $C_{\mathrm{gd,p}}$ have origins analogous to the gate-to-source capacitances. In the case of C_{gd}, the parasitic contribution often dominates, so that the intrinsic $C_{\mathrm{gd,i}}$

can be disregarded. For the same reason, an intrinsic resistance R_j, analogous to the intrinsic R_i, is disregarded. In the more detailed model later on, R_j will be included.

- For the controlled current source $y_\mathrm{m} = g_\mathrm{m} - \mathrm{j}\omega C_\mathrm{m}$, g_m is the transconductance and C_m the mutual differential capacitance, which balances the charge in the channel and takes into account delay in the transconductance.
- C_sd is the source-to-drain capacitance. The same explanation applies as for C_gd.
- g_ds is the output conductance of the transistor.

A Short Derivation of R_i [2]
In DC operation and at low frequencies, the movement of charge carriers in the transistor is much faster than the change in the bias conditions so that it can be assumed that the charge carriers inside the transistor adapt to a change in the bias conditions instantaneously. If the frequencies of operation approach the inverse of the time it takes for charge carriers to traverse the channel, this cannot be assumed any longer. The delay due to the finite velocity of the charge carriers can be modelled with R_i and the intrinsic capacitance $C_\mathrm{gs,i}$.

The time it takes charge carriers to traverse half the channel of a MOSFET is approximately $t_{1/2} = L_\mathrm{G}/(2v^+)$, where L_G is the gate length and v^+ is the average velocity of the charge carriers. This can be equated with the time $t_\mathrm{RC} = \ln(2)\tau = 0.7R_\mathrm{i}C_\mathrm{gs,i}L_\mathrm{G}W_\mathrm{G}$ that it takes to halfway charge the RC element consisting of R_i and $C_\mathrm{gs,i}$. W_G is the width of the transistor. Equating $t_{1/2}$ and t_RC yields $R_\mathrm{i} = 1/(1.4W_\mathrm{G}C_\mathrm{gs,i}v^+)$. This can be further related to the transconductance g_m since $g_\mathrm{m} = (\partial Q/\partial v_\mathrm{gs})v^+ + Q(\partial v^+/\partial v_\mathrm{gs}) \approx (\partial Q/\partial v_\mathrm{gs})v^+ \approx W_\mathrm{G}C_\mathrm{gs,i}v^+$, so that $R_\mathrm{i} \approx 1/(1.4g_\mathrm{m})$.

4.2 Off-State, Access Resistances, and Parasitic Capacitances

In order to be able to use the expressions for the intrinsic y-parameters in (8), the external resistances R_g, R_s, and R_d have to removed. One way to determine their values is a measurement of the s-parameters in the off-state of the transistor, a so-called *cold-FET measurement*. For such a measurement, if the gate voltage V_g is chosen far in the off-state of the transistor and the drain voltage is set to 0 V, the small-signal model of Fig. 3 reduces to that in Fig. 4. The elements, which are missing compared with Fig. 3, disappear because there is no current flowing through the device.

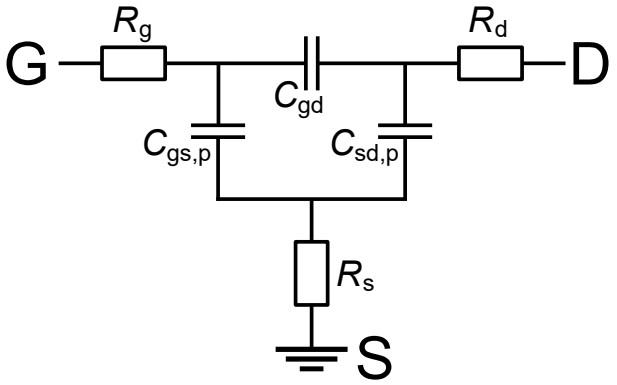

Figure 4: Small-signal model in the off-state.

At sufficiently high frequencies (above about a few GHz), the capacitances in Fig. 4 will short-circuit possibly still present intrinsic conductances so that the real part of the impedance will be dominated by the resistances R_g, R_s, and R_d. The real part of the matrix of impedance parameters (z-parameters) then becomes

$$\boldsymbol{Z_\mathrm{R}} = \mathrm{Re}(\boldsymbol{Z_{\textbf{off-state}}}) = \begin{bmatrix} R_\mathrm{s} + R_\mathrm{g} & R_\mathrm{s} \\ R_\mathrm{s} & R_\mathrm{s} + R_\mathrm{d} \end{bmatrix}, \tag{9}$$

so that the resistances can be read easily from plotting the real parts of the z-parameters. Just as for the y-parameters, the set of z-parameters is completely equivalent to the measured s-parameters, so that the z-parameters can be calculated directly from the measured s-parameters. For completeness, the corresponding

equations are listed below [1]:

$$z'_{11} = \frac{(1+s_{11})(1-s_{ss})+s_{12}s_{21}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} \tag{10a}$$

$$z'_{12} = \frac{2s_{12}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} \tag{10b}$$

$$z'_{21} = \frac{2s_{21}}{(1-s_{11})(1-s_{22})-s_{12}s_{21}} \tag{10c}$$

$$z'_{22} = \frac{(1-s_{11})(1+s_{22}+s_{12}s_{21})}{(1-s_{11})(1-s_{22})-s_{12}s_{21}}. \tag{10d}$$

Again, just as for the y-parameters, the z' are normalised to the characteristic impedance of the system so that the unnormalised $z = z'Z_0$.

An example of a (semi-successful) cold-FET measurement is presented in Fig. 5. As can be seen, the curves do not completely saturate at high frequencies, which can can be a sign of parasitic leakage paths, which are not taken into account in the simple model in Fig. 4. In this example, the obtained values for the resistances seem reasonable, which can be verified by comparison with DC measurements of the transistor. Sometimes, however, if parasitic leakages are too large, or the transistor does not turn off properly, the curves in Fig. 5 do not saturate at all. In that case, estimations from DC measurements are the only way to obtain approximate values for R_g, R_s, and R_d. In any case, comparison with the values obtained from DC measurements is good practice.

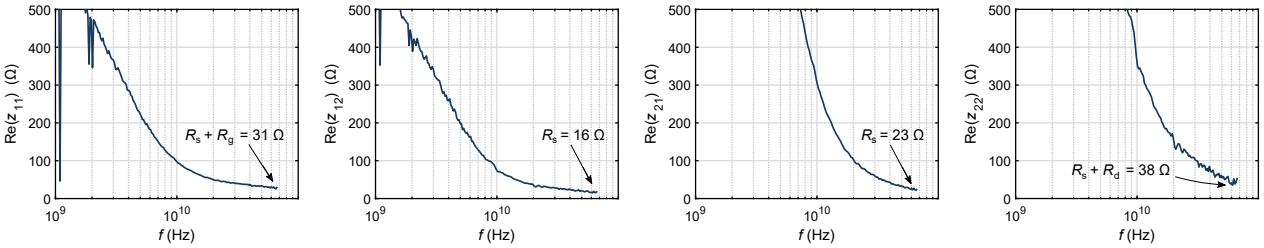

Figure 5: Real parts of the off-state impedances to determine R_g, R_s, and R_d.

Once R_g, R_s, and R_d are determined, they can be subtracted from the complete measured z-matrix of the transistor just as $\boldsymbol{Z}_\mathbf{removed} = \boldsymbol{Z}_\mathbf{measured} - \boldsymbol{Z}_\mathbf{R}$. In the next step, the z-parameters $\boldsymbol{Z}_\mathbf{removed}$ without the resistances can be transformed to y-parameters, which now only consist of the parasitic capacitances in Fig. 4 as

$$\boldsymbol{Y}_\mathbf{removed} = \begin{bmatrix} \mathrm{j}\omega(C_\mathrm{gs,p}+C_\mathrm{gd}) & -\mathrm{j}\omega C_\mathrm{gd} \\ -\mathrm{j}\omega C_\mathrm{gd} & \mathrm{j}\omega(C_\mathrm{gd}+C_\mathrm{sd}) \end{bmatrix}. \tag{11}$$

Once, values for the access resistances and the parasitic capacitances are obtained, the y-parameters from (8) can be used to determine the remaining small-signal parameters.

References

[1] W. Liu, *Fundamentals of III-V Devices.* Wiley Interscience, 1999. ISBN 0-471-29700-3

[2] E. Lind, "Lecture slides High-Speed Devices," January 2018.