Experimental Determination of Electrical, Metallurgical, and Physical Gate Lengths of Submicron MOSFET's

Xing Zhou* and Khee Yong Lim†

*School of Electrical & Electronic Engineering, Nanyang Technological University,
Nanyang Avenue, Singapore 639798. exzhou@ntu.edu.sg
†Chartered Semiconductor Manufacturing Ltd, 60 Woodlands Industrial Park D,
Street 2, Singapore 738406

ABSTRACT

A simple, empirically-based method is developed for extraction of submicron surface-channel MOSFET's effective channel length (L_{eff}) with critical-dimension correction to poly-gate length (L_g) and correlation to metallurgical channel length (L_{met}). A self-consistent compact model for the lightly-doped drain (LDD) lateral diffusion is proposed, which can be correlated to the extracted L_{eff}. The combined experimental determination of L_{eff}, L_{met}, and L_g further validates the proposed "critical-current at linear-threshold" ("$I_{crit}@V_{t0}$") L_{eff}-extraction method, and provides important applications in statistical process control and monitoring as well as deep-submicron (DSM) technology characterization and device modeling.

Keywords: Effective channel length, metallurgical channel length, LDD lateral diffusion, critical-dimension correction, critical-current at linear-threshold, MOSFET.

1 INTRODUCTION

MOSFET channel length is the most important parameter for transistor design, technology scaling, and device modeling, which becomes increasingly difficult to measure and model as the technology is driven into the deep-submicron (DSM) regime. There are many "lengths" discussed in the literature, such as drawn gate length (L_{drawn}), mask length (L_{mask}), physical (or poly) gate length (L_g or L_{poly}), metallurgical channel length (or junction-to-junction spacing) (L_{met} or L_{jj}), and effective channel length (L_{eff}). Among them, only L_{drawn} (drawn on the layout) is well and accurately defined. L_g is a physical quantity, but is subject to (lithographical and etching) process uncertainties during mask making (ΔL_{mask}) and gate patterning (ΔL_{g1}) and etching (ΔL_{g2}). L_{met} is also a physical quantity, but is difficult to measure due to unavailability of two-dimensional (2-D) channel profiling. L_{eff} is an electrical parameter whose extraction strongly depends on its definition. However, if properly defined and accurately extracted, L_{eff} can be extremely useful for providing information on process controllability and device electrical characteristics.

Although accurate (and independent) determination of L_{eff} is important, it should not be overemphasized since the channel resistance, $R_{ch} = r_{ch}(V_{gt})L_{eff}$, is in principle inseparable from the S/D series resistance, $R_{sd}(V_{gt})$, where $V_{gt} \equiv V_{gs} - V_{t0}$ is the gate overdrive [1]. The decoupled C–V method [2] is supposed to be able to extract an accurate L_{eff} (although femto-farad accuracy is required for submicron devices). However, when relating to I–V data, fitting is unavoidable to model $R_{sd}(V_{gt})$, which is just another way of partitioning R_{ch} and R_{sd}. As recently studied [3], process uncertainties cannot be eliminated for all conventional L-array methods and it was suggested that single-transistor methods are able to extract L_{eff} values at low V_{gs} (near threshold) that are closer to L_{met}. Metallurgical length, however, is not practically measurable (except for 2-D numerical studies [4]). Although strongly definition dependent, deviation (not necessarily error [5]) of L_{eff} from L_{met}, $\delta_L \equiv L_{eff} - L_{met}$, can provide important information on the lightly doped drain (LDD) device structures for process control and monitoring [4].

Fig. 1 summarizes common L_{eff}-extraction methods reported in the literature. Conventional methods are all based on the "ideal" model [6], with the measured total resistance (R_{tot}) in linear mode (small V_{ds}) partitioned into two parts, $R_{tot} = V_{ds}/I_{ds} = R_{sd}(V_{gt}) + r_{ch}(V_{gt})L_{eff}$, in the hope that R_{tot} versus L_g at different gate overdrive V_{gt} would "merge" to one point. However, this linear relationship starts to deviate in the DSM regime [7], which, in principle, invalidates all the methods based on the conventional method unless some kind of averaging method is adopted [1], [4], [8] to minimize (or neglect) the bias dependence of S/D series resistance, channel resistance, or both (as depicted in Fig. 1). The observed nonscaling $R_{tot} - L_{drawn}$ behavior [7] is most pronounced at low V_{gt}, at which L_{eff} is known to be close to the bias-independent L_{met} [1], [3], [5], however, the linear-mode assumption will be violated at low V_{gt}. Another common concern is to avoid, or to correct, the influence of R_{sd} in the measured linear threshold voltage (V_{t0}) with the maximum-g_m definition [6]. On the other hand, the newly proposed "$I_{crit}@V_{t0}$" L_{eff}-extraction method [9] takes advantage of the fact that the information on R_{sd} is contained in the measured $I_{crit}@V_{t0}$ data, and L_{eff} is extracted at zero V_{gt} based on a different algorithm from all conventional methods.

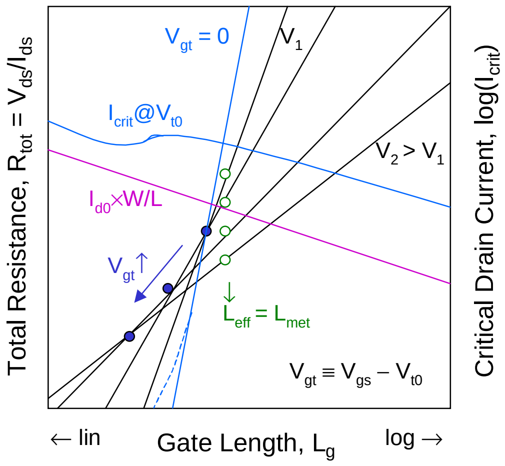

Figure 1: Comparison of the I_{crit}@V_{t0} method with existing L_{eff}-extraction methods (solid circle: [1]; open circle: [8]; dotted line: equivalence of this work, similar to [7]). L_{eff} reduction due to 2-D short-channel effects is assumed to be contained in the total linear drain current I_{ds0} at $V_{gs} = V_{t0}$ with the maximum-g_m definition.

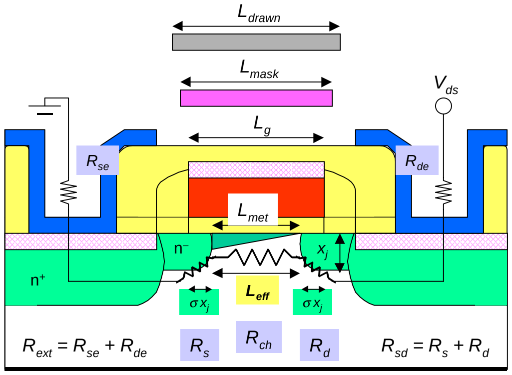

Figure 2: The MOSFET model depicting the various lengths and the parameters used in the "I_{crit}@V_{t0}" L_{eff}-extraction method.

This paper aims to provide a simple approach to experimentally determining the electrical channel length and correlating to the physical and metallurgical gate lengths based on the newly proposed "I_{crit}@V_{t0}" extraction method [9]. For the first time, the electrical, metallurgical, and physical lengths of deep-submicron MOSFET's are determined experimentally with a very simple algorithm based on one *I–V* measurement (of course, within the validity of the definitions proposed). The extraction algorithm is performed at zero gate overdrive ($V_{gt} = 0$) and, hence, much closer extracted L_{eff} to L_{met} can be expected [1], [3], [5]. The method can be easily integrated into automatic wafer test systems for process monitoring and device characterization.

2 MEASUREMENT, EXTRACTION, AND MODELING

To apply the "I_{crit}@V_{t0}" method [9], it is important to measure accurately the linear threshold voltage V_{t0} from the maximum-g_m definition as well as the critical drain current I_{crit} at $V_{gs} = V_{t0}$ for each device of length L_{drawn} on the same wafer [9]. In fact, only the $I_{ds} - V_{gs}$ data (at low V_{ds}) is needed for each MOSFET, and extraction of V_{t0} and I_{crit} can be automated. The devices under test should be selected on the same die to minimize gate oxide nonuniformities and doping variations. For better accuracy, inclusion of very long devices (> 10 µm) is recommended. The set of $I_{crit} - L_{drawn}$ data for this work was obtained from the experimental test wafer of a 0.25-µm CMOS process with L_{drawn} ranging from 10 µm down to 0.2 µm [9], [10].

To have a simple method, one must have a simple model (simple mental image of reality) of a MOSFET for easy measurement, extraction, and modeling. Our simple conceptual MOSFET model is shown in Fig. 2, in which the terminal drain current ($I_{ds} = I_{crit}$) at $V_{gs} = V_{t0}$ flows through the channel (L_{eff}) and S/D lateral diffusion regions ($2\sigma x_j$) under a given external drain voltage $V_{ds} = V_{d0}$. (The external contact resistance can be ignored compared to the LDD resistance when $V_{gs} = V_{t0}$ [11].) Since L_{eff} is extracted from electrical *I–V* on devices of gate length L_g (not L_{drawn}), and LDD is formed after poly definition (L_g), which determines L_{met}, a model for the actual L_g and L_{met} is needed. The following simple model is proposed for the physical gate length:

$$L_g = L_{drawn} - \Delta_{CD} \tag{1}$$

where Δ_{CD} is the *critical-dimension correction* that accounts for process variations in mask/gate lithography and poly etching (ΔL_{mask}, ΔL_{g1}, and ΔL_{g2} mentioned in Introduction). In principle, Δ_{CD} can be measured by SEM for each device, although a model for its nonuniformity requires substantial research efforts. As a first-order approximation, it is assumed that Δ_{CD} is L_{drawn} independent (i.e., all devices on the same wafer have the same amount of Δ_{CD}).

Since L_{met} is not easily measurable (although physically exists), the following simple model is used:

$$L_{met} = L_g - 2\sigma x_j \tag{2}$$

where x_j is the LDD junction depth and σ is a parameter for the *LDD lateral diffusion*. So, L_{met} is subject to three process variations: Δ_{CD}, x_j, and σ. x_j can be determined from SIMS profiles on test structures of the same wafer, but σ is the most difficult one to determine.

For L_{eff}, the simple "calibration–extraction" algorithm [9] is used on the measured $I_{crit} - L_{drawn}$ data, with L_g corrected by (1). Since no knowledge of Δ_{CD} is available, six values of Δ_{CD} (ranging from 0 to 25 nm) are assumed. The channel sheet resistance

$$r_{ch0} = \rho(V_{ds}/L_g)^{\gamma} \tag{3a}$$

is first calibrated for long-channel data:

$$I_{crit} = V_{ds}/(r_{ch0}L_g), \tag{3b}$$

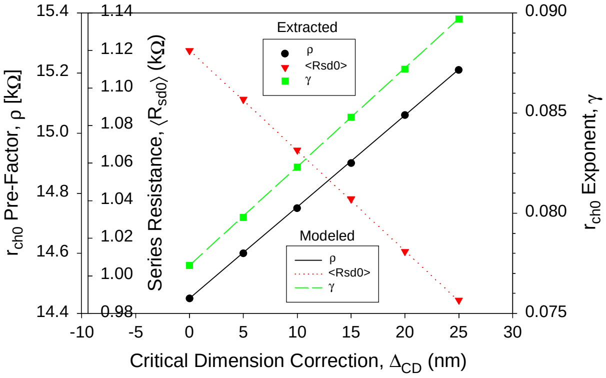

Figure 3: Extracted (symbols) and modeled (lines) parameters ρ, γ, and $\langle R_{sd0} \rangle$ as a function of Δ_{CD} by fitting the $I_{crit} - L_g$ data.

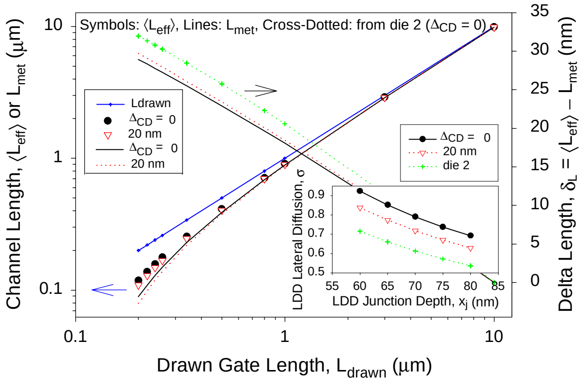

Figure 4: *Left axis*: $\langle L_{eff} \rangle$ vs. L_{drawn} (*symbols*) and L_{met} vs. L_{drawn} (*lines*) based on calibrating $\langle L_{eff} \rangle = L_{met}$ at long channel. The difference $\delta_L = \langle L_{eff} \rangle - L_{met}$ is shown on the right axis. The modeled σ is shown in the inset with two values of Δ_{CD}. δ_L and σ from the second die (Δ_{CD} = 0) are shown in *cross-dotted lines*.

and then, an averaged series resistance $\langle R_{sd0} \rangle$ is obtained from nonlinear regression on all data:

$$I_{crit} = V_{ds}/(\langle R_{sd0} \rangle + r_{ch0} L_g). \tag{3c}$$

The extracted $\langle L_{eff} \rangle$ is given by

$$\langle L_{eff} \rangle = L_g - \langle \Delta L \rangle = L_g - \langle R_{sd0} \rangle / r_{ch0}. \tag{3d}$$

As shown in Fig. 3, the extracted ρ, γ, and $\langle R_{sd0} \rangle$ (*symbols*) as a function of Δ_{CD} exhibit excellent linearity. This means that empirical linear relationships can be easily obtained with only two extractions at Δ_{CD} = 0 and 25 nm, as follows:

$$\rho = \rho_0 + \rho_1 \Delta_{CD} = 14.45 + 3.05\times10^{-2} \Delta_{CD} \text{ [k}\Omega\text{]} \qquad \text{(4a)}$$
$$\gamma = \gamma_0 + \gamma_1 \Delta_{CD} = 0.0774 + 4.93\times10^{-4} \Delta_{CD} \qquad \text{(4b)}$$
$$\langle R_{sd0} \rangle = R_0 + R_1 \Delta_{CD} = 1.12 - 5.34\times10^{-3} \Delta_{CD} \text{ (k}\Omega\text{)} \qquad \text{(4a)}$$

where Δ_{CD} is in nm, as shown in Fig. 3 (*lines*). Then, $\langle L_{eff} \rangle$ is fully modeled by (3), (4), and (1) as a function of L_{drawn} with Δ_{CD} as a parameter.

To observe the deviation of $\langle L_{eff} \rangle$ from L_{met}:

$$\delta_L = \langle L_{eff} \rangle - L_{met} = 2\sigma x_j - \langle \Delta L \rangle, \tag{5}$$

the LDD lateral diffusion (or gate-to-S/D overlap, σx_j) is modeled as a whole, with the assumption that $\langle L_{eff} \rangle = L_{met}$ at long channel. This is consistent with the basic assumption of the "I_{crit}@V_{t0}" method (i.e., calibration at long channel). Then, from (5) with $\delta_L = 0$ at $L_{drawn} = L_\infty$ = 10 μm (the longest drawn length in the data):

$$\sigma = \frac{(\Delta L)_\infty}{2x_j} = \frac{1}{2x_j} \left[\frac{\langle R_{sd0} \rangle}{\rho \left[V_{d0} / (L_\infty - \Delta_{CD}) \right]^\gamma} \right] \tag{6}$$

where ρ, γ, and $\langle R_{sd0} \rangle$ are from (4). Then,

$$\delta_L = \langle L_{eff} \rangle - L_{met} = (\Delta L)_\infty - \langle \Delta L \rangle, \tag{7}$$

which is independent of x_j, represents the difference of $\langle L_{eff} \rangle$ with respect to that of the long channel. The δ_L such defined provides a measure of the short-channel effects of LDD structures at decreasing L_{drawn} since σ is proportional to $\langle R_{sd0} \rangle$, an averaged R_{sd} contribution contained in the I_{crit} data [9]. The total LDD lateral diffusion, $2\sigma x_j = (\Delta L)_\infty$, is fixed for all L_{drawn} at a given Δ_{CD}, which is determined from the long-channel electrical ΔL. The extracted $\langle L_{eff} \rangle$, L_{met}, and δ_L are shown in Fig. 4, with the $\sigma - x_j$ plot shown in the inset, at two values of Δ_{CD}. The extracted values of σ (0.7~0.75) are found to be reasonable with the estimated x_j = 70~75 nm and Δ_{CD} = 5~15 nm.

3 DISCUSSIONS

The proposed model is similar to what was studied in [4], in which r_{ch} is extracted from the spatial derivative of quasi-Fermi potential along the channel normalized by the terminal current per device width with bias-independent R_{sd}. In the "I_{crit}@V_{t0}" method, V_{gs} is fixed at V_{t0} and L_{drawn} is used as the averaging variable. Since V_{ds} is fixed at 0.1 V in all devices, the effective parallel field is increased at shorter gate lengths, equivalent to an increased $(dV/dx)/(I_{ds}/W)$ as defined in [4], which has been modeled empirically by the parameter γ in r_{ch0} in our model. This model is also consistent with [5], in which one source of error in δ_L was attributed to the bias-dependent R_{sd} due to a change in r_{ch}.

Since I–V data is used in any L_{eff} extraction methods, to minimize the error, it is important that all the devices under test have consistent operating conditions. In conventional methods, averaging is performed over a range of V_{gt} values. If, e.g., the constant-current definition is used in V_{t0} extraction (as is often done to avoid the V_{t0} dependence on R_{sd} in the maximum-g_m definition), short- and long-channel transistors would be operating under different conditions since the critical current is unphysically scaled [10]. Depending on the definition of V_{t0}, mobility degradation may also be different for different length devices since the electrons would experience different perpendicular fields. When V_{ds} is fixed for all the devices, which is thus far a common practice, different lateral fields result in different mobilities (hence, different currents). Our proposed method essentially eliminates the effect of mobility degradation since zero gate overdrive is consistently applied to all devices and channel charge inversion is consistently achieved based on the maximum-g_m definition. Lateral-field variation is also empirically modeled.

The validity of the proposed definitions and extraction approach should be judged by its self-consistency and application to providing a meaningful guide to technology developers. The physical interpretation of the defined quantities (such as R_{sd0}, r_{ch0}, σ and δ_L) must be evaluated within their definitions, which are quite different from the conventional ones.

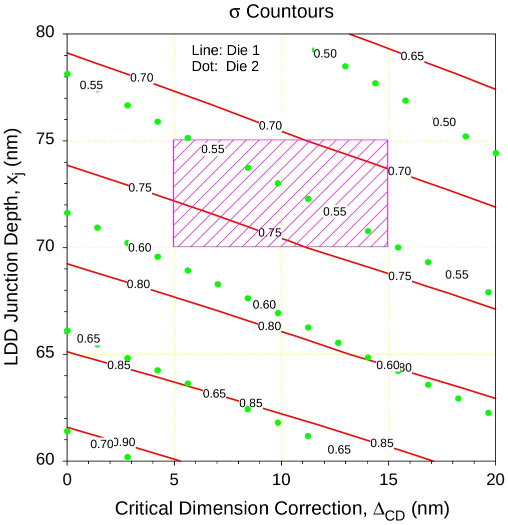

Figure 5: σ contours from die 1 (solid line) and die 2 (dot) as a function of x_j and Δ_{CD}. Estimated x_j and Δ_{CD} uncertainties are shown with the shaded region.

One application of the proposed model is for statistical analysis of process variations based on contour plots with process uncertainties as parameters. For example, based on (6), σ contours are plotted against Δ_{CD} and x_j, as shown in Fig. 5. Any fluctuations of the σ values obtained from different sets of devices on the same wafer are indicative of the associated statistical uncertainties. On the other hand, if Δ_{CD} and x_j (or their distributions) can be determined experimentally (e.g., shaded region in Fig. 5), a smaller σ suggests a smaller $\langle R_{sd0} \rangle$ contribution (6). The corresponding changes in the $\delta_L - L_{drawn}$ curves would be a reflection of the parallel-field effect, since δ_L such defined (7) is due to the change in r_{ch0} relative to the long-channel one (similar to dV/dx in [4]). Results obtained from another die on the same wafer (assuming $\Delta_{CD} = 0$, extracted $\rho = 16.57$ kΩ, $\gamma = 0.119$, $\langle R_{sd0} \rangle = 0.82$ kΩ) are shown in Fig. 4 (*cross-dotted line*) for δ_L and σ, and in Fig. 5 (*dot*) for σ contours. Assuming $\Delta_{CD} = 0$, the deviation from the first die could be due to process fluctuations in $x_j \pm \Delta x_j = 70 \pm 9$ nm if the same $\sigma = 0.7$ is to be used; or due to variations in 2-D LDD lateral diffusion ($\sigma \pm \Delta\sigma = 0.7 \pm 0.1$) if the same $x_j = 70$ nm can be confirmed.

4 CONCLUSION

The most attractive feature of the proposed method is its simplicity, which requires only one *I*–*V* measurement. This is achieved by the simple, yet physical, conceptual model, combined with empirical calibration to the well-defined long-channel device characteristics. The developed compact models and extraction approach can be (and have been) easily automated, which will prove to be extremely useful for statistical process analysis and device characterization.

REFERENCES

[1] G. J. Hu, C. Chang, and Y.-T. Chia, “Gate-Voltage-Dependent Effective Channel Length and Series Resistance of LDD MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 34, pp. 2469–2475, Dec. 1987.

[2] J.-C. Guo, S. S.-S. Chung, and C. C.-H. Hsu, “A New Approach to Determine the Effective Channel Length and the Drain-and-Source Series Resistance of Miniaturized MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 41, pp. 1811–1818, Oct. 1994.

[3] S. Biesemans, M. Hendriks, S. Kubicek, and K. D. Meyer, “Practical Accuracy Analysis of Some Existing Effective Channel Length and Series Resistance Extraction Methods for MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 45, pp. 1310–1316, June 1998.

[4] Y. Taur, Y.-J. Mii, R. Logan, and H.-S. Wong, “On “Effective Channel Length” in 0.1-μm MOSFET’s,” *IEEE Electron Devices Lett.*, vol. 16, pp. 136–138, Apr. 1992.

[5] J. Y.-C. Sun, M. R. Wordeman, and S. E. Laux, “On the Accuracy of Channel Length Characterization of LDD MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 33, pp. 1556–1562, Oct. 1986.

[6] K. Terada and H. Muta, “A new method to determine effective MOSFET channel length,” *Japan. J. Appl. Phys.*, vol. 18, p. 935, 1979.

[7] D. Esseni, H. Iwai, M. Saito, and B. Ricco, “Nonscaling of MOSFET’s linear resistance in the deep submicrometer regime,” *IEEE Electron Device Lett.*, vol. 19, pp. 131–133, Apr. 1998.

[8] K. Takeuchi, N. Kasai, T. Kunio, and K. Terada, “An effective channel length determination method for LDD MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 43, pp. 580–587, Apr. 1996.

[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,” *IEEE Trans. Electron Devices*, vol. 46, pp. 1492–1494, July 1999.

[10] X. Zhou, K. Y. Lim, and D. Lim, “A Simple and Unambiguous Definition of Threshold Voltage and Its Implications in Deep-Submicron MOS Device Modeling,” *IEEE Trans. Electron Devices*, vol. 46, pp. 807–809, Apr. 1999.

[11] K. K. Ng and W. T. Lynch, “Analysis of the gate-voltage-dependent series resistance of MOSFET’s,” *IEEE Trans. Electron Devices*, vol. ED-33, pp. 965–972, July 1986.