A Simple and Unambiguous Definition of Threshold Voltage and Its Implications in Deep-Submicron MOS Device Modeling

X. Zhou, K. Y. Lim, and D. Lim

***Abstract*—A new definition of MOSFET threshold voltage is proposed, namely, the "critical-current at linear-threshold" method, which has a unique solution and is very simple to measure. This definition gives consistent values of threshold voltage for different regions of operation at long channel, and contains the information on short-channel effects at short channel, which is very useful for deep-submicron MOS device characterization and modeling. The proposed method effectively removes ambiguity of *de facto* industry standard of the constant-current method for MOS threshold voltage.**

***Index Terms*— Deep-submicron MOSFET modeling, DIBL, MOS threshold voltage.**

I. Introduction

The threshold voltage (V_t) is a key parameter in MOSFET design and modeling. There are numerous definitions and extraction methods [1]–[6], each of them is proposed with a focus on different aspects. There are three major criteria for the definition and extraction of V_t: simplicity, unambiguity, and consistency, which should apply to various gate lengths (L_g) and operating conditions (linear or saturation).

The theoretical definition of V_t is based on the "strong-inversion" condition at which the surface potential is twice of the bulk Fermi potential ($\phi_s = 2\phi_B$)

$$V_t = V_{\rm FB} + 2\phi_B + \gamma\sqrt{2\phi_B - V_{\rm bs}} \tag{1}$$

where $V_{\rm FB}$ is the flat-band voltage and γ is the body factor. This definition is not practical for measurement since ϕ_s is not a measurable parameter. The constant-current (CC) method, although not physical, is the simplest and most popular to measure V_t. The major drawback is the arbitrary choice of the critical drain current, I_{d0} (usually scaled by W/L_g), at which the value of V_t is measured. Another popular definition is the maximum-g_m method in which the drain current is linearly extrapolated to zero at maximum transconductance. This method is unambiguous but only valid for linear region of operation (low $V_{\rm ds}$). It, however, depends strongly on S/D series resistance. For a MOSFET operating in the saturation region (high $V_{\rm ds}$), the drain current is quadratically extrapolated to zero (at maximum slope) to obtain the saturation threshold voltage, $V_{\rm tsat}$. For intermediate drain voltages, neither method is appropriate since the device operation crosses different regions for the full range of $I_{\rm ds} - V_{\rm gs}$. There are other less popular methods such as the second-derivative method [4] and the ratio method [5], which are developed to avoid the dependence on the series resistance. However, in our opinion, V_t definition, measurement, extraction and modeling are different things. One may define V_t with very simple measurement (such as the CC method) and develop a model to account for small-geometry effects and bias dependencies [6]; or define V_t with all the short-channel effects built in, but it usually requires more complex procedures for measuring V_t [3]. In any case, V_t modeling (formulation) is dependent upon its definition. In addition, it does not make justice to gauge the accuracy of a V_t model against the "strong-inversion" condition [3] since it is known [7] that for short-channel devices the surface-potential barrier is lower than that of the long-channel one.

With technologies going into the deep-submicron regime, it is increasingly important to model the short-channel effects such as V_t roll-off (at decreasing L_g) and drain-induced barrier-lowering (DIBL) (at increasing $V_{\rm ds}$). Hence, a simple, unambiguous, and consistent definition of V_t for all values of L_g and $V_{\rm ds}$ is desirable. In this brief, we propose a new threshold-voltage definition, namely, the "critical-current at linear-threshold" (or "$I_{\rm crit}@V_{t0}$" for short), which combines the unambiguity of the maximum-g_m definition and the simplicity of the CC definition. The method applies consistently to all values of L_g and $V_{\rm ds}$. In this brief, all data presented are measured from the same 0.25-μm CMOS process wafer. Complete extraction and modeling of V_t based on this definition, including gate length, drain- and substrate-bias dependencies, has been presented elsewhere [6].

II. Definition, Measurement, and Discussions

In the "$I_{\rm crit}@V_{t0}$" definition of V_t, the simple CC definition is adopted for any given L_g and $V_{\rm ds}$. However, the critical current is uniquely defined as the drain current when the gate voltage is the threshold voltage from the maximum-g_m definition in linear region: $I_{\rm crit} = I_{\rm ds}$ ($@V_{\rm gs} = V_{t0}$). Although it appears to have no direct relationship between the maximum-g_m definition and the theoretical "strong-inversion" definition of V_t, it has been shown [8] through two-dimensional (2-D) numerical simulations that the two definitions give the same threshold voltages for long-channel uniformly-doped MOSFET's. This provides a basis for the proposed V_t definition. This definition solves the problem of arbitrary choice of the critical current in the conventional CC method while retaining the simple definition and measurement of V_t at different $V_{\rm ds}$. It also provides consistent values of long-channel V_t in linear and saturation regions (since DIBL effect is minimal at long channel), which may not be true with a combined linear- and quadratic-extrapolation method. If this definition is to be applied to numerical simulations, V_t can be easily obtained based on the CC extraction algorithm.

With this definition, V_t measurement is simple. Only the linear $I_{\rm ds} - V_{\rm gs}$ curve (at some small $V_{\rm ds}$) is required to extract V_{t0} at maximum slope (g_m), as well as measuring the drain current $I_{\rm ds} = I_{\rm crit}$ when $V_{\rm gs} = V_{t0}$. The threshold voltage at any other drain bias is *defined* (and measured) as the gate voltage when $I_{\rm ds} = I_{\rm crit}$ (for the same transistor).

Fig. 1 plots the critical drain current against drawn gate length at $V_{\rm ds} = 0.1$ V extracted from the same wafer based on the above definition, together with the constant current scaled by $I_{d0} \times (W/L_g)$ where I_{d0} is chosen to be 0.1 μA and $W = 20$ μm. The same data is plotted on logarithmic scales in the inset of Fig. 1. It is observed that $I_{\rm crit}@V_{t0}$ has a similar $1/L_g$ dependence at longer channel lengths, but changes to a sub-$(1/L_g)$ dependence for short-channel devices. This is exactly due to the increased contribution of the S/D series resistance at shorter channel lengths, a fact commonly considered as a drawback of the maximum-g_m definition of V_t. However, it can be considered as an advantage since the threshold voltage defined this way contains the information on the 2-D short-channel effects. A technique for direct measuring the effective channel length and series resistance of submicron MOSFET's has been developed based on this V_t definition [9]. A clear (if not "fatal") drawback of the

Manuscript received September 8, 1998; revised November 2, 1998. The review of this brief was arranged by Editor C. Y. Yang.

X. Zhou and K. Y. Lim are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

D. Lim is with Chartered Semiconductor Manufacturing Ltd., Singapore 738406.

Publisher Item Identifier S 0018-9383(99)02639-8.

0018–9383/99$10.00 © 1999 IEEE

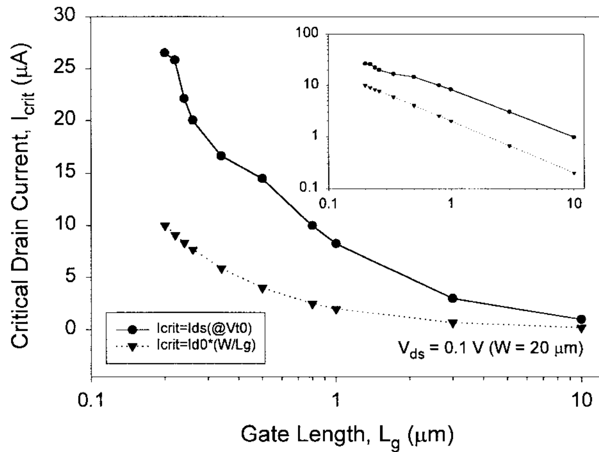

Fig. 1. Critical drain current (•) measured at $V_{gs} = V_{t0}$ against drawn gate length, where V_{t0} is the linear threshold voltage based on the maximum-g_m definition. Constant current (▲) is also plotted as a comparison. The inset shows the same data on a log–log scale.

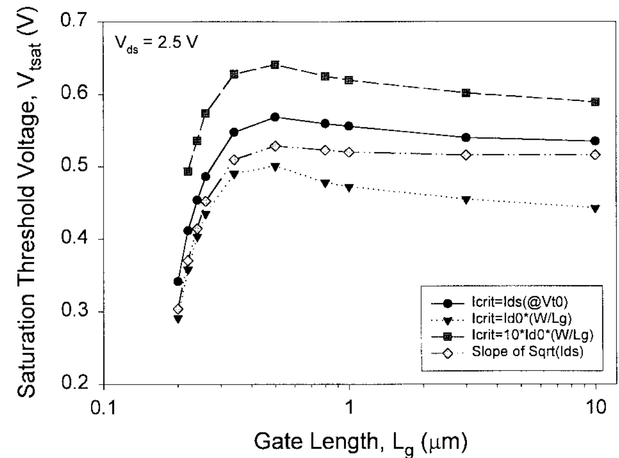

Fig. 3. Measured saturation threshold voltages against drawn gate length for the "I_{crit} @V_{t0}" definition (•), the CC definition with $I_{d0} = 0.1$ μA (▲) and 1 μA (■), and the quadratic-extrapolation definition (♦).

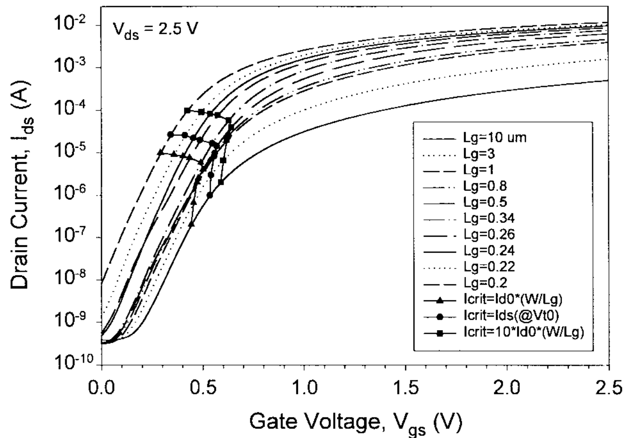

Fig. 2. $I_{ds} - V_{gs}$ curves in saturation ($V_{ds} = 2.5$ V) for different gate-length devices as indicated. The critical currents and the corresponding threshold voltages are shown in symbols for the "I_{crit} @V_{t0}" definition (•) and the CC definition with $I_{d0} = 0.1$ μA (▲) and 1 μA (■).

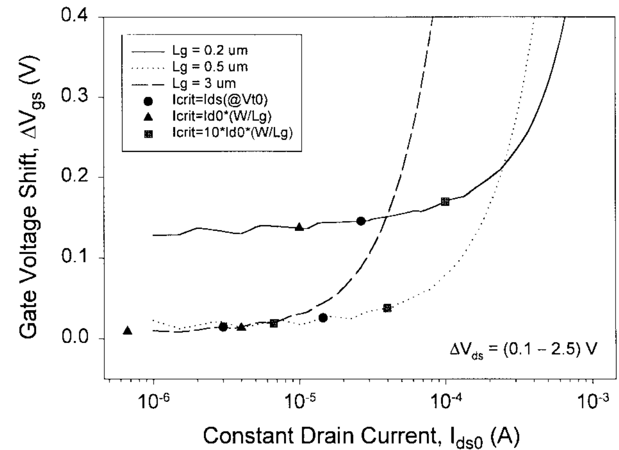

Fig. 4. Gate voltage shift due to a change in $\Delta V_{ds} = (0.1\text{–}2.5\ \text{V})$ measured at different drain currents for three transistors with $L_g = 3, 0.5,$ and 0.2 μm.

CC method is made apparent from the inset of Fig. 1: Even if the same long-channel current I_{crit} is chosen as the value of I_{d0}, the extracted V_t at short channel length will not be a correct and consistent representation of the short-channel V_t since the current is unphysically forced to have an exact $1/L_g$ behavior. This difference (compared to the "I_{crit} @V_{t0}" definition) can be significant for deep-submicron MOSFET's since a unique solution is extremely important due to unavoidable process variations.

The measured $I_{ds} - V_{gs}$ curves in saturation ($V_{ds} = 2.5$ V) for channel lengths ranging from 10 μm down to 0.2 μm are shown in Fig. 2. The gate voltages at the critical currents (from Fig. 1), which are the saturation threshold voltages, are shown together with two constant-current values, $I_{d0} = 0.1$ and 1 μA. The extracted V_{tsat} versus L_g is shown in Fig. 3, together with the V_{tsat} extracted based on the quadratic-extrapolation method as a comparison. In fact, the shape of the $V_{tsat} - L_g$ curves in Fig. 3 is a result of the transformation of Fig. 2 data (*symbols*) with respect to Fig. 1 (inset): $V_{tsat} = f^{-1}[g(L_g)]$ where $I_{crit} = f(V_{tsat})$ from Fig. 2 and $I_{crit} = g(L_g)$ from Fig. 1, and the transfer function, $\log(I_{crit}) - \log(L_g)$, is linear. It is noteworthy from Fig. 3 that the arbitrary choice of I_{d0} in the conventional CC definition not only results in a large shift in V_t, but also different shape of the V_t roll-up (larger roll-up at lower I_{d0} is observed), which is due to the reverse short-channel effect. This implies that V_t modeling is dependent upon the choice of I_{d0}, which is undesirable, e.g., for the application of inverse modeling. The "I_{crit} @V_{t0}" definition, however, has a unique solution as well as a simple extraction.

Another important criterion for V_t definition is its consistency at different values of V_{ds}. It is known that at a given gate length, V_t reduction at increasing V_{ds} is mainly due to the DIBL effect. If the linear V_{t0} is not chosen appropriately, the threshold shift ΔV_t due to ΔV_{ds} may not be a correct representation of the DIBL voltage. To justify the validity of the "I_{crit} @V_{t0}" definition with respect to this concern, the gate voltage shift (ΔV_{gs}) at a constant drain current due to a change in the drain voltage $\Delta V_{ds} = (0.1\text{–}2.5)$ V is plotted against the drain-current level at which ΔV_{gs} is extracted from the measured linear and saturation $I_{ds} - V_{gs}$ curves for three different gate lengths, as shown in Fig. 4. Our definition is shown by the circles, together with two values of the CC definition with $I_{d0} = 0.1$ and 1 μA, shown by the triangles and squares, respectively. It is seen that for long-channel devices, DIBL effect is insignificant. At short channel, if the current level is too high (e.g., $I_{d0} > 1$ μA), the error

in ΔV_t would be large. The "$I_{\rm crit}@V_{t0}$" definition is shown to be in a region for correct interpretation of the DIBL voltage. This definition has been employed in the modeling of the DIBL voltage at various $V_{\rm ds}$ values [10].

III. Conclusion

In conclusion, a simple, unambiguous, and consistent threshold voltage is defined, which applies to MOSFET's with any gate length, drain- and substrate-bias conditions. With this definition, the threshold voltage has a unique solution, and its measurement is as simple as the conventional CC method for all values of $V_{\rm ds}$ and $V_{\rm bs}$. Extraction and modeling of the small-geometry effects is easier using this definition since the 2-D short-channel effects (such as series resistance) are supposed to have been contained in the measured V_t data. It will prove to be a meaningful and useful definition in both industry applications and theoretical modeling of deep-submicron MOSFET's.

References

[1] A. Akers and J. J. Sanchez, "Threshold voltage models of short, narrow, and small geometry MOSFET's: A review," *Solid-State Electron.,* vol. 25, no. 7, pp. 621–641, 1982.

[2] J. J. Liou, A. Ortiz-Conde, and F. G. Sanchez, "Extraction of the threshold voltage of MOSFET's: An overview," in *Proc. IEEE HKEDM'97,* Hong Kong, 1997, pp. 31–38.

[3] Z. X. Yan and M. J. Deen, "Physically-based method for measuring the threshold voltage of MOSFET's," in *Proc. Inst. Elect. Eng. G,* 1991, vol. 138, no. 3, pp. 351–357.

[4] H. S. Wong, M. H. White, T. J. Krutsick, and R. V. Booth, "Modeling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFET's," *Solid-State Electron.,* vol. 30, no. 9, pp. 953–968, 1987.

[5] S. Jain, "Measurement of threshold voltage and channel length of submicron MOSFET's," in *Proc. Inst. Elect. Eng. G,* 1988, vol. 135, p. 162.

[6] X. Zhou, K. Y. Lim, and D. Lim, "A general approach to compact threshold voltage formulation based on 2-D numerical simulation and experimental correlation for deep-submicron ULSI technology development," submitted for publication.

[7] Z.-H. Liu, C. Hu, J.-H. Huang, T.-Y. Chan, M.-C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer MOSFET's," *IEEE Trans. Electron Devices,* vol. 40, pp. 86–94, Jan. 1993.

[8] K. Y. Lim and X. Zhou, "Modeling of threshold voltage with nonuniform substrate doping," in *Proc. IEEE ICSE,* Malaysia, Nov. 1998, pp. 27–31.

[9] X. Zhou, K. Y. Lim, and D. Lim, "A new 'critical-current at linear-threshold' method for direct extraction of deep-submicron MOSFET effective channel length," submitted for publication.

[10] X. Zhou and W. Long, "A novel hetero-material gate (HMG) MOSFET for deep-submicron ULSI technology," *IEEE Trans. Electron Devices,* vol. 45, pp. 2546–2548, Dec. 1998.

Assessment of Beryllium Out-Diffusion in AlGaAs/GaAs Heterojunction Bipolar Transistors Using Low-Temperature Photoluminescence Technique

Hong Wang, Geok Ing Ng, Haiqun Zheng, and Penghua Zhang

***Abstract*— Low-temperature photoluminescence (PL) is used for the investigation of beryllium (Be) dopant out-diffusion in AlGaAs/GaAs abrupt single-heterojunction bipolar transistors (HBT's). The degree of Be out-diffusion into the emitter from a Be-doped base can be estimated based on the band gap narrowing effect (BGNE). The measured current gain and emitter-base turn-on voltage of HBT's fabricated on wafers with different growth conditions were found to correlate well with the PL results.**

***Index Terms*— Beryllium, heterojunction bipolar transistors, photoluminescence.**

I. Introduction

The heterojunction bipolar transistor (HBT) uses a wider band gap (e.g., AlGaAs) material for the emitter region and a narrower band gap material (e.g., GaAs) for the base region. A good quality hetero-interface between the emitter and base layers is critical in obtaining good device performance [1]. Molecular beam epitaxial (MBE) grown HBT material with a beryllium-doped (Be-doped) base is often used for commercial devices [2]. A redistribution of the p-type Be dopant during MBE growth has been found at conventional growth condition. SIMS measurements have been intensively used to investigate such Be redistribution. However, the measurement resolution is limited by typical SIMS effects such as ion mixing and surface roughness [3].

Photoluminescence (PL) is a very powerful and nondestructive method and has been widely used for the characterization of III–V compound materials [4], [5]. In an early work, 77 K PL recombination peaks were identified and assigned to the various layers within AlGaAs/GaAs HBT multiple-layer structures [6]. Recently, the hole density in the base was investigated from the low temperature PL spectrum in the p^+-GaAs layer using the band gap narrowing effect [7], [8]. However, to the authors' knowledge, no study has been made thus far using the low-temperature PL technique to study Be out-diffusion behavior in Be-doped HBT's. In this work, we present a simple method to monitor Be out-diffusion in Be-doped AlGaAs/GaAs HBT's using low temperature PL measurements based on the band gap narrowing effect. The reliability of this measurement technique has been confirmed by electrical characterization on fabricated HBT's.

II. Experiments

The AlGaAs/GaAs abrupt HBT structures used in this study were grown by MBE on (100) semi-insulating GaAs substrates. Si and Be were used for n- and p-type doping, respectively. The structures are composed of an n^+-GaAs cap (100 nm, 5×10^{18} cm^{-3}), an n-$Al_{0.3}Ga_{0.7}As$ emitter (180 nm, 5×10^{17} cm^{-3}), a p^+-GaAs base (150 nm, 1×10^{19} cm^{-3}), an n-GaAs collector (500 nm, 5×10^{16} cm^{-3}),

Manuscript received August 26, 1998; revised November 10, 1998. The review of this brief was arrangaed by Editor J. N. Hollenhorst. This work was supported by the National Science and Technology Board of Singapore (NSTB/17/2/1).

The authors are with the Microelectronics Centre, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798.

Publisher Item Identifier S 0018-9383(99)02401-6.

0018–9383/99$10.00 © 1999 IEEE