A New Method for Layout-Dependent Parasitic Capacitance Analysis and Effective Mobility Extraction in Nanoscale Multifinger MOSFETs

Kuo-Liang Yeh, *Member, IEEE*, and Jyh-Chyurn Guo, *Senior Member, IEEE*

***Abstract*—The impact of layout-dependent parasitic capacitances on extraction of inversion carrier density $Q_{\rm inv}$ and effective mobility $\mu_{\rm eff}$ has been investigated on multifinger MOSFETs. An improved open deembedding method can eliminate the extrinsic parasitic capacitance, and 3-D interconnect simulation is necessary for extraction of intrinsic parasitic capacitances such as gate finger sidewall and finger-end fringing capacitances, i.e., $C_{\rm of}$ and $C_{f\rm (poly\text{-}end)}$, respectively. Both categories of parasitic capacitance lead to overestimated $Q_{\rm inv}$ and underestimated $\mu_{\rm eff}$. The increase in effective channel width $W_{\rm eff}$ due to ΔW from shallow trench isolation (STI) top-corner rounding may compensate $\mu_{\rm eff}$ degradation due to STI stress. The tradeoff between $\mu_{\rm eff}$ and $W_{\rm eff}$ determines the impact of width scaling on $I_{\rm DS}$ and G_m. A new method based on the measured S-parameters, open-M1 deembedding, and Raphael simulation can precisely determine the mentioned parameters associated with the intrinsic channel and realize accurate extraction of $\mu_{\rm eff}$ in multifinger MOSFETs with various layouts and narrow widths down to 0.125 μm.**

***Index Terms*—Effective mobility, effective width, fringing capacitance, open deembedding, parasitic capacitances.**

I. Introduction

THE LAYOUT-dependent stress effect has been known as a critical factor responsible for the variation of channel carrier mobility. Most of previous works limited the observation through dc characterization and assumed the variations of channel current $I_{\rm DS}$ and transconductance G_m simply from that of effective mobility $\mu_{\rm eff}$. This assumption may become invalid for nanoscale MOSFETs, in which the variation of effective channel width $W_{\rm eff}$ emerges as a more important variable [1]. To investigate the stress effect on $\mu_{\rm eff}$, an accurate and reliable method for $\mu_{\rm eff}$ extraction becomes a prerequisite, but this has been increasingly challenging in nanoscale MOSFETs. The split C–V method has been frequently used for the determination of inversion carrier density $Q_{\rm inv}$, which is one of basic parameters for $\mu_{\rm eff}$ extraction [2], [3]. However, the conventional split C–V method is not valid for miniaturized devices because the measured gate capacitance cannot represent the intrinsic channel capacitance. Recently, several improved techniques have been proposed, trying to extend the split C–V method to scaled CMOS devices [4]–[7]. Unfortunately, the proposed methods leave a number of critical problems without a concrete solution. First, the conventional C–V measurement cannot be applied to nanoscale MOSFETs with ultrathin gate oxide due to gate leakage current and abnormal C–V falloff under the strong inversion condition. Second, an appropriate open deembedding method for removing parasitic capacitances from the pads, interconnection lines, and lossy substrate [8], [9] is not available for the published methods relying on the conventional C–V measurement. The mentioned parasitic capacitances do not make any contribution to $I_{\rm DS}$ but lead to an overestimation of $Q_{\rm inv}$ and then an underestimation of $\mu_{\rm eff}$. Moreover, the existing methods are limited to single gate-finger devices and not applicable to multifinger MOSFETs, which have been widely used in RF and analog circuits. A more critical problem is that the gate sidewall and gate finger-end fringing capacitances were not considered in the existing methods. Most of the early works are limited to long-channel devices in which the gate sidewall fringing capacitance is negligibly small compared with the intrinsic channel capacitance. Recent works, even with an improvement for short-channel devices, have been limited to single-finger and wide-channel-width MOSFETs in which the gate finger-end fringing capacitance is insignificant [6]. In our recent work, the 3-D parasitic gate capacitances have been accurately calculated by a 3-D integral model and Raphael simulation. Our research foresees an important impact that the parasitic gate capacitances are not scalable with gate length L_g scaling and may dominate that from the intrinsic channel region in nanoscale MOSFETs [10]. However, the previous models [10]–[14] limit the focus on the gate sidewall capacitance and leave the finger-end fringing capacitance an open question. This topic deserves more research effort to explore the layout-dependent mechanisms and the methodology for precise extraction of intrinsic gate capacitance density $C_{\rm ox(inv)}$. The last point of special concern is that the shallow trench isolation (STI) profile and top-corner rounding (TCR) effects on $W_{\rm eff}$ were not considered in previous works [4]–[7]. The approximation of $W_{\rm eff}$ by the total width on layout, i.e., $W_{\rm tot}$, becomes invalid for multifinger and narrow-width MOSFETs [1].

In this paper, a new method based on high-frequency S-parameter measurement, an improved open deembedding

Manuscript received March 8, 2011; revised April 27, 2011 and May 17, 2011; accepted May 20, 2011. Date of publication July 25, 2011; date of current version August 24, 2011. This work was supported in part by the National Science Council under Grant NSC 98-2221-E009-166-MY3. The review of this paper was arranged by Editor Y.-H. Shih.

K.-L. Yeh is with the Institute of Electronics Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, and also with Silicon Motion Technology Corporation, Hsinchu 300, Taiwan.

J.-C. Guo is with the Institute of Electronics Engineering, National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: jcguo@mail.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2011.2158105

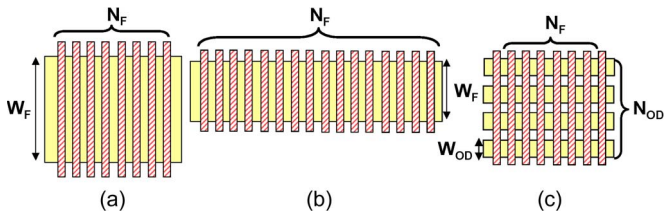

Fig. 1. Schematic of multifinger MOSFETs with three different layouts. (a) Standard multifinger device: $W_F \times N_F = 2\ \mu\text{m} \times 16$ (W2N16). (b) Narrow-OD devices: $W_F \times N_F = 1\ \mu\text{m} \times 32$ (W1N32) and $W_F \times N_F = 0.5\ \mu\text{m} \times 16$ (W05N64). (c) Multi-OD devices: $W_{\text{OD}} \times N_{\text{OD}} = 2\ \mu\text{m} \times 1$ (OD1), $W_{\text{OD}} \times N_{\text{OD}} = 0.25\ \mu\text{m} \times 8$ (OD8), and $W_{\text{OD}} \times N_{\text{OD}} = 0.125\ \mu\text{m} \times 16$ (OD16).

process, and Raphael simulation has been developed to precisely extract and eliminate all of the parasitic capacitances outside the intrinsic channel. In addition, STI TCR-induced ΔW can be determined, giving W_{eff} with high precision, which is critical for multifinger devices with extremely narrow width. Accordingly, effective mobility μ_{eff} can be extracted with improved accuracy, which is attributed to the ensured accuracy of basic device parameters such as $C_{\text{ox(inv)}}$, Q_{inv}, L_g, and W_{eff}.

II. Device Fabrication and Characterization

In our recent work, multifinger MOSFETs have been fabricated in a 90-nm logic CMOS process [1]. Target gate length L_g is 80 nm, and total gate width $W_{\text{tot}} = W_F \times N_F$ is specified at 32 μm. Two kinds of layouts derived from the standard multifinger MOSFET, namely, narrow-OD and multi-OD, were implemented for this study. Note that OD means oxide diffusion, which is equivalent to the active area, generally denoted by AA. Fig. 1(a)–(c) displays the device layouts for standard, narrow-OD, and multi-OD, respectively. The narrow-OD devices illustrated in Fig. 1(b) were designed with simultaneously varied N_F and W_F under fixed total width, i.e., $W_{\text{tot}} = 32\ \mu$m, and $W_F \times N_F = 2\ \mu\text{m} \times 16$, $1\ \mu\text{m} \times 32$, and $0.5\ \mu\text{m} \times 64$, which were denoted by W2N16, W1N32, and W05N64, respectively. The multi-OD devices shown in Fig. 1(c) represent multiple OD fingers with simultaneously varied OD finger width W_{OD} and OD finger number N_{OD} under a specified finger width, i.e., $W_F = W_{\text{OD}} \times N_{\text{OD}}$. In this paper, the multi-OD devices were implemented with $N_F = 16$ and $W_{\text{OD}} \times N_{\text{OD}} = 2\ \mu\text{m} \times 1$, $0.25\ \mu\text{m} \times 8$, and $0.125\ \mu\text{m} \times 16$, corresponding to $W_F = 2\ \mu$m, namely, OD1, OD8, and OD16, respectively.

S-parameters were measured by an Agilent E8364B network analyzer for high-frequency characterization up to 40 GHz. Two-step deembedding, i.e., open and short deembedding, was performed via dummy open and dummy short test structures. Open deembedding was carried out on the measured two-port S-parameters to remove the parasitic capacitances from the pads, interconnection lines, and lossy substrate. Moreover, short deembedding was done to eliminate the parasitic resistances and inductances originated from the metal interconnection [8]. Two kinds of dummy open test structures, namely, open-M3 and open-M1, were designed, as shown in Fig. 2(a) and (b), respectively. Note that the dummy open-M1 created in this paper incorporates metal-3/metal-2/metal-1 (M3–M2–M1) stacked layers, as shown in Fig. 2(b), to remove all of the

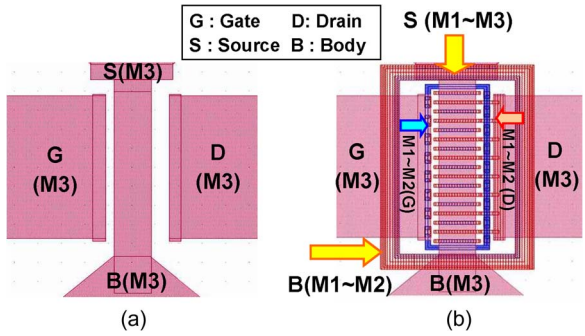

Fig. 2. Dummy open test structures. (a) Open-M3 for open deembedding to M3. (b) Open-M1 for open deembedding to M1.

parasitic capacitances from the stacked metals from top to bottom, i.e., M1. On the other hand, open-M3, which is known as the conventional and most frequently used open deembedding method, cannot remove the parasitic capacitances from the metals below M3, which are not scalable with device scaling and indeed impose a significant influence on miniaturized devices in high-frequency performance. In this paper, 3-D interconnect simulator Raphael was utilized to simulate the 3-D parasitic capacitances from the metals and gate-related fringing capacitances.

III. Intrinsic Gate Capacitance Extraction

A. Open Deembedding for Intrinsic Gate Capacitance Extraction

The S-parameters measured from the devices, the dummy open-M3, and the dummy open-M1, which were denoted by DUT, OM3, and OM1, were converted to Y-parameters, as shown in (1.1)–(1.3) for open deembedding. Following the process given by (1)–(3), open-M3 and open-M1 deembedding were simultaneously performed on the same device to investigate the differences in the extracted gate capacitances and the final impact on μ_{eff}. Thus

$$[S]_{\text{DUT}} \Rightarrow [Y]_{\text{DUT}} \Rightarrow C_{gg(\text{DUT})} = \frac{\text{Im}\left(Y_{11}^{\text{DUT}}\right)}{\omega} \tag{1.1}$$

$$[S]_{\text{OM3}} \Rightarrow [Y]_{\text{OM3}} \Rightarrow C_{gg(\text{OM3})} = \frac{\text{Im}\left(Y_{11}^{\text{OM3}}\right)}{\omega} \tag{1.2}$$

$$[S]_{\text{OM1}} \Rightarrow [Y]_{\text{OM1}} \Rightarrow C_{gg(\text{OM1})} = \frac{\text{Im}\left(Y_{11}^{\text{OM1}}\right)}{\omega} \tag{1.3}$$

$$C_{gg(\text{DUT,OM3})} = C_{gg(\text{DUT})} - C_{gg(\text{OM3})} \tag{2}$$

$$C_{gg(\text{DUT,OM1})} = C_{gg(\text{DUT})} - C_{gg(\text{OM1})} \tag{3}$$

$$C_{gg(\text{OM1})} = C_{gg(\text{OM3})} + C_{M123} \tag{4}$$

where

$C_{g,\text{DUT}}$	gate capacitance measured from DUT with pads;
$C_{gg(\text{OM3})}$	parasitic capacitance measured from open-M3;
$C_{gg(\text{OM1})}$	parasitic capacitance measured from open-M1;
C_{M123}	parasitic capacitance from M3–M2–M1;

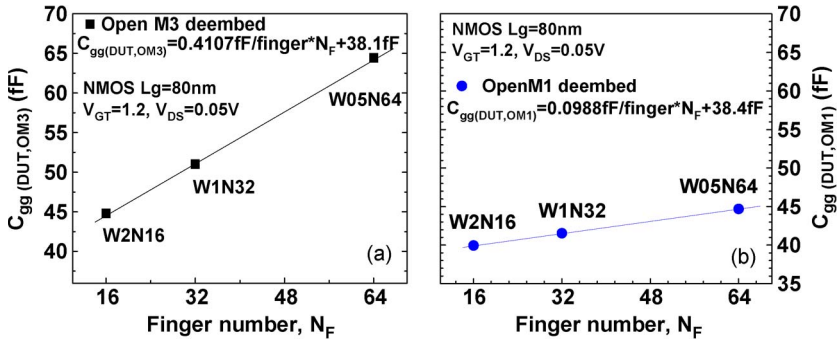

Fig. 3. Gate capacitances $C_{gg(\mathrm{DUT,OM3})}$ and $C_{gg(\mathrm{DUT,OM1})}$ extracted from open-M3 and open-M1 deembedding on multifinger NMOS (W2N16, W1N32, and W05N64) with fixed $W_F \times N_F$. (a) $C_{gg(\mathrm{DUT,OM3})}$ versus N_F. (b) $C_{gg(\mathrm{DUT,OM1})}$ versus N_F.

$C_{gg(\mathrm{DUT,OM3})}$ gate capacitance after open-M3 deembedding;

$C_{gg(\mathrm{DUT,OM1})}$ gate capacitance after open-M1 deembedding.

It was expected that the multifinger MOSFETs with fixed $W_{\mathrm{tot}} = W_F \times N_F$ should have a constant intrinsic gate capacitance, which is independent of the variation of N_F and W_F. To verify this point, $C_{gg(\mathrm{DUT,OM3})}$ and $C_{gg(\mathrm{DUT,OM1})}$ extracted by using open-M3 and open-M1 deembedding are presented versus N_F, as shown in Fig. 3(a) and (b), respectively. Both $C_{gg(\mathrm{DUT,OM3})}$ and $C_{gg(\mathrm{DUT,OM1})}$ indicate a linear function of N_F, and $C_{gg(\mathrm{DUT,OM1})}$ versus N_F presents a much smaller slope than that of $C_{gg(\mathrm{DUT,OM3})}$. The experimental results are out of conventional expectation, and the nonzero slope, i.e., the variation w.r.t. N_F, reveals parasitic capacitances outside the intrinsic channel, which can be significantly reduced using open-M1 deembedding but cannot be eliminated to zero. The linear increase in $C_{gg(\mathrm{DUT,OM1})}$ with N_F, under fixed W_{tot}, suggests the existence of some other components of parasitic capacitance, which are generated from certain sources other than the pads, interconnection lines, and substrate and cannot be removed, even using open-M1 deembedding, i.e., the most effective open deembedding method.

B. Gate Fringing Capacitance Simulation and Analysis

To explore the mechanism responsible for this new observation, a rigorous analysis was performed by using Raphael simulation, and the results suggest that the parasitic capacitances lumped into $C_{gg(\mathrm{DUT})}$ can be classified into two categories: one is contributed from the pads, interconnection lines, and substrate, and the other is fringing capacitances from the gate sidewall and finger ends, which are denoted by C_{of} and $C_{f(\mathrm{poly\text{-}end})}$, respectively. The former is a kind of extrinsic parasitic capacitances and can be eliminated through a dedicated open deembedding method, such as open-M1. However, the latter is actually a kind of intrinsic parasitic capacitance arising from the gate fingers and the surrounding conductors, such as the source/drain (S/D) diffusion regions and the contact plugs, and cannot be removed using any existing deembedding methods. To solve this problem, 3-D capacitance simulation was performed using Raphael to calculate C_{of} and $C_{f(\mathrm{poly\text{-}end})}$.

Fig. 4(a) and (b) illustrates the 3-D MOSFET structure for Raphael simulation. This 3-D structure incorporates four conducting regions, namely, a polygate, a channel region, an S/D diffusion region, and contact plugs to the S/D region. As an individual electrode is specified for each conducting region as defined, the three components of coupling capacitances from the gate to other three regions can be calculated. Among the three components, the gate-to-channel region is the intrinsic gate capacitance responsible for Q_{inv}, and the other two components, i.e., gate to S/D diffusion and gate to contact, which are denoted by $C_{g,\mathrm{Diff}}$ and $C_{g,\mathrm{CT}}$, respectively, constitute the gate sidewall fringing capacitance, given by $C_{\mathrm{of}} = C_{g,\mathrm{Diff}} + C_{g,\mathrm{CT}}$. Fig. 5 presents $C_{g,\mathrm{Diff}}$, $C_{g,\mathrm{CT}}$, and C_{of} calculated for MOSFETs with various L_g. Note that all of the components of the sidewall fringing capacitance indicate very weak dependence on L_g (80–160 nm), and C_{of} is 0.28 fF/μm, corresponding to 90-nm CMOS design rule and technology parameters. The results suggest that the gate sidewall fringing capacitance is not scalable with L_g scaling, and its weighting factor will rapidly increase with device scaling.

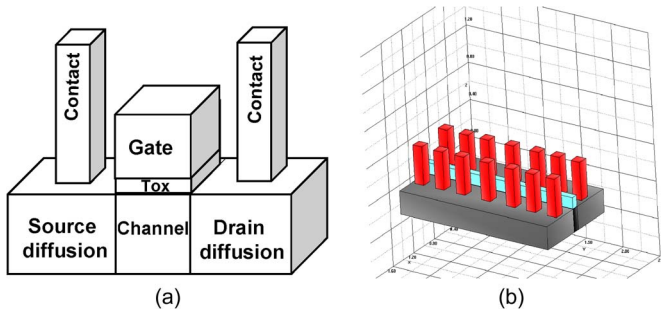

Fig. 4. Schematic of the 3-D MOSFET structure for Raphael simulation. (a) The 3-D structure incorporates four conducting regions, namely, a polygate, a channel, an S/D diffusion region, and contact plugs to the S/D. (b) Cross-sectional view of the multifinger MOSFET.

In addition to gate sidewall fringing capacitances C_{of}, gate finger-end fringing capacitance $C_{f(\mathrm{poly\text{-}end})}$ is another key element of the intrinsic parasitic capacitances, which always exists in MOSFETs but cannot be removed by the existing open deembedding methods. Again, Raphael simulation was employed to calculate $C_{f(\mathrm{poly\text{-}end})}$, which emerges as a critical component for accurate extraction of truly intrinsic gate capacitance and STI TCR-induced ΔW. Both C_{of} and $C_{f(\mathrm{poly\text{-}end})}$ are not scalable with device scaling and may dominate the intrinsic gate capacitance of miniaturized MOSFETs. As a result, C_{of} and $C_{f(\mathrm{poly\text{-}end})}$ appear as critical parameters to

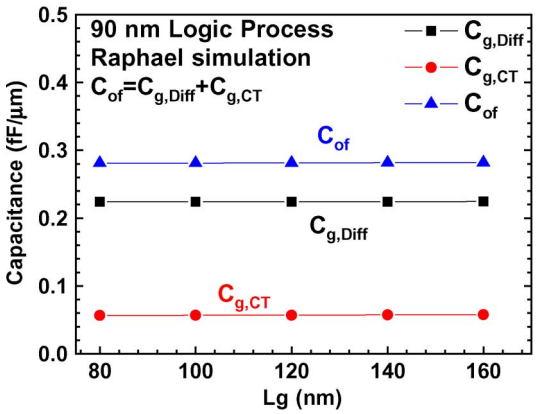

Fig. 5. Gate sidewall fringing capacitances simulated by Raphael for MOSFETs with various L_g, $C_{g,\mathrm{Diff}}$, $C_{g,\mathrm{CT}}$, and $C_{\mathrm{of}} = C_{g,\mathrm{Diff}} + C_{g,\mathrm{CT}}$ versus L_g.

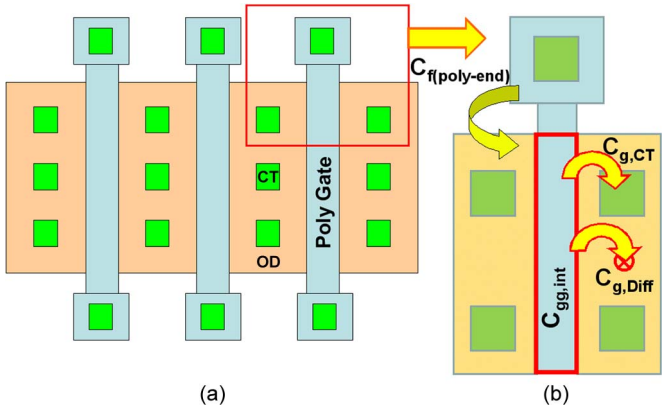

Fig. 6. (a) Schematic of the multifinger MOSFET layout. (b) Three components of fringing capacitances associated with each gate finger: $C_{f(\mathrm{poly\text{-}end})}$ is the finger-end fringing capacitance, and $C_{g,\mathrm{Diff}}$ and $C_{g,\mathrm{CT}}$ are the gate sidewall fringing capacitances.

be known for precise determination of Q_{inv} and then accurate extraction of μ_{eff} in multifinger MOSFETs with narrow width. Fig. 6 illustrates the planar view of a multifinger MOSFET in which three components of fringing capacitances, such as $C_{f(\mathrm{poly\text{-}end})}$, $C_{g,\mathrm{Diff}}$, and $C_{g,\mathrm{CT}}$, are depicted. This graphical analysis explains that $C_{f(\mathrm{poly\text{-}end})}$ is proportional to N_F but is independent of W_F. On the other hand, sidewall fringing capacitances, such as $C_{g,\mathrm{Diff}}$ and $C_{g,\mathrm{CT}}$ are determined by both W_F and N_F and in a linear function proportional to $W_{\mathrm{tot}} = W_F \times N_F$. Note that $C_{f(\mathrm{poly\text{-}end})}$ can be decomposed into the fringing capacitances from the polygate finger ends on STI to the S/D diffusion region and that to the contacts on the S/D region. For narrow-OD devices [see Fig. 1(b)] with simultaneously varied W_F and N_F under a specified $W_{\mathrm{tot}} = W_F \times N_F$, $C_{f(\mathrm{poly\text{-}end})}$ will increase with increasing N_F, and its weighting factor dramatically increases in MOSFETs with very large N_F and extremely narrow W_F. According to 90-nm CMOS design rule and technology parameters, $C_{f(\mathrm{poly\text{-}end})}$ is calculated to be 0.064 fF/finger by Raphael simulation.

Based on an extensive simulation and analysis on the aforementioned parasitic capacitances from pads, metals, lossy substrate, and gate-related fringing effects, the intrinsic gate capacitance extraction flow and analysis can be derived as follows:

$$C_{gg(\mathrm{DUT,OM1})} = C_{gg,\mathrm{int}} \times N_F(W_F + \Delta W) + C_{gg,\mathrm{ext}} \times W_{\mathrm{tot}} + C_{f(\mathrm{poly\text{-}end})} \times N_F \quad (5)$$

where $C_{gg,\mathrm{int}}$ and $C_{gg,\mathrm{ext}}$ are the intrinsic and extrinsic components of the gate capacitance density per unit width, respectively, which are defined as

$$C_{gg,\mathrm{int}} = C_{\mathrm{ox(inv)}} L_g \quad (6.1)$$

$$C_{\mathrm{ox(inv)}} = \varepsilon_{\mathrm{ox}} \varepsilon_0 / T_{\mathrm{ox(inv)}} \quad (6.2)$$

$$C_{gg,\mathrm{ext}} = C_{\mathrm{of}} = C_{g,\mathrm{CT}} + C_{g,\mathrm{Diff}} \quad (7)$$

$$W_{\mathrm{tot}} = W_F \times N_F \quad (8)$$

$T_{\mathrm{ox(inv)}}$ electrical thickness of the oxide under strong inversion;

$C_{\mathrm{ox(inv)}}$ intrinsic gate capacitance density of the inversion channel;

ΔW increase in effective width from STI TCR.

According to (6)–(8), (5) can be rewritten as two parts, one of which is proportional to N_F and the other is independent of N_F, i.e.,

$$C_{gg(\mathrm{DUT,OM1})} = N_F \left[(\Delta W \cdot L_g) C_{\mathrm{ox(inv)}} + C_{f(\mathrm{poly\text{-}end})}\right] + \left(C_{\mathrm{ox(inv)}} L_g + C_{\mathrm{of}}\right) W_{\mathrm{tot}} \quad (9)$$

then $C_{gg(\mathrm{DUT,OM1})}$ of the multifinger MOSFETs with various N_F but fixed $W_{\mathrm{tot}} = W_F \times N_F$ can be expressed as a linear function of N_F with the slope and intercept defined as α and β, respectively, given by

$$C_{gg(\mathrm{DUT,OM1})} = \alpha N_F + \beta \quad (10)$$

where

$$\alpha = (\Delta W \cdot L_g) C_{\mathrm{ox(inv)}} + C_{f(\mathrm{poly\text{-}end})} \quad (11)$$

$$\beta = \left(C_{\mathrm{ox(inv)}} L_g + C_{\mathrm{of}}\right) W_{\mathrm{tot}} = (C_{gg,\mathrm{int}} + C_{\mathrm{of}}) W_{\mathrm{tot}}. \quad (12)$$

Note that the first term of intercept β in (12), i.e., $C_{gg,\mathrm{int}} W_{\mathrm{tot}}$, is the intrinsic gate capacitance, which contributes the inversion carriers $Q_{\mathrm{inv}} W_{\mathrm{tot}}$ that are responsible for channel current I_{DS}. The intrinsic gate capacitance density $C_{gg,\mathrm{int}} = C_{\mathrm{ox(inv)}} L_g$ defined by (6.1) can be extracted from intercept β given by (13) converted from (12), with C_{of} determined by Raphael simulation. In another approach, C_{of} can be extracted from β given by (14.1), under varying L_g and specified $C_{\mathrm{ox(inv)}}$ ($T_{\mathrm{ox(inv)}}$). The C_{of} determined by two independent approaches, one from simulation and the other from measurement, should be self-consistent to validate the accuracy. In addition, L_g can be extracted from (14.2) as a function of C_{of} and $C_{\mathrm{ox(inv)}}$. Finally, ΔW can be extracted from slope α given by (15) derived from (11), in which L_g and $C_{\mathrm{ox(inv)}}$ have been extracted, and $C_{f(\mathrm{poly\text{-}end})}$ is provided by Raphael simulation. The accurate determination of $C_{gg,\mathrm{int}}$ and W_{eff} incorporating ΔW enables

TABLE I
MULTIFINGER MOSFET DEVICE PARAMETERS EXTRACTED FROM NARROW-OD NMOS (W2N16, W1N32, AND W05N32)

NMOS, V_{DS}=0.05V, V_{GS}=1.2V Device Parameters	unit	Extraction after Open-M1 deembedding
L_g	μm	0.08
$T_{ox(inv)}$	nm	3
$C_{ox(inv)}$	fF/μm^2	11.5102
α	fF/finger	0.0988
β	fF	38.37
$C_{of,sim}$	fF/μm	0.2812
$C_{f(polyend)}$	fF/finger	0.064
$C_{ox(inv)}$ =(β/$W_F N_F$-C_{of})/L_g	fF/μm^2	11.4700
$T_{ox(inv)}$=$\varepsilon_0\varepsilon_{ox}$/$C_{ox(inv)}$	nm	3.0100
ΔW=(α-$C_{f(polyend)}$)/($C_{ox(inv)}$*L_g)	μm	0.0379

precise calculation of inversion carrier density $Q_{\rm inv}$ and accurate extraction of effective mobility $\mu_{\rm eff}$ in the intrinsic channel region. The aforementioned flow is described as follows:

$$C_{\rm ox(inv)}L_g = \frac{\beta}{W_{\rm tot}} - C_{\rm of} \tag{13}$$

$$C_{\rm of} = \frac{\beta}{W_{\rm tot}} - C_{\rm ox(inv)}L_g \tag{14.1}$$

$$L_g = \frac{\beta}{W_{\rm tot}C_{\rm ox(inv)}} - \frac{C_{\rm of}}{C_{\rm ox(inv)}} \tag{14.2}$$

$$\Delta W = \frac{\alpha - C_{f(\text{poly-end})}}{C_{\rm ox(inv)}L_g}. \tag{15}$$

Following the extraction flow, open-M1 deembedding, and Raphael simulation can facilitate thorough and precise extraction of the 3-D parasitic capacitances and accurate determination of the intrinsic gate capacitance. Taking the developed characterization and analysis process, the fundamental device parameters such as $C_{\rm ox(inv)}$, L_g, ΔW, $C_{\rm of}$, and $C_{f(\text{poly-end})}$ can be determined with proven accuracy for nanoscale MOSFETs. Table I summarizes the key parameters extracted from narrow-OD NMOS. Note that the slope α of 0.0988 fF/finger is contributed from two portions given by (11), one from STI TCR-induced ΔW, i.e., $C_{\rm ox(inv)}L_g\Delta W = 0.0348$ fF/finger, and the other from the finger-end fringing capacitance, i.e., $C_{f(\text{poly-end})} = 0.064$ fF/finger. The former resulted from $\Delta W = 37.9$ nm occupies around 35% of α, and the latter contributes the remaining 65%. Both terms cannot be eliminated using any deembedding methods. The analysis proves the results shown in Fig. 3 and explains why it is impossible to achieve zero slope in $C_{gg(\rm DUT,OM1)}$ versus N_F under fixed $W_{\rm tot} = W_F \times N_F$, even through open-M1 deembedding. As for multi-OD NMOS, the extracted ΔW is 42.9 nm. It is larger than that of narrow-OD NMOS due to the sharper trench profile associated with the denser OD/STI layout.

IV. EFFECTIVE MOBILITY EXTRACTION AND ANALYSIS

A. Effective Mobility $\mu_{\rm eff}$ Extraction Method

In the following, effective mobility $\mu_{\rm eff}$ can be extracted from linear I–V characteristics according to (16)–(22), in which effective channel length $L_{\rm eff}$ and series drain/source resistances R_D and R_S can be determined by the decoupled C–V method from our previous work [15] based on the precisely determined L_g. Furthermore, effective channel width $W_{\rm eff}$ in (19) and inversion carrier density $Q_{\rm inv}$ in (20) can be obtained with accurately extracted ΔW and $C_{\rm ox(inv)}$, as shown in Table I. Note that $Q_{\rm inv}$ is calculated by $C_{\rm ox(inv)}$ and gate overdrive $(V_{\rm GS} - V_T - \lambda V_{\rm DS})$ in the linear region. Herein, $C_{\rm ox(inv)}$ is the intrinsic gate capacitance density per unit area of the inversion channel, given by $C_{gg,\rm int}/L_g$, and $C_{gg,\rm int}$ is the intrinsic gate capacitance density determined by (22) [or (13)], in which intercept β has been extracted from the linear function of $C_{gg(\rm DUT,OM1)}$ versus N_F under fixed $W_{\rm tot} = W_F \times N_F$ [see Fig. 3(b)], and $C_{\rm of}$ is the gate sidewall fringing capacitance calculated by Raphael simulation (see Fig. 5).

$$I_{\rm DS} = W_{\rm eff}Q_{\rm inv}\mu_{\rm eff}\frac{V_{\rm DS0}}{L_{\rm eff}} \tag{16}$$

$$\mu_{\rm eff} = \frac{I_{\rm DS}}{V_{\rm DS0}}\frac{1}{\frac{W_{\rm eff}}{L_{\rm eff}}Q_{\rm inv}} \tag{17}$$

$$V_{\rm DS0} = V_{\rm DS} - I_{\rm DS}(R_D + R_S) \tag{18}$$

$$W_{\rm eff} = N_F(W_F + \Delta W) \tag{19}$$

$$Q_{\rm inv} = C_{\rm ox(inv)}(V_{\rm GS} - V_T - \lambda V_{\rm DS}) \tag{20}$$

$$C_{\rm ox(inv)} = \frac{C_{gg,\rm int}}{L_g} \tag{21}$$

$$C_{gg,\rm int} = \frac{\beta}{N_F \times W_F} - C_{\rm of} \tag{22}$$

$$\lambda \le \frac{1}{2} \quad \text{for short-channel MOSFETs in the linear region.}$$

Note that the differences between the new and conventional methods appear in $C_{\rm ox(inv)}$ and $W_{\rm eff}$, and there is nearly no difference in $(V_{\rm GT} - \lambda V_{\rm DS})$. The influence from λ variation on $Q_{\rm inv}$ and the extracted $\mu_{\rm eff}$ is negligible in the strong inversion region due to $V_{\rm GT} \gg \lambda V_{\rm DS}$. However, the influence may become significant in the weak inversion region when $V_{\rm GT}$ approaches $\lambda V_{\rm DS}$.

B. Effective Mobility $\mu_{\rm eff}$ Extraction and Analysis for Narrow-OD NMOS

Fig. 7 presents the linear $I_{\rm DS}$ and G_m measured from narrow-OD NMOS under varying $V_{\rm GT}$ $(V_{\rm GT} = V_{\rm GS} - V_T)$. The results indicate that the narrower width ($W_F = W_{\rm OD}$ for narrow-OD devices) leads to $I_{\rm DS}$ and G_m degradation. As shown in Fig. 8(a), V_T versus $W_{\rm OD}$ for narrow-OD NMOS presents an obvious inverse narrow-width effect (INWE), which is generally originated from the STI process [16]. Fig. 8(b) reveals a monotonic decrease in maximum G_m $(G_{m,\max})$ with $W_{\rm OD}$ scaling. As compared with the standard reference (W2N16), $G_{m,\max}$ degradation is around 2% for W1N32, and it is significantly increased to as large as 8% for W05N64, with four times narrower width than W2N16. The monotonic

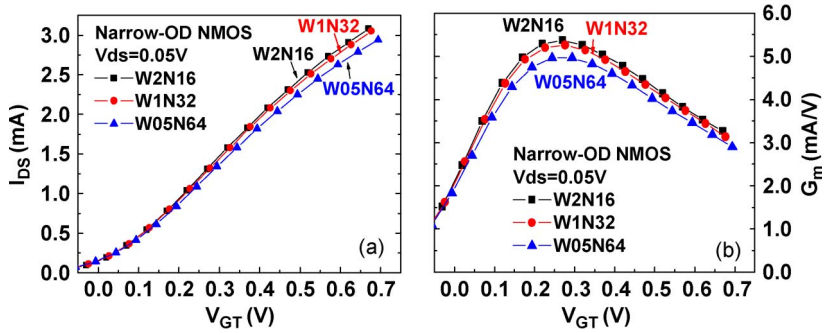

Fig. 7. (a) Drain current $I_{\rm DS}$ versus $V_{\rm GT}$. (b) Transconductance G_m versus $V_{\rm GT}$ in the linear region ($V_{\rm DS} = 50$ mV), as measured from narrow-OD NMOS (W1N32 and W05N64) and standard multifinger NMOS (W2N16). All of the devices have the same total width, i.e., $W_F \times N_F = 32\ \mu$m.

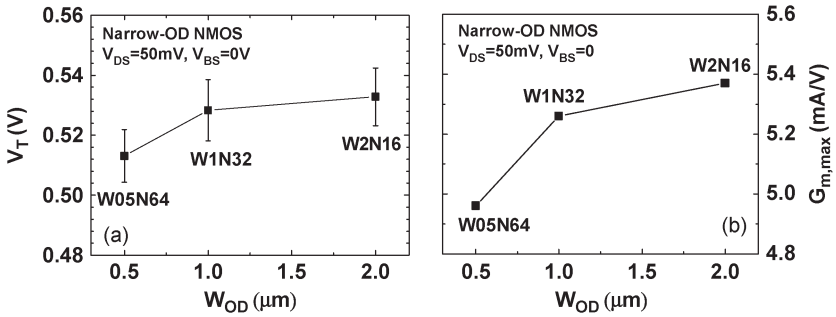

Fig. 8. (a) Linear V_T versus $W_{\rm OD}$. (b) $G_{m,\max}$ versus $W_{\rm OD}$ for narrow-OD NMOS (W1N32 and W05N64) and standard multifinger NMOS (W2N16), under the biases of $V_{\rm DS} = 50$ mV and $V_{\rm BS} = 0$ V.

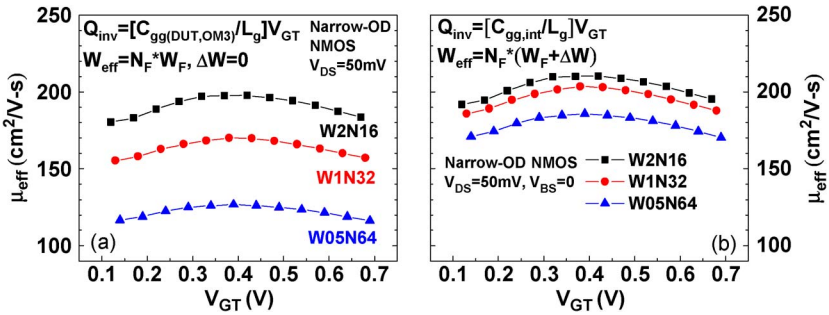

Fig. 9. Effective mobility $\mu_{\rm eff}$ versus $V_{\rm GT}$ for narrow-OD NMOS. (a) Conventional method with approximation: $Q_{\rm inv} = C_{gg(\rm DUT,OM3)}/L_g \times V_{\rm GT}$, $\Delta W = 0$, and $W_{\rm eff} = W_F \times N_F$. (b) New method with open-M1 deembedding and extraction of $C_{\rm of}$ and $C_{f(\rm poly\text{-}end)}$ for $C_{gg,\rm int}$ and ΔW for $W_{\rm eff}$: $Q_{\rm inv} = C_{gg,\rm int}/L_g \times V_{\rm GT}$ and $W_{\rm eff} = N_F(W_F + \Delta W)$.

degradation of G_m with $W_{\rm OD}$ scaling suggests that the increase in STI transverse stress $\sigma_\perp$ along the direction of width, which is generally compressive, may be the dominant factor responsible for $\mu_{\rm eff}$ degradation and the resulted G_m degradation.

To verify the layout dependence of $\mu_{\rm eff}$, the conventional and new methods were applied to each specified device for a comparison. Note that the new method has been described by (16)–(22), in which $Q_{\rm inv}$ is determined by $(C_{gg,\rm int}/L_g) \times (V_{\rm GT} - \lambda V_{\rm DS})$, and $C_{gg,\rm int}$ is calculated by (22), with intercept β extracted from $C_{gg(\rm DUT,OM1)}$ versus N_F and $C_{\rm of}$ extracted from Raphael simulation. As for the conventional method, $Q_{\rm inv}$ is approximated by $(C_{gg(\rm DUT,OM3)}/L_g \times (V_{\rm GT} - \lambda V_{\rm DS})$ without consideration of $C_{\rm of}$ and $C_{f(\rm poly\text{-}end}$, and $W_{\rm eff}$ is assumed equal to $W_{\rm tot} = W_F \times N_F$, without ΔW. The difference between $C_{gg(\rm DUT,OM1)}$ and $C_{gg(\rm DUT,OM3)}$ can be referred to in Fig. 3(a) and (b), in which $C_{gg(\rm DUT,OM3)}$ appears much larger than $C_{gg(\rm DUT,OM1)}$ and leads to an overestimation of $Q_{\rm inv}$. As shown in Fig. 9(a) and (b), the $\mu_{\rm eff}$ extracted by the conventional and new methods indicates the same trend of $\mu_{\rm eff}$ variation with decreasing W_F, i.e., the narrower W_F leading to the smaller $\mu_{\rm eff}$. However, the degradation of $\mu_{\rm eff}$ in W1N32 and W05N64 compared with W2N16 is much larger for those extracted by the conventional method than those determined by the new method. The results suggest that the $\mu_{\rm eff}$ extracted by the conventional method is underestimated due to an overestimation of $Q_{\rm inv}$ from invalid approximation of $C_{\rm ox(inv)}$ by $C_{gg(\rm DUT,OM3)}/L_g$. This deviation increases with width scaling and becomes significant for $W_F = W_{\rm OD} = 1\ \mu$m and below. Using the new method, $\mu_{\rm eff}$ degradation is as small as 1.3%–2.3% for W1N32 and increases to 5.5%–8% for W05N64 compared with W2N16. However, the conventional method leads to an overestimation of $\mu_{\rm eff}$ degradation to 14.6%–15.9% for W1N64 and even much worse to 42.5%–45% for W05N64.

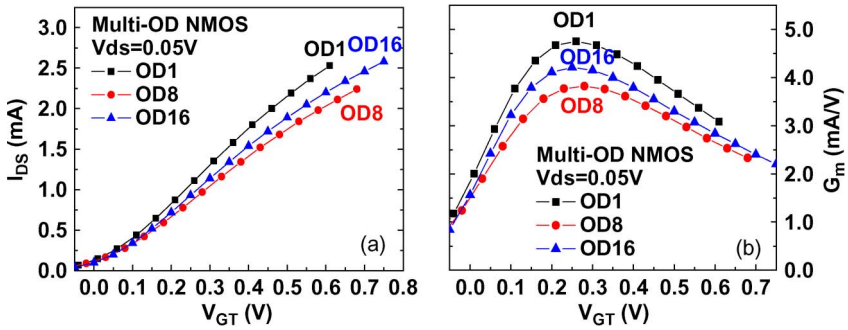

Fig. 10. (a) Drain current $I_{\rm DS}$ versus $V_{\rm GT}$. (b) Transconductance G_m versus $V_{\rm GT}$ in the linear region ($V_{\rm DS} = 50$ mV), as measured from multi-OD NMOS OD8 ($N_{\rm OD} = 8$, $W_{\rm OD} = 0.25$ μm) and OD16 ($N_{\rm OD} = 16$, $W_{\rm OD} = 0.125$ μm) and standard reference OD1 ($N_{\rm OD} = 1$, $W_{\rm OD} = 2$ μm). OD1 = W2N16. All of the devices have the same finger number, i.e., $N_F = 16$.

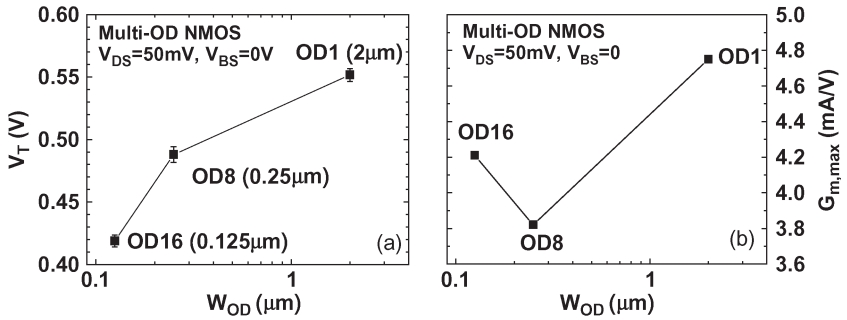

Fig. 11. (a) Linear V_T versus $W_{\rm OD}$. (b) $G_{m,\max}$ versus $W_{\rm OD}$ for multi-OD NMOS OD8 and OD16 and standard multifinger NMOS OD1(W2N16), under the biases of $V_{\rm DS} = 50$ mV and $V_{\rm BS} = 0$ V.

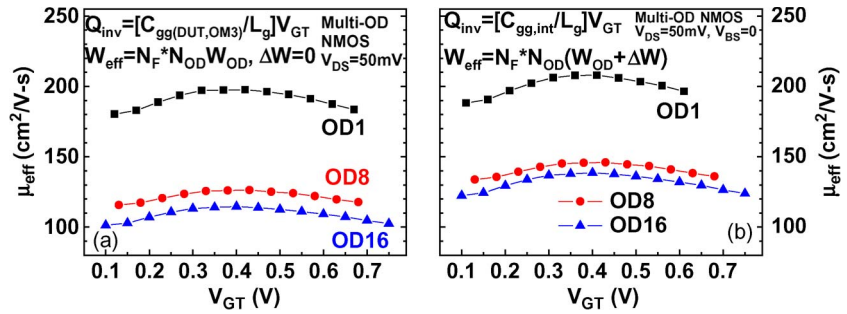

Fig. 12. Effective mobility $\mu_{\rm eff}$ versus $V_{\rm GT}$ for multi-OD NMOS. (a) Conventional method with approximation: $Q_{\rm inv} = C_{gg(\rm DUT,OM3)}/L_g \times V_{\rm GT}$, $\Delta W = 0$, and $W_{\rm eff} = N_F \times N_{\rm OD} W_{\rm OD}$. (b) New method with open-M1 deembedding and extraction of $C_{\rm of}$ and $C_{f(\rm poly\text{-}end)}$ for $C_{gg,\rm int}$ and ΔW for $W_{\rm eff}$: $Q_{\rm inv} = C_{gg,\rm int}/L_g \times V_{\rm GT}$ and $W_{\rm eff} = N_F \times N_{\rm OD}(W_{\rm OD} + \Delta W)$.

C. Effective Mobility $\mu_{\rm eff}$ Extraction and Analysis for Multi-OD NMOS

Fig. 10 presents $I_{\rm DS}$ and G_m versus $V_{\rm GT}$ ($V_{\rm DS} = 0.05$ V) measured from multi-OD NMOS, such as OD8 ($N_{\rm OD} = 8, W_{\rm OD} = 0.25$ μm) and OD16 ($N_{\rm OD} = 16, W_{\rm OD} = 0.125$ μm), and the standard reference, such as OD1 ($N_{\rm OD} = 1, W_{\rm OD} = 2$ μm). The results indicate degradation of both $I_{\rm DS}$ and G_m in OD8 compared with OD1, but further scaling of $W_{\rm OD}$ to 0.125 μm in OD16 leads to an increase in $I_{\rm DS}$ and G_m compared with OD8, and the degradation compared with OD1 is reduced. As for V_T versus $W_{\rm OD}$ of the multi-OD NMOS shown in Fig. 11(a), it presents an obvious INWE, which is similar to narrow-OD NMOS [see Fig. 8(a)] and will actually appear in a universal curve when plotted together with those of narrow-OD NMOS [1]. However, $G_{m,\max}$ versus $W_{\rm OD}$ shown in Fig. 11(b) reveals a nonmonotonic variation of $G_{m,\max}$ with $W_{\rm OD}$ scaling. $G_{m,\max}$ degradation is around 21.7% for OD8 compared with OD1, but further scaling of $W_{\rm OD}$ to 0.125 μm for OD16 results in $G_{m,\max}$ increase of around 9.7% compared with OD8. This layout dependence of $I_{\rm DS}$ and G_m appears unusual and cannot be explained by STI compressive stress alone.

To explore the mechanism responsible for the unusual results measured from multi-OD NMOS with extremely narrow $W_{\rm OD}$, $\mu_{\rm eff}$ was extracted from multi-OD NMOS to verify the layout dependence and identify the impact on $I_{\rm DS}$ and G_m from $\mu_{\rm eff}$ or other parameters. As shown in Fig. 12(a) and (b), the $\mu_{\rm eff}$ extracted by the conventional and new methods for multi-OD NMOS (OD8 and OD16) and the standard reference (OD1 = W2N16) indicate the same trend of $\mu_{\rm eff}$ variation with decreasing $W_{\rm OD}$, i.e., the smaller $\mu_{\rm eff}$ associated with the narrower $W_{\rm OD}$. However, the degradation of $\mu_{\rm eff}$ in OD8 and OD16 compared with OD1 is larger for those extracted by the conventional method than those determined by the new method. Again, the results suggest that the $\mu_{\rm eff}$ extracted by the conventional method is underestimated due to an overestimation of $Q_{\rm inv}$

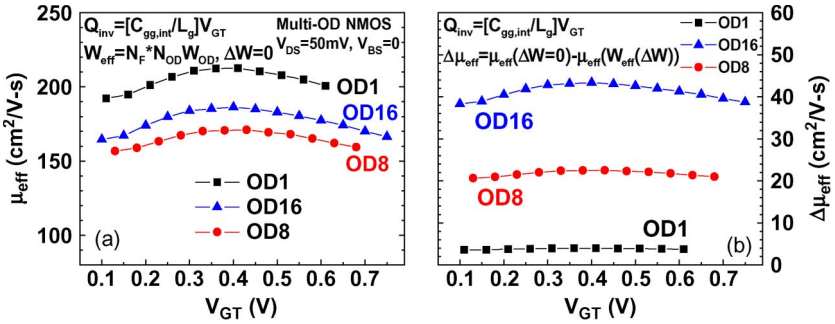

Fig. 13. Effective mobility $\mu_{\rm eff}$ versus $V_{\rm GT}$ for multi-OD NMOS $Q_{\rm inv} = C_{gg,{\rm int}}/L_g \times V_{\rm GT}$ in the new method. (a) $\mu_{\rm eff}(\Delta W = 0)$: ΔW is removed, and $W_{\rm eff} = N_F \times N_{\rm OD} W_{\rm OD}$. (b) $\Delta\mu_{\rm eff} = \mu_{\rm eff}(\Delta W = 0) - \mu_{\rm eff}(\Delta W)$, $\mu_{\rm eff}(\Delta W)$: $W_{\rm eff} = N_F \times N_{\rm OD}(W_{\rm OD} + \Delta W)$.

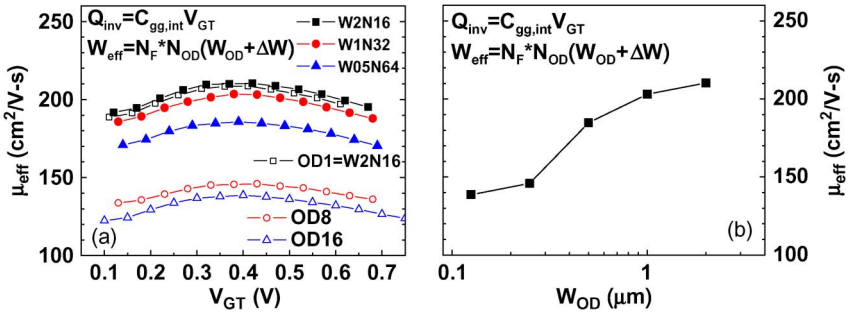

Fig. 14. Effective mobility $\mu_{\rm eff}$ extracted by the new method. (a) $\mu_{\rm eff}$ versus $V_{\rm GT}$. (b) $\mu_{\rm eff}$ versus $W_{\rm OD}$ for narrow-OD NMOS (W2N16, W1N32, and W05N64) and multi-OD NMOS (OD1, OD8, and OD16).

originated from invalid use of $C_{\rm ox(inv)}$ by $C_{gg({\rm DUT,OM3})}/L_g$. This deviation increases with width scaling and becomes significant for $W_{\rm OD} = 0.25$ μm and below. Using the new method, $\mu_{\rm eff}$ degradation is around 28%–33.6% for OD8 and increased to 36.5%–39.6% for OD16 compared with OD1. However, the conventional method leads to an overestimation of $\mu_{\rm eff}$ degradation to 43.6%–44.3% for OD8 and 52.5%–56.2% for OD16. Note that the $\mu_{\rm eff}$ extracted by the new method indicates a monotonic decrease with $W_{\rm OD}$ scaling, which suggests that $\mu_{\rm eff}$ is not the factor causing the nonmonotonic variations of $I_{\rm DS}$ and G_m under decreasing $W_{\rm OD}$ (see Fig. 10). The increase in $W_{\rm eff}$ from ΔW is considered as the primary parameter responsible for the increase in $I_{\rm DS}$ and G_m for OD16 with extremely narrow $W_{\rm OD}$ to 0.125 μm.

Under the condition that ΔW is removed from $W_{\rm eff}$ given by (19), the $\mu_{\rm eff}$ extracted from multi-OD NMOS shown in Fig. 13(a) just reproduces a nonmonotonic variation w.r.t. $W_{\rm OD}$ scaling, which is the same trend as that of $I_{\rm DS}$ and G_m. The impact of ΔW on the extracted $\mu_{\rm eff}$ can be verified by the difference between those with and without taking ΔW into $W_{\rm eff}$, i.e., $\Delta\mu_{\rm eff} = \mu_{\rm eff}(\Delta W = 0) - \mu_{\rm eff}(\Delta W)$, as shown in Fig. 13(b). The positive $\Delta\mu_{\rm eff}$ suggests that $\mu_{\rm eff}(\Delta W = 0)$ is overestimated by neglecting ΔW. Note that $\Delta\mu_{\rm eff}$ is less than 4 cm^2/V $\cdot$ s for OD1 ($W_{\rm OD} = 2$ μm), then increases to around 22 cm^2/V $\cdot$ s for OD8 with $W_{\rm OD} = 0.25$ μm, and further increases to more than 40 cm^2/V $\cdot$ s for OD16 with $W_{\rm OD} = 0.125$ μm. The above analysis explains how the increase in $W_{\rm eff}$ from ΔW can dominate the impact of $\mu_{\rm eff}$ degradation and lead to the increase in $I_{\rm DS}$ and G_m in multifinger MOSFETs with extremely narrow width.

D. Layout Dependence of $\mu_{\rm eff}$ in Multifinger MOSFETs

The effective mobility $\mu_{\rm eff}$ extracted by the new method for both narrow- and multi-OD devices are presented together in Fig. 14(a) to verify the layout dependence of $\mu_{\rm eff}$ in multifinger MOSFETs. The comparison indicates that $\mu_{\rm eff}$ degradation compared with the standard reference (W2N16) is not serious for narrow-OD NMOS (W1N32 and W05N64), with channel-width scaling limited to $W_F = W_{\rm OD} = 0.5$ μm, but becomes significantly worse for multi-OD NMOS, such as OD8 and OD16, with more aggressively scaled width to $W_{\rm OD} = 0.25{-}0.125$ μm. The results suggest that $\mu_{\rm eff}$ is a universal function of channel width $W_{\rm OD}$, no matter whether in the same or different layouts such as narrow- or multi-OD. Fig. 14(b) presents $\mu_{\rm eff}$ versus $W_{\rm OD}(V_{\rm GT} = 0.38$ V) based on the data from narrow- and multi-OD NMOS and indicate a monotonic degradation of $\mu_{\rm eff}$ with decreasing $W_{\rm OD}$. The degradation is not significant in the region of $W_{\rm OD} \geqq 0.5$ μm, but a sharp degradation happens in the transition region from $W_{\rm OD} = 0.5$ μm to 0.25 μm. The significant degradation of $\mu_{\rm eff}$ when scaling $W_{\rm OD}$ from 0.5 to 0.25 μm suggests that the penetration of STI stress from the STI edge into the channel center gradually covers the whole active channel region. With regard to the gradual level off of $\mu_{\rm eff}$ degradation in the region below 0.25 μm, it indicates that STI stress has covered the whole active channel for $W_{\rm OD} \leq 0.25$ μm, and the influence from further $W_{\rm OD}$ scaling becomes very minor. The results provide a helpful guideline for multifinger MOSFET layout optimization with a tradeoff among $\mu_{\rm eff}$, R_g, and parasitic capacitances, such as $C_{f({\rm poly\text{-}end})}$ and $C_{\rm of}$.

V. Conclusion

A new method for layout-dependent parasitic capacitance analysis and $\mu_{\rm eff}$ extraction has been developed for multifinger MOSFETs with narrow width. An improved open deembedding technique, i.e., open-M1 deembedding, can eliminate the extrinsic parasitic capacitances from the pads, interconnection lines, and lossy substrate. However, a 3-D interconnect simulation is required to extract the intrinsic parasitic capacitances such as gate sidewall and finger-end fringing capacitances, i.e., $C_{\rm of}$ and $C_{f(\text{poly-end})}$, respectively. The $C_{gg(\rm DUT,OM1)}$ achieved by open-M1 deembedding indicate a linear function of N_F in which intercept β is contributed from two terms: one is intrinsic gate capacitance $C_{gg,\rm int}W_{\rm tot}$, and the other is $C_{\rm of}W_{\rm tot}$. The former is responsible for the inversion carriers and channel current $I_{\rm DS}$ and can be extracted from β with known $C_{\rm of}W_{\rm tot}$. The slope α of the linear function can be decomposed into two terms: one is the gate capacitance contributed from ΔW, which is given by $C_{\rm ox(inv)}\Delta {\rm WL}_g$, and the other is $C_{f(\text{poly-end})}$. Again, the former is an additional component of the intrinsic gate capacitance with contribution to the inversion carriers and may compensate the $\mu_{\rm eff}$ degradation effect on $I_{\rm DS}$ and G_m, resulting to a nonmonotonic variation of $I_{\rm DS}$ and G_m with $W_{\rm OD}$ scaling, in extremely narrow devices. The $\mu_{\rm eff}$ extracted by this new method for narrow- and multi-OD devices indicates a universal function of $W_{\rm OD}$ and a monotonic degradation of $\mu_{\rm eff}$ with decreasing $W_{\rm OD}$. The degradation is not significant in the region of $W_{\rm OD} \geqq 0.5$ μm, but a sharp degradation appears in the transition region from $W_{\rm OD} = 0.5$ μm to 0.25 μm. The results provide a useful guideline for multifinger MOSFET layout optimization with a tradeoff among $\mu_{\rm eff}$, G_m, R_g, and parasitic capacitances, which is important for RF and analog circuit design.

Acknowledgment

The authors would like to thank the National Nano Device Laboratories for RF measurement and the Chip Implementation Center for device fabrication.

References

[1] K.-L. Yeh and J.-C. Guo, "The impact of layout dependent STI stress and effective width on low frequency noise and high frequency performance in nanoscale nMOSFETs," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 3092–3100, Nov. 2010.

[2] J. Koomen, "Investigation of the MOST channel conductance in weak inversion," *Solid State Electron.*, vol. 16, no. 7, pp. 801–810, Jul. 1973.

[3] C. Sodini, T. Ekstedit, and J. Moll, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid State Electron.*, vol. 25, no. 9, pp. 833–841, Sep. 1982.

[4] S. Takagi and M. Takayanagi, "Experimental evidence of inversion-layer mobility lowering in ultrathin gate oxide metal–oxide–semiconductor field-effect-transistors with direct tunneling current," *Jpn. J. Appl. Phys.*, vol. 41, no. 4B, pp. 2348–2352, 2002.

[5] F. Lime, C. Guiducci, R. Clerc, G. Ghibaudo, C. Leroux, and T. Ernst, "Characterization of effective mobility by split $C(V)$ technique in N-MOSFETs with ultra thin gate oxide," *Solid State Electron.*, vol. 47, no. 7, pp. 1147–1153, Jul. 2003.

[6] K. Romanjek, F. Andrieu, T. Ernst, and G. Ghibaudo, "Improved split C–V method for effective mobility extraction in sub-0.1 μm Si MOSFETs," *IEEE Electron Device Lett.*, vol. 25, no. 8, pp. 583–585, Aug. 2004.

[7] Y.-W. Chang, H.-W. Chang, T.-C. Lu, Y.-C. King, K.-C. Chen, and C.-Y. Lu, "Combining a novel charge-based capacitance measurement (CBCM) technique and split C–V method to specifically characterize the STI stress effect along the width direction of MOSFET devices," *IEEE Electron Device Lett.*, vol. 29, no. 6, pp. 641–644, Jun. 2008.

[8] J.-C. Guo and Y.-M. Lin, "A lossy substrate de-embedding method for sub-100 nm RF CMOS noise extraction and modeling," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 339–347, Feb. 2006.

[9] J.-C. Guo and Y.-H. Tsai, "A broadband and scalable lossy substrate model for RF noise simulation and analysis in nanoscale MOSFETs with various pad structures," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no. 2, pp. 271–281, Feb. 2009.

[10] J.-C. Guo and C.-T. Yeh, "A new three-dimensional capacitor model for accurate simulation of parasitic capacitances in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 8, pp. 1598–1607, Aug. 2009.

[11] N. R. Mohapatra, M. P. Desai, S. G. Narendra, and V. R. Rao, "Modeling of parasitic capacitances in deep submicrometer conventional and high-K dielectric MOS transistors," *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 959–966, Apr. 2003.

[12] K. Suzuki, "Parasitic capacitance of submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 46, no. 9, pp. 1895–1900, Sep. 1999.

[13] R. Shrivastava and K. Fitzpatrick, "A simple model for the overlap capacitance of a VLSIMOS device," *IEEE Trans. Electron Devices*, vol. ED-29, no. 12, pp. 1870–1875, Dec. 1982.

[14] M. I. Elmasry, "Capacitance calculations in MOSFET VLSI," *IEEE Electron Device Lett.*, vol. EDL-3, no. 1, pp. 6–7, Jan. 1982.

[15] J.-C. Guo, S.-S. Chung, and C.-H. Hsu, "A new approach to determine the effective channel length and the drain-and-source series resistance of miniaturized MOSFETs," *IEEE Trans. Electron Devices*, vol. 41, no. 10, pp. 1811–1818, Oct. 1994.

[16] A. Ono, R. Ueno, and I. Sakai, "TED control technology for suppression of reverse narrow channel effect in 0.1 μm MOS devices," in *IEDM Tech. Dig.*, 1997, pp. 227–230.

Kuo-Liang Yeh (M'09) received the B.S.E.E. degree from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1995 and the M.S.E.E. degree from National Taiwan University, Taipei, Taiwan, in 1997. He is currently working toward the Ph.D. degree in electronics engineering at NCTU.

In 1999, he joined Taiwan Semiconductor Manufactory Company, Inc., Hsinchu, where he worked on process integration and yield improvement. From 2004 to 2007, he was a Senior Engineer with MediaTek Inc., Hsinchu. He is currently a Senior Manager with Silicon Motion Technology Corporation, Hsinchu. He has authored more than ten technical publications in international journals and conference proceedings. His research interests include the characterization and parameter extraction of CMOS devices for modeling and circuit simulation, as well as the protection of intellectual property rights.

Jyh-Chyurn Guo (M'06–SM'07) received the B.S.E.E. and M.S.E.E. degrees from National Tsing Hua University, Hsinchu, Taiwan, in 1982 and 1984, respectively, and the Ph.D. degree in electronics engineering from National Chiao Tung University (NCTU), Hsinchu, in 1994.

For more than 19 years, she was with the semiconductor industry, where her major focus was on device design and VLSI technology development. In 1984, she joined ERSO/ITRI, where she was engaged in semiconductor integrated circuit technologies with a broad scope that covers high voltage, high power, submicrometer project, high-speed SRAM technologies, etc. From 1994 to 1998, she was with Macronix International Corporation, Hsinchu, where she engaged in high-density and low-power Flash memory technology development. In 1998, she joined Vanguard International Semiconductor Corporation, where she assumed responsibility of Device Department Manager for advanced DRAM device technology development. In 2000, she joined Taiwan Semiconductor Manufacturing Company (TSMC), Hsinchu, where she served as a Program Manager in charge of 100-nm logic CMOS FEOL, high-performance analog (HPA), and RF CMOS technology development. In 2003, she joined as an Associate Professor with the Department of Electronics Engineering, NCTU, where she has been a Full Professor since 2008. She has authored or coauthored more than 60 technical papers. She is the holder of 19 U.S. patents in her professional field. Her current research interests cover RF/MS CMOS device design and modeling for low power and low noise, nanoscale CMOS noise modeling and strain engineering effects, broadband and scalable inductors modeling, novel nonvolatile memory technologies, and device integration technologies for SoC.