EE2C11

Lab Report

D. Stijns
D.J.Stijns@student.tudelft.nl

R. van der Leer
R.C.vanderLeer@student.tudelft.nl

Submicrometer MOS parameter extraction

October 18, 2019

TUDelft

Abstract

Estimating the operation of a MOSFET by heart is an important skill for any engineer in digital integrated circuit design. This report describes a method for extracting the model parameters for the Unified Model from data simulated in a more robust and more intricate model, Berkeley Short-channel IGFET Model 3 (BSIM3).

We simulated the DC response of an N-channel enhancement-mode MOSFET with randomized parameters in Linear Technology spice (LTspice) with the BSIM3 model. From the gathered data, we derived the parameters necessary to model the transistor's behavior with the Unified Model. Additionally, we considered some methods that are used in the semiconductor industry to account for manufacturing tolerances.

We discovered that the quadratic model is reasonably accurate as a first-order approximation over the entire operating region except at the transition to velocity saturation.

1. Introduction

In the era of personal computers, doing calculation by hand may seem like an unnecessary skill. However, to understand the behavior of any system it is important to be able to reason about that system without needing the aid of a computer. Having the skill of manual calculation also allows an engineer to quickly evaluate a design in their head.

In order to do manual calculation, a model must be chosen such that it is simple enough to apply yet accurate enough to provide meaningful results. That is to say, the results should match the real behavior of a device.

A model that suits manual calculation of MOSFETs is the Unified Model (UM), also known as the Quadratic Model (QM). The UM predicts the DC characteristics of a MOSFET with a single quadratic equation. This model requires several parameters to be extracted from data before any calculation can be done. This data was acquired by using BSIM3, a more exhaustive model, in LTSpice. The MOSFET that was simulated had a channel length $L = 1.8\,\mu\text{m}$ and a channel width of $W = 180\,\text{nm}$. Certain parameters of the BSIM3 model were randomized in MatLab with a seed value of 4911210. The rest of the parameters are taken from the Predictive Technology Model website[4] The full parameter listing can be seen in Appendix A.

The method of extracting the parameters is discussed, the results are evaluated by comparing the the I_D-V_{DS} curves, and methods for mitigating process variability in production is also discussed.

2. Theory

The QM models a MOSFET by Formula 1. Here, W and L are the width and length of the modeled transistor and V_{min} is the minimum of $V_{GS} - V_T$, V_{DS} and V_{DSAT}. V_{GS} is the gate-to-source voltage and V_{DS} is the drain-to-source voltage. V_T, $k = k' \cdot \frac{W}{L}$, V_{DSAT}, and λ are the parameters to be determined.

$$I_D = \frac{W}{L} k' \left((V_{GS} - V_T) \cdot V_{min} - \frac{1}{2} V_{min}^2 \right) \cdot (1 + \lambda \cdot V_{DS}) \tag{1}$$

Since this is a quadratic formula it is relatively easy to solve for an unknown by hand.

BSIM3 has a myriad of formulas to model the same transistor with more accuracy. This model is part of a simulation computer program called Simulation Program with Integrated Circuit Emphasis (SPICE). There exist multiple variations of the original Berkeley SPICE. These all use the same BSIM3 model but may be extended with even more sophisticated models. This work only uses BSIM3, so which version of SPICE is used is irrelevant.

3. Extraction Method

To use the QM four parameters must be determined: V_T, k, λ and V_{DSAT}. We derive values for these parameters from measurements of the drain current at a certain fixed gate-to-source and another fixed drain-to-source voltage and a bulk-to-source voltage of 0 V.

3.1. Extracting V_T

V_T is an important parameter for describing the operation of the transistor [6][2]. As such, much care must be taken to find an appropriate method for extracting its value. One method is to linearly extrapolate $\sqrt{I_D}$ versus V_{GS} [2, section 2.3]. Dobrescu (2000) also argues that it is best to take the linearization at the point of inflection, the point with the maximum slope. The intersection between the linearization and the horizontal axis would be taken as the value of V_T.

It is also possible to use the same technique with a I_D versus V_{GS} plot [2, section 2.1]. This method produces a significantly different value for V_T which does not fit previously obtained values. The reason is, according to our interpretation, that the part that is extrapolated is not very linear when not taking the square root. This justifies the square root technique.

From the square root method, different values of V_T can be obtained by varying V_{DS}. This is not desirable, but can be explained with the DIBL-effect [7, section 3.3.3]. This effect reduces the apparent V_T when V_{DS} is increased. Therefore, the maximum apparent value found for V_T is used as the real value.

3.2. Extracting k

We will discuss two ways of determining k: using Formula 2 when $V_{min} = V_{GS} - V_T$ or using Formula 3 when $V_{min} = V_{DS}$. The use of Formula 2 might be more compelling when the extraction of V_T as mentioned in Section 3.1 has already been set up. This is because the maximum value of V_T relative to V_{DS} occurs at the inflection point where $V_{DS} = V_{GS} - V_T$; the boundary point between the linear and saturation regions. Here, V_{min} can be chosen as $V_{GS} - V_T$, and combining this with Formula 1 and deriving as shown in Appendix B.1 gives Formula 2 with the slope at the inflection point as its only variable.

$$k = 2 \cdot \left(\frac{\partial \sqrt{I_D}}{\partial V_{GS}} \right)^2 \quad (2)$$

The most significant limitation of Formula 2 is that it can only be used in the saturation region, which means the value of k obtained will be relatively high such that the QM will not be accurate in the linear region. To get a different, generally lower value of k, Formula 3 can be used, which is obtained from differentiating Formula 1 to V_{DS} with $V_{min} = V_{DS}$, at the point $V_{DS} = 0$.

$$k = \frac{\partial I_D}{\partial V_{DS}} \cdot (V_{GS} - V_T)^{-1} \quad (3)$$

The derivation of Formula 3 is in Appendix B.2. To use Formula 3, a value of V_{GS}, refferd to in this work as $V_{GS,k}$, must be chosen, and thus a single I_D-V_{DS} curve, with which to determine k. This choice of $V_{GS,k}$ determines around which V_{GS} the QM will be most accurate, so the choice can depend on the application, but it is commonly chosen around 1.5 V.

As said before, determining k using Formula 3 generally results in a lower value for k than Formula 2. This is because of an effect called surface scattering [9], which can be explained as the charge carriers being pulled toward and colliding with the gate oxide by high vertical electric field strength, caused by high V_{GS}. This reduces the mobility of the charge carriers, and since k depends on the mobility it is also affected. What this means is that one must use Formula 3 to determine k for a good modeling of the linear region and Formula 2 for the saturation region. However, as I_D (relative to V_{DS}) becomes linear in the saturation region and the slope in the saturation region is proportional to λ, there is not much to be gained by using Formula 2 to determine k for use in the saturation region. Therefore we will use Formula 3 to determine k, thus improving model accuracy in the triode region.

3.3. Extracting λ

Lambda can be calculated from the slope of I_{DS} versus V_{DS} in the saturation region: $\lambda = \frac{\partial I_D}{\partial V_{DS}}$.

When one linearly extrapolates I_D versus V_{DS} from the saturation region, this line intersects the V_{DS}-axis at a voltage called V_E, the early voltage. From assuming $I_{DS} = 0$ in Formula 1, it follows that $V_E = \frac{-1}{\lambda}$ The early voltage is independent of V_{GS}, because it is the axis of all extrapolated lines. In other words, if V_{GS} is larger the slope of the line is larger, but the offset is also larger. These two effects should cancel each other out according to the QM.

In reality, V_E and therefore λ are dependent of V_{GS}. Since we are modelling digital circuits, it is best if our model is accurate when either $V_{GS} = V_{DD}$, or $V_{GS} = 0$.

3.4. Extracting V_{DSAT}

It is difficult to determine V_{DSAT} mathematically or algorithmically. However, at this point all other essential parameters for the Unified Model have been determined. It is therefore possible to determine V_{DSAT} by inspection. That is, trying different values to get the best overall fit.

4. Extraction Results

The extracted parameter values are in Table 1. A comparison between the BSIM3 simulation and the Unified Model with our extracted parameters is shown in Figure 1.

The QM roughly follows the BSIM3 model, which is to be expected. The two models diverge significantly around the transition between the linear and velocity saturation regions. The curve of the QM has a sharp bend at the point where $V_{DS} - V_T$ becomes greater than V_{DSAT}. The BSIM3 curve is continuous in this transition point. Of course, BSIM3 is more accurate; this is clearly a limitation of the QM. This is because velocity saturation is very roughly modeled in the QM, which makes it impossible to have a continuous curve in the transition point.

One could ask: is velocity saturation occurring at all, given that including velocity saturation makes the outcome of the QM and BSIM3 diverge around the transition point?

Answering this question does not have to be complicated: if velocity saturation is not included ($V_{DSAT} \geq V_{DD}$), the parameters in Table 1 in combination with the QM do not produce a set of I_D-V_{DS} curves resembling those of BSIM3 for any value of k. Furthermore, it is known that velocity saturation does occur in devices with a drawn length of 180 nm, such as the one this work is about [3]. In conclusion, the discrepancy around the transition between linear mode and velocity saturation mode is a limitation of the QM model and not an error of the determined parameters themselves.

The Unified Model in Figure 1 has a curve in saturation mode that best matches with the $V_{GS} = 1.5$V

curve. This makes sense, because k was determined from the data set where V_{GS} is equal to 1.5 V. For analyzing digital circuits, it is important to have a model that works well for the case where the supply voltages are applied. One should consider a different set of parameters when analyzing analog circuits.

V_T	0.373 V
k	2.339 mA/V^2
λ	0.185 V^{-1}
V_{DSAT}	0.39 V

Table 1: Extracted Unified Model Parameters

4.1. Further optimization

The parameters determined in Section 4 produce a fairly good result when compared to BSIM3. It is possible to adjust $V_{GS,k}$ and V_{DSAT} in order to decrease the deviation from BSIM3 in the saturation region, but this has as a greater deviation in the triode region as a trade-off.

This tweaking is done by decreasing $V_{GS,k}$ which results in an increase in k, which scales I_D linearly. At the same time, decreasing V_{DSAT} decreases the quadratic increase of I_D with V_{GS}. This way, the spacing between the I_D curves in the saturation region is slightly decreased, placing them closer to the BSIM3 curves. Table 2 contains the parameters after manual optimization. Figure 2 shows the result of the optimization: closer resemblance in the saturation region but greater deviation in the triode region.

V_T	0.373 V
k	2.968 mA/V^2
λ	0.185 V^{-1}
V_{DSAT}	0.29 V

Table 2: Parameters after manual optimization

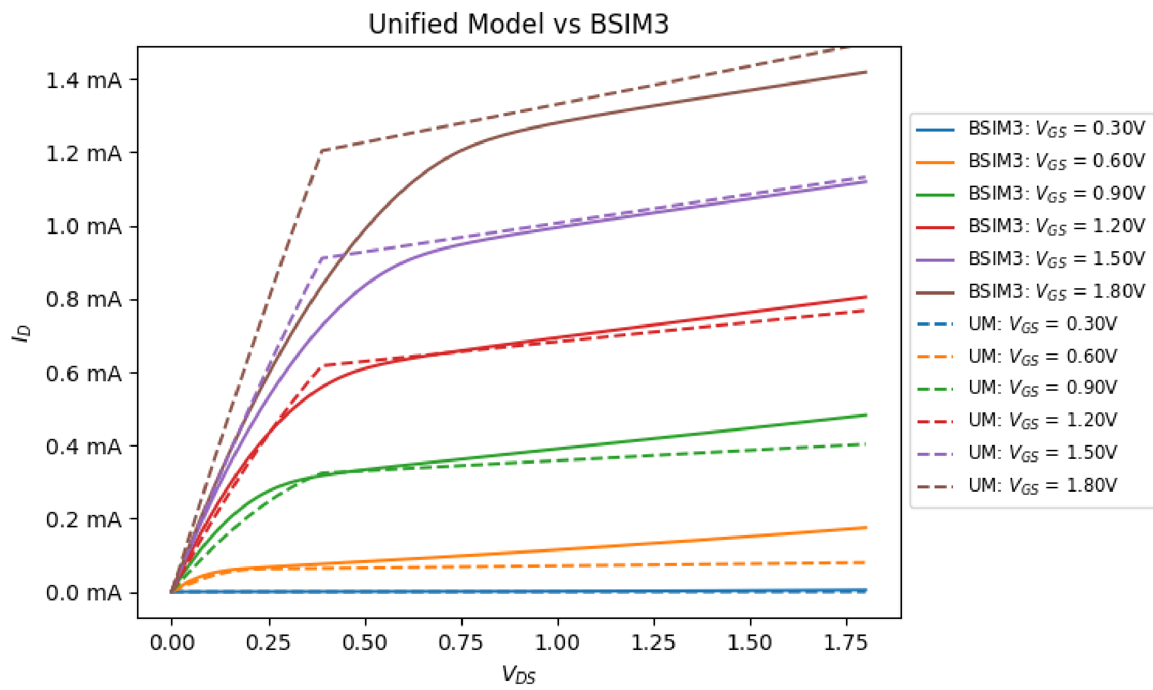

Figure 1: Comparison between BSIM3 and UM with parameters from Table 1

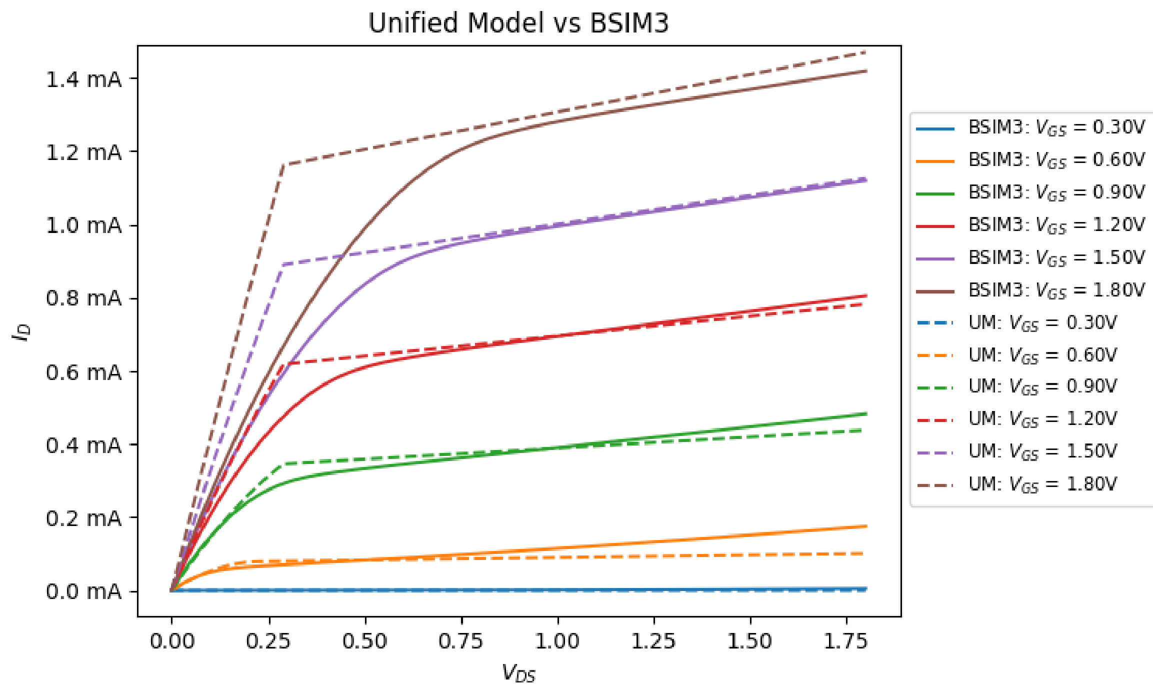

Figure 2: Comparison between BSIM3 and UM with optimized parameters from Table 2

5. Accounting for Variability

Modern ICs are designed with feature sizes well below the 180 nm scale presented in this work. These ICs have feature sizes of less than 50 nm. At this scale, statistical variability becomes significant. The variability affects the yield and performance of ICs. There is a distinction between two types of variability: random and systematic. Both of these have an effect on the parameters of a MOSFET in any IC with sub 50 nm sizes [5].

When designing with small MOSFETs, an engineer must take this variability into account. This is done with the aid of a computer simulator. There exist multiple types of simulators with a trade-off between speed and accuracy. Two examples are Statistical Compact Models (SCMs) and atomistic models (AMs). An SCM uses analytic equations based on empirical data to model variability and an AM uses more universal physics-based equations [5]. An SCM can model the standard deviation of propagation delay with about 4% accuracy if the atomistic model is assumed to be perfect.

When using any statistical simulator, variations must first be characterized. This is done with a Characterisation Array (CA) [8]. The CA contains a grid of transistors which can each be measured by probing the pins of the CA IC. From this, data is extracted as a basis for statistical computing.

SCMs can be used inside a design flow [1]. This allows a designer to consider the effect of variation while designing an IC and make a better informed choice. For example, when a certain logical block's performance affects the throughput of the entire circuit, it makes sense to ensure that block will have low variability at the cost of area. Where tolerances are larger, the parts can be optimized for size instead.

6. Conclusion

In this work, the parameters of a BSIM3-simulated 180 nm transistor were extracted for use in the Universal Model. The method for extracting these parameters were discussed and the two models were compared. Finally, the methods used to account for variability in current IC design were discussed.

The Universal Model and BSIM3 have been found to differ significantly in the region where velocity saturation begins, but otherwise the Universal Model has been found to be a reasonable one for a first order approximation.

Bibliography

[1] B. Cheng, D. Dideban, N. Moezi, C. Millar, G. Roy, X. Wang, S. Roy, and A. Asenov. Statistical-variability compact-modeling strategies for bsim4 and psp. *IEEE Design Test of Computers*, 27(2):26–35, March 2010. doi: 10.1109/MDT.2010.53.

[2] Lidia Dobrescu, M. Petrov, D. Dobrescu, and C. Ravariu. Threshold voltage extrection methods for mos transistors, 2000.

[3] Kiran Agarwal Gupta, Dinesh K. Anvekar, and V. Venkateswarlu. Comparative study and analysis of short channel effects for 180nm and 45nm transistors. In Natarajan Meghanathan, Dhinaharan Nagamalai, and Nabendu Chaki, editors, *Advances in Computing and Information Technology*, pages 707–715, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-31600-5.

[4] Arizona State University Nanoscale Integration and Modeling Group. Predictive technology model. URL `http://ptm.asu.edu`.

[5] H. Jooypa and D. Dideban. Embedding statistical variability into propagation delay time compact models using different parameter sets: A comparative study in 35-nm technology. *IEEE Transactions on Electron Devices*, 65(7):2714–2720, July 2018. doi: 10.1109/TED.2018.2833879.

[6] A. Ortiz-Conde, F.J. García Sánchez, J.J. Liou, A. Cerdeira, M. Estrada, and Y. Yue. A review of recent mosfet threshold voltage extraction methods. *Microelectronics Reliability*, 42:583–596, 04-05 2002. doi: 10.1016/S0026-2714(02)00027-6A.

[7] Jan M. Rabaey, Anatha Chandrakasan, and Borivoje Nikolic. *Digital Integrated Circuits*. Pearson, 2002.

[8] M. Simicic, V. Putcha, B. Parvais, P. Weckx, B. Kaczer, G. Groeseneken, G. Gielen, D. Linten, and A. Thean. Advanced mosfet variability and reliability characterization array. In *2015 IEEE International Integrated Reliability Workshop (IIRW)*, pages 73–76, Oct 2015. doi: 10.1109/IIRW.2015.7437071.

[9] N.P. van der Meijs. *EE2C11 Integrated Circuits Course Lab 1*. TU Delft, 2019.

A. BSIM3 Parameters

```
*$
* Predictive Technology Model Beta Version
* 180nm NMOS SPICE Parametersv (normal one)
* BSIM3 Version 3.1
* From http://ptm.asu.edu/latest.html, 2018-09
* Cao, Yu, et al. "New paradigm of predictive MOSFET and interconnect
* modeling for early circuit simulation." Custom Integrated Circuits
* Conference, 2000. CICC. Proceedings of the IEEE 2000. IEEE, 2000
*
* Modified for TU Delft EE2C11 by NvdM, 2018-09
* removed Tref, binflag, php parameters
* BSIM3v3
* PSpice and LTSpice
* PSpice: level = 7
* LTSpice: level = 8
*
* Modified by Daniel Stijns, 2019-10
* randomized u0, Tox, Lint


.model nenh NMOS
+Level = 8

+Lint = 4.0438e-08 Tox = 4.41702e-09
+Vth0 = 0.3999 Rdsw = 250

+lmin=1.8e-7 lmax=1.8e-7 wmin=1.8e-7 wmax=1.0e-4 version =3.1
+Xj= 6.0000000E-08          Nch= 5.9500000E+17
+lln= 1.0000000             lwn= 1.0000000              wln= 0.00
+wwn= 0.00                  ll= 0.00
+lw= 0.00                   lwl= 0.00                   wint= 0.00
+wl= 0.00                   ww= 0.00                    wwl= 0.00
+Mobmod=  1                 binunit= 2                  xl=  0
+xw=  0
+Dwg= 0.00                  Dwb= 0.00

+K1= 0.5613000                K2= 1.0000000E-02
+K3= 0.00                   Dvt0= 8.0000000             Dvt1= 0.7500000
+Dvt2= 8.0000000E-03        Dvt0w= 0.00                 Dvt1w= 0.00
+Dvt2w= 0.00                Nlx= 1.6500000E-07          W0= 0.00
+K3b= 0.00                  Ngate= 5.0000000E+20

+Vsat= 1.3800000E+05        Ua= -7.0000000E-10          Ub= 3.5000000E-18
+Uc= -5.2500000E-11         Prwb= 0.00
+Prwg= 0.00                 Wr= 1.0000000               U0= 3.5951000E-02
+A0= 1.1000000              Keta= 4.0000000E-02         A1= 0.00
+A2= 1.0000000              Ags= -1.0000000E-02         B0= 0.00
+B1= 0.00

+Voff= -0.12350000           NFactor= 0.9000000           Cit= 0.00
+Cdsc= 0.00                 Cdscb= 0.00                 Cdscd= 0.00
+Eta0= 0.2200000            Etab= 0.00                  Dsub= 0.8000000

+Pclm= 5.0000000E-02        Pdiblc1= 1.2000000E-02      Pdiblc2= 7.5000000E-03
```

```
+Pdiblcb= -1.3500000E-02    Drout= 1.7999999E-02    Pscbe1= 8.6600000E+08
+Pscbe2= 1.0000000E-20      Pvag= -0.2800000        Delta= 1.0000000E-02
+Alpha0= 0.00               Beta0= 30.0000000

+kt1= -0.3700000            kt2= -4.0000000E-02     At= 5.5000000E+04
+Ute= -1.4800000            Ua1= 9.5829000E-10      Ub1= -3.3473000E-19
+Uc1= 0.00                  Kt1l= 4.0000000E-09     Prt= 0.00

+Cj= 0.00365                Mj= 0.54                Pb= 0.982
+Cjsw= 7.9E-10              Mjsw= 0.31
+Cta= 0                     Ctp= 0                  Pta= 0
+Ptp= 0                     JS=1.50E-08             JSW=2.50E-13
+N=1.0                      Xti=3.0                 Cgdo=2.786E-10
+Cgso=2.786E-10             Cgbo=0.0E+00            Capmod= 2
+NQSMOD= 0                  Elm= 5                  Xpart= 1
+Cgsl= 1.6E-10              Cgdl= 1.6E-10           Ckappa= 2.886
+Cf= 1.069e-10              Clc= 0.0000001          Cle= 0.6
+Dlc= 4E-08                 Dwc= 0                  Vfbcv= -1
*
*$
```

B. Formula Derivation

B.1. Formula 2

$$I_D = \frac{1}{2}k(V_{GS} - V_T)^2$$

$$\sqrt{I_D} = \sqrt{\frac{1}{2}k} \cdot (V_{GS} - V_T)$$

$$\frac{\partial\sqrt{I_D}}{\partial V_{GS}} = \sqrt{\frac{1}{2}k}$$

$$k = 2 \cdot \left(\frac{\partial\sqrt{I_D}}{\partial V_{GS}}\right)^2$$

B.2. Formula 3

$$I_D = k\left((V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2\right)$$

$$\frac{\partial I_D}{\partial V_{DS}} = k(V_{GS} - V_T) - V_{DS}$$

Assume $V_{DS} = 0$

$$k = \frac{\partial I_D}{\partial V_{DS}} \cdot (V_{GS} - V_T)^{-1}$$