Narrow-Width Effect on High-Frequency Performance and RF Noise of Sub-40-nm Multifinger nMOSFETs and pMOSFETs

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

***Abstract*—The impact of narrow-width effects on high-frequency performance like f_T, $f_{\rm MAX}$, and RF noise parameters, such as $NF_{\min}$ and R_n, in sub-40-nm multifinger CMOS devices is investigated in this paper. Narrow-oxide-diffusion (OD) MOSFET with smaller finger width and larger finger number can achieve lower R_g and higher $f_{\rm MAX}$. However, these narrow-OD devices suffer f_T degradation and higher $NF_{\min}$, even with the advantage of lower R_g. The mechanisms responsible for the tradeoff between different parameters will be presented to provide an important guideline of multifinger MOSFET layout for RF circuit design using nanoscale CMOS technology.**

***Index Terms*—$f_{\rm MAX}$, f_T, multifinger, nanoscale CMOS, narrow width, $NF_{\min}$, RF noise, R_g.**

I. Introduction

NANOSCALE CMOS devices adopting multifinger layout have been extensively used in modern RF circuits and proven successful to improve high-frequency performance, such as higher f_T and $f_{\rm MAX}$, driven by gate length scaling, and lower RF noise due to lower gate resistance (R_g) from multifinger structure [1]–[5]. Unfortunately, the continuous increase of finger number (N_F) and reduction of finger width (W_F) for smaller R_g may lead to the penalty of lower transconductance (g_m) and larger parasitic capacitances. The former one comes from stress-induced mobility degradation, and the latter one stems from gate-related fringing capacitances [6], [7]. Both cannot be scalable with device scaling, and the impact may dominate high-frequency characteristics in nanoscale devices.

In our previous work on multifinger devices in 90-nm CMOS technology [7], we achieved the important finding that the kind of parasitic capacitances, which cannot be removed by existing open deembedding, was contributed from the poly finger sidewall and finger-end fringing capacitances, namely, $C_{\rm of}$ and $C_{f({\rm poly\ end})}$. Note that $C_{f({\rm poly\ end})}$ increases linearly with N_F and the impact on high-frequency performance may

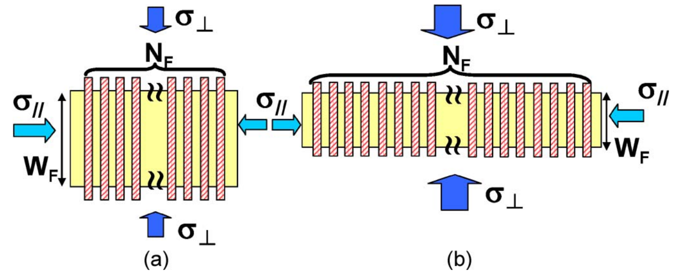

Fig. 1. Schematics of multifinger MOSFETs with different layouts. (a) Standard multifinger device: $W_F \times N_F = 2\ \mu$m × 32 (W2N32). (b) Narrow-OD device: $W_F \times N_F = 1\ \mu$m × 64 (W1N64) and 0.5 μm × 128 (W05N128).

become significant in multifinger devices with very narrow W_F and very large N_F. Meanwhile, ΔW caused by STI top corner rounding (TCR) becomes a critical parameter when trying to minimize R_g by continuously increasing N_F and reducing W_F [7]. Both $C_{f({\rm poly\ end})}$ and ΔW appear as two key factors for an accurate extraction of effective mobility $(\mu_{\rm eff})$ in multifinger devices with narrow width. Also, ΔW may offset the impact of STI stress on $\mu_{\rm eff}$. However, the mentioned study was focused on dc characteristics and flicker noise limited to nMOSFETs, and the potential impact from narrow width on high-frequency performance and RF noise remains unknown and deserves further research effort.

In this paper, the impact of aggressive width scaling on f_T, $f_{\rm MAX}$, and RF noise parameters in multifinger devices will be explored. The investigation will be carried out on both nMOS and pMOS, which were fabricated by 65-nm CMOS process, with gate length aggressively scaled to below 40 nm.

II. Device Fabrication and Characterization

In this paper, multifinger MOSFETs were fabricated in 65-nm CMOS process with gate oxide of 1.4-nm physical thickness. The gate length drawn on the layout is 60 nm, i.e., $L_{\rm drawn} = 60$ nm, and the total gate width $W_{\rm tot}$ is fixed at 64 μm. Fig. 1(a) and (b) shows multifinger MOSFET layouts, namely, standard and narrow-oxide-diffusion (OD) devices in which $\sigma_{//}$ and $\sigma_\perp$ denote the longitudinal and transverse stresses introduced from STI (OD means oxide diffusion, i.e., active area). The finger width was reduced from $W_F = 2\ \mu$m for standard device (W2N32) to $W_F = 1\ \mu$m and $W_F = 0.5\ \mu$m for narrow-OD devices (W1N64 and W05N128) to investigate the impact from layout-dependent stress $(\sigma_\perp)$,

Manuscript received June 14, 2012; revised September 21, 2012; accepted November 7, 2012. Date of current version December 19, 2012. 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 J. C. S. Woo.

K.-L. Yeh is with the The institute of Electronics Engineering, National Chiao Tung University, Hsinchu 30010, Taiwan, and also with Silicon Motion Technology Corporation, Hsinchu 30265, Taiwan (e-mail: rexyeh@yahoo.com).

J.-C. Guo is with the Department of Electronics Engineering and Institute of Electronics, National Chiao-Tung University, Hsinchu 30010, 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.2012.2228196

0018-9383/$31.00 © 2012 IEEE

TABLE I
STRESS FAVORABLE FOR MOBILITY ENHANCEMENT IN nMOS AND pMOS, ALONG LONGITUDINAL AND TRANSVERSE DIRECTIONS [12]

Directions	Stress favorable for mobility enhancement	
	NMOS	PMOS
Longitudinal ($\sigma_{//}$)	Tensile	Compressive
Transverse ($\sigma_{\perp}$)	Tensile	Tensile

parasitic capacitances, and parasitic resistances on high-frequency performance.

S-parameters were measured up to 40 GHz by Agilent network analyzer E8364B for high-frequency characterization and device parameter extraction. An open deembedding to the bottom metal (M1), namely, openM1, was performed to remove the extrinsic parasitic capacitances from the pads, interconnection lines, and substrate. Three-dimensional interconnect simulation was carried out using Raphael to extract the intrinsic parasitic capacitances, such as $C_{\rm of}$ and $C_{f(\rm poly\ end)}$ [7]. In this way, taking openM1 deembedding and Raphael simulation, both the extrinsic and intrinsic parasitic capacitances can be eliminated to obtain the truly intrinsic gate capacitances. Also, shortM1 deembedding was done to remove the parasitic resistances and inductances. H-parameters and unilateral gain (U) obtained from S-parameters after the mentioned openM1 and shortM1 deembedding can achieve f_T and $f_{\rm MAX}$ of the intrinsic device [8]–[11]. Also, Y- and Z-parameters can be used to determine the intrinsic device parameters, such as gate capacitances ($C_{\rm gg}$ and $C_{\rm gd}$), g_m, and R_g associated with the multifinger MOSFETs.

III. LAYOUT-DEPENDENT EFFECTS ON DEVICE CHARACTERISTICS OF MULTIFINGER MOSFET

Layout-dependent stress effect becomes increasingly significant with device scaling and has been known as an important factor influencing the carrier mobility, transconductance (g_m), and channel current $(I_{\rm DS})$. Table I shows a brief summary of the stress in different directions, which can favor mobility enhancement in nMOS and pMOS, respectively [8]. The tensile stress in longitudinal direction $(\sigma_{//})$ can enhance electron mobility in nMOS but degrades hole mobility in pMOS. As for the transverse stress $(\sigma_{\perp})$ of our major interest, the compressive stress from STI always leads to mobility degradation in both nMOS and pMOS [12], [13]. In this paper, narrow-OD layouts were implemented to enhance $\sigma_{\perp}$ and investigate the impact on device characteristics, with major focus extended to high-frequency performance and RF noise.

A. STI Transverse Stress Effect on g_m of Narrow-OD nMOS and pMOS

Transconductance g_m has been known as one of the key parameters governing high-frequency performance, noise, and also dc gain $(A_v = g_m/g_{\rm ds})$, which are of special concern in RF and analog circuit design. Fig. 2(a) and (b) shows g_m-versus-$V_{\rm GT}$ in linear region measured from nMOS and pMOS with different multifinger layouts, such as W2N32 (standard) and W1N64 and W05N128 (narrow OD), shown in Fig. 1.

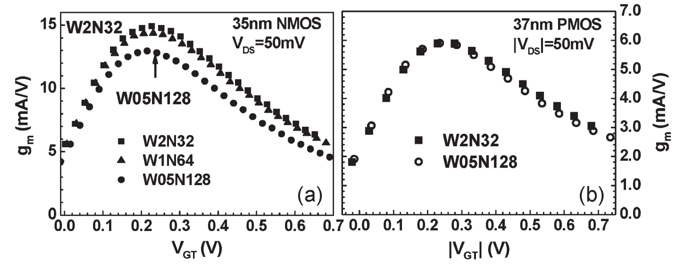

Fig. 2. Transconductance g_m-versus-$V_{\rm GT}$ in linear region, measured from (a) nMOS with W2N32, W1N64, and W05N128 and (b) pMOS with W2N32 and W1N64. $|V_{\rm DS}| = 50$ mV and $V_{\rm GT} = V_{\rm GS} - V_T$.

Note that $V_{\rm GT} = V_{\rm GS} - V_T$ is used to offset V_T variations due to inverse narrow-width effect [14]. The result for nMOS [Fig. 2(a)] indicates a monotonic degradation of g_m with W_F scaling and suggests that $\mu_{\rm eff}$ degradation from STI $\sigma_{\perp}$ (along the width) is the dominant factor responsible for the lower g_m suffered by narrow-OD nMOS. However, pMOS shown in Fig. 2(b) reveals a dramatically different result that W2N32 and W05N128 have nearly the same g_m over the full range of $V_{\rm GT}$. This interesting phenomenon suggests that the lower $\mu_{\rm eff}$ from narrower W_F is no longer the dominant factor determining g_m and other parameters may offset $\mu_{\rm eff}$ degradation and recover g_m to a comparable value. To explore the mechanism underlying the unusual narrow-width effect on g_m, we propose to revisit the fundamental I–V model for MOSFET in which the width scaling effect on the parameters other than $\mu_{\rm eff}$ should be taken into consideration.

B. Device Parameter Extraction and for Narrow-Width Effect Analysis in Multifinger nMOS and pMOS

According to the linear I–V model used for CMOS devices given by (1), g_m can be derived from the differential of $I_{\rm DS}$ w.r.t. $V_{\rm GS}$ and written as (2). Note that the effective width $(W_{\rm eff})$ for multifinger MOSFETs is expressed by (3) in which ΔW represents width extension created by STI TCR. We can understand from this model that the narrower W_F may lead to lower g_m due to $\mu_{\rm eff}$ degradation caused by larger STI $\sigma_{\perp}$, but an aggressive reduction of W_F may result in a significant increase of $W_{\rm eff}$ due to nonscalable ΔW and then contributes to higher g_m. The narrow-width effect on g_m is determined by the tradeoff between the lower $\mu_{\rm eff}$ and wider $W_{\rm eff}$. To verify and justify the mentioned argument, $\mu_{\rm eff}$ and ΔW appear as two key parameters and can be extracted by using our proprietary capacitance method [7]

$$I_{\rm DS} = W_{\rm eff} C_{\rm ox(inv)} (V_{\rm GS} - V_T - \lambda V_{\rm DS}) \mu_{\rm eff} \frac{V_{\rm DS}}{L_{\rm eff}} \quad (1)$$

$$g_m = \frac{\partial I_{\rm DS}}{\partial V_{\rm GS}} = W_{\rm eff} C_{\rm ox(inv)} \mu_{\rm eff} \frac{V_{\rm DS}}{L_{\rm eff}} \quad (2)$$

$$W_{\rm eff} = (W_F + \Delta W) \times N_F \quad (3)$$

where $C_{\rm ox(inv)}$ is the equivalent oxide thickness under inversion and $L_{\rm eff}$ is the effective channel length.

In the following, the basic device parameters, such as L_g, $T_{\rm ox(inv)}$ or $C_{\rm ox(inv)}$, $C_{\rm of}$, and ΔW, will be determined, and

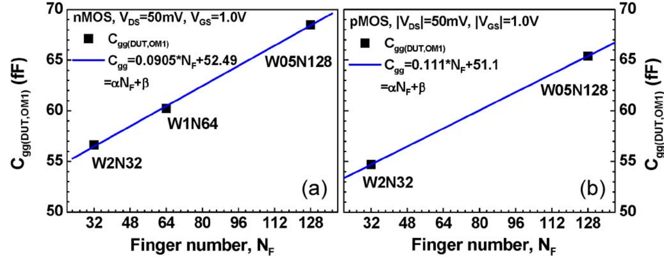

Fig. 3. $C_{\mathrm{gg(DUT,OM1)}}$-versus-N_F extracted by using openM1 deembedding on multifinger MOSFETs with various W_F and N_F under fixed $W_F \times N_F = 64\ \mu$m. (a) nMOS: W2N32, W1N64, and W05N128. (b) pMOS: W2N32 and W05N128.

then, μ_{eff} can be extracted with sufficient accuracy. First, Fig. 3(a) and (b) shows $C_{\mathrm{gg(DUT,OM1)}}$ extracted using openM1 deembedding for nMOS and pMOS with various W_F and N_F as specified. The $C_{\mathrm{gg(DUT,OM1)}}$-versus-N_F reveal a linearly increasing function. Theoretically, the truly intrinsic C_{gg} of multifinger devices with fixed $W_{\mathrm{tot}} = W_F \times N_F$ should remain constant under various N_F. The increase of $C_{\mathrm{gg(DUT,OM1)}}$ with N_F suggests the existence of some parasitic capacitances, which cannot be removed even using openM1 deembedding. Following our previous work [7], two components of gate-related fringing capacitance, such as C_{of} (finger sidewall) and $C_{f\mathrm{(poly\ end)}}$ (finger end), have been identified as the kind of intrinsic parasitic capacitances, which cannot be eliminated using existing open deembedding. According to our capacitance analysis method, the linear function of $C_{\mathrm{gg(DUT,OM1)}}$ versus N_F can be modeled by (4)–(6), in which $C_{f\mathrm{(poly\ end)}}$ and C_{of} contribute to the slope α and intercept β, respectively. Both C_{of} and $C_{f\mathrm{(poly\ end)}}$ can be calculated by Raphael. The simulation based on 65-nm CMOS technology and layout parameters achieves $C_{\mathrm{of}} = 0.23$ fF/μm and $C_{f\mathrm{(poly\ end)}} = 0.06764$ fF/finger. Referring to Fig. 3, α and β extracted from $C_{\mathrm{gg(DUT,OM1)}}$ versus N_F are 0.0905 fF/finger and 52.49 fF for nMOS and 0.111 fF/finger and 51.1 fF for pMOS. The physical gate length (L_g) can be extracted by (7), with known β and C_{of}, under specified W_F and N_F. The results indicate $L_g = 35$ and 37 nm for nMOS and pMOS, respectively. Finally, ΔW can be extracted from (8) with given α and $C_{f\mathrm{(poly\ end)}}$ and $C_{\mathrm{ox(inv)}}L_g$ determined by (6). Interestingly, ΔW extracted from pMOS is 77.1 nm, which is around two times larger than $\Delta W = 38.7$ nm for nMOS. This large ΔW can contribute more than 10% W_{eff} in W05N128, and it appears as a key factor offsetting μ_{eff} reduction due to STI $\sigma_{\perp}$. It can explain why the pMOS reveals nearly the same g_m between W2N32 and W05N128 [Fig. 2(b)]. Table II summarizes the device parameters extracted from the multifinger nMOS and pMOS following the mentioned extraction flow. Note that the extracted $T_{\mathrm{ox(inv)}}$ approaches the target $T_{\mathrm{ox(inv)}}$ with a difference as small as $0.5-1.0$ Å. It is important to justify the accuracy and precision of the basic parameters like $C_{\mathrm{ox(inv)}}$, L_g, and ΔW to ensure an accurate determination of μ_{eff}

$$C_{\mathrm{gg(DUT,OM1)}} = \alpha N_F + \beta \tag{4}$$

$$\alpha = (\Delta W \cdot L_g) C_{\mathrm{ox(inv)}} + C_{f\mathrm{(poly\ end)}} \tag{5}$$

$$\beta = \left(C_{\mathrm{ox(inv)}} L_g + C_{\mathrm{of}}\right) W_F N_F \tag{6}$$

TABLE II
MULTIFINGER nMOS AND pMOS DEVICE PARAMETERS

Device types		NMOS	PMOS
α	fF/finger	0.0905	0.1115
β	fF	52.49	51.1
L_g	μm	0.035	0.037
$T_{ox(inv)}$ target	A	20.0	21.5
$C_{of,sim}$	fF/μm	0.23	0.23
$C_{Polyend,sim}$	fF	0.06764	0.06764
$C_{ox(inv)}=(\beta/W_FN_F-C_{of})/L_g$	fF/μm^2	16.805	15.363
$T_{ox(inv)}=\varepsilon_0\varepsilon_{ox}/C_{ox(inv)}$	A	20.548	22.476
$\Delta W=(\alpha-C_{polyend})/(C_{ox(inv)}*L_g)$	μm	0.0387	0.0771

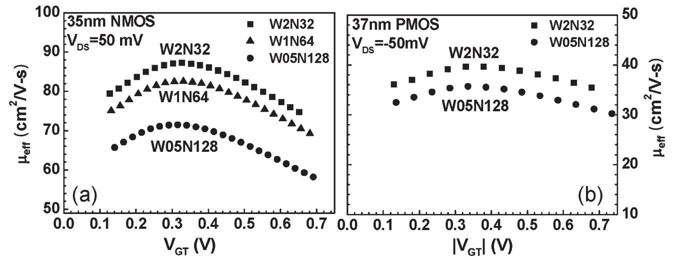

Fig. 4. μ_{eff}-versus-V_{GT} extracted from linear I–V for multifinger devices. (a) nMOS: W2N21, W1N64, and W05N128. (b) pMOS: W2N32 and W05N128. $W_{\mathrm{eff}} = N_F(W_F + \Delta W)$. $|V_{\mathrm{DS}}| = 50$ mV and $V_{\mathrm{GT}} = V_{\mathrm{GS}} - V_T$.

$$L_g = \frac{\beta}{W_F N_F C_{\mathrm{ox(inv)}}} - \frac{C_{\mathrm{of}}}{C_{\mathrm{ox(inv)}}} \tag{7}$$

$$\Delta W = \frac{\alpha - C_{f\mathrm{(poly\ end)}}}{C_{\mathrm{ox(inv)}} L_g}. \tag{8}$$

C. Effective Mobility Extraction and Narrow-Width Effect

The effective mobility μ_{eff} can be extracted from linear I–V model given by (1) and written as (9) in which W_{eff} has been accurately determined by including ΔW. Note that V_{DS} is the drain bias applied to the effective channel region defined by L_{eff} and obtained by subtracting the voltage drop across the parasitic S/D resistances, R_S and R_D. Herein, L_{eff}, R_S, and R_D can be determined by our decoupled C–V method [15] based on the precisely extracted L_g. Fig. 4(a) and (b) shows μ_{eff}-versus-V_{GT} extracted from nMOS and pMOS with specified split of W_F and N_F. The results indicate a monotonic reduction of μ_{eff} with W_F scaling (both nMOS and pMOS) and manifest the increasing impact of STI $\sigma_{\perp}$ on μ_{eff} in narrow-OD devices. To verify the accuracy of extracted μ_{eff} and the influence of ΔW, μ_{eff} extracted by assuming $\Delta W = 0$ is added into the same plot for a comparison, as shown in Fig. 5(a) and (b) for nMOS and pMOS, respectively. We can see that μ_{eff} determined by using extracted ΔW can achieve a good match with μ_{eff} calculated by $\mu_0[1 - k * \log(W_{\mathrm{ref}}/W_F)]$ [13], [16]. However, the assumption of $\Delta W = 0$ leads to certain deviation, particularly large for pMOS. Again, this verification justifies

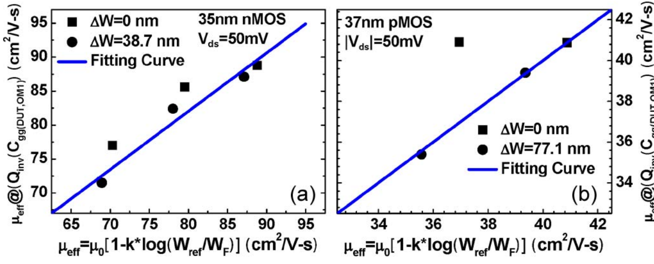

Fig. 5. Extracted $\mu_{\rm eff}$ with different ΔW versus calculated $\mu_{\rm eff}$ with width-dependent stress effect. (a) nMOS: $\Delta W = 0$ and 38.7 nm, and $k = 0.3465$. (b) pMOS $\Delta W = 0$ and 77.1 nm, and $k = 0.1598$.

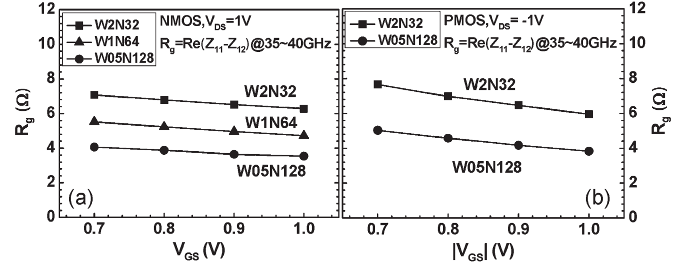

Fig. 6. R_g-versus-$V_{\rm GS}$ extracted from $\mathrm{Re}(Z_{11} - Z_{12})$ at very high frequency for multifinger devices. (a) nMOS (W2N32, W1N64, and W05N128) and (b) pMOS (W2N32 and W05N128).

the accuracy of extracted ΔW. The difference of ΔW between nMOS and pMOS suggests that STI TCR is dependent not only on STI etching profile and postetching thermal budget but also on implantation and postannealing process. The differences in gate oxide thinning and gate depletion near the STI top corner, between nMOS and pMOS, may be one more reason responsible for the ΔW difference. Note that this difference is determined not only by STI process but also by the gate stack process

$$\mu_{\rm eff} = \frac{I_{\rm DS}}{V_{\rm DS}} \frac{1}{C_{\rm ox(inv)}(V_{\rm GS} - V_T - \lambda V_{\rm DS}) W_{\rm eff}/L_{\rm eff}} \quad (9)$$

$$V_{\rm DS} = V_{\rm DS,ext} - I_{\rm DS}(R_D + R_S). \quad (10)$$

D. Gate Resistance Extraction and Layout Dependence in Multifinger MOSFETs

Gate resistance R_g has been recognized as the most critical parameter determining $f_{\rm MAX}$ and RF noise parameters, such as $NF_{\rm min}$ and R_n [8]–[10]. However, an accurate extraction and modeling of R_g remain a challenging subject. Most of the previous works relied on curve fitting to the measured input impedance for R_g extraction. As a result, it is difficult to figure out a scalable R_g model, which can predict layout and bias dependence. The mentioned problem adds the challenge to high-frequency simulation accuracy, particularly for the design with special concern of $f_{\rm MAX}$ and RF noise [17]. First, an analytical expression with Z-parameters, expressed by (11)–(13), was proposed for R_g extraction [18]. This Z-method requiring measurement up to infinitely high frequency faces the limitation of existing vector network analyzer. An alternative approach by curve fitting was proposed based on the assumption that the frequency-dependent term in (11) with A_g and B given by (12) and (13) becomes negligible at sufficiently high frequency. Afterward, another R_g extraction method using Y-parameters was presented and given by a simple formula (14) [19]. This Y-method looks much simpler and may avoid the difficulty in measurement to infinite frequency. Both Z- and Y-methods were applied to our multifinger devices to verify the accuracy and limitations in R_g extraction. The results indicate that R_g extracted by these two methods reveals a dramatic difference at lower frequency but tends to converge to nearly the same value at very high frequency, up to 35–40 GHz, expressed by (15). However, this convergence is achievable only for W2N32 (standard) and becomes invalid for narrow-OD devices, such as W1N64 and W05N128. The major problem happens when trying to apply the Y-method to narrow-OD devices, in which the extracted R_g is underestimated due to overestimated $\mathrm{Im}(Y_{11})$, by including $C_{\rm of}$ and $C_{f(\rm poly\ end)}$. Note that both Z- and Y-parameters went through openM1 and shortM1 deembedding, but the openM1 cannot remove $C_{\rm of}$ and $C_{f(\rm poly\ end)}$. The impact on $\mathrm{Im}(Y_{11})$ increases with increasing N_F and leads to underestimated R_g. The mentioned verification suggests that Y-method cannot be applied to narrow-OD devices with very narrow W_F and very larger N_F

$$R_g = \mathrm{Re}(Z_{11} - Z_{12}) - \frac{A_g}{\omega^2 + B} \quad (11)$$

$$A_g = \frac{C_{\rm ds}\left[g_{\rm ds}(C_{\rm gs} + C_{\rm gd}) + g_m C_{\rm gd}\right]}{(C_{\rm gs}C_{\rm ds} + C_{\rm gs}C_{\rm gd} + C_{\rm gd}C_{\rm ds})^2} - \frac{g_{\rm ds}}{(C_{\rm gs}C_{\rm ds} + C_{\rm gs}C_{\rm gd} + C_{\rm gd}C_{\rm ds})} \quad (12)$$

$$B = \left[\frac{g_{\rm ds}(C_{\rm gs} + C_{\rm gd}) + g_m C_{\rm gd}}{(C_{\rm gs}C_{\rm ds} + C_{\rm gs}C_{\rm gd} + C_{\rm gd}C_{\rm ds})}\right]^2 \quad (13)$$

$$R_g = \frac{\mathrm{Re}(Y_{11})}{[\mathrm{Im}(Y_{11})]^2}. \quad (14)$$

For W2N32 at very high frequency up to 35–40 GHz

$$R_g \cong \mathrm{Re}(Z_{11} - Z_{12})|_{35-40\ \rm GHz} \cong \frac{\mathrm{Re}(Y_{11})}{[\mathrm{Im}(Y_{11})]^2}|_{35-40\ \rm GHz}. \quad (15)$$

According to the aforementioned verification and identified mechanism, Z-method was adopted as a reliable solution. Fig. 6(a) and (b) shows R_g extracted from nMOS and pMOS, respectively. The narrow-OD devices can achieve smaller R_g, i.e., the expected benefit from smaller W_F and larger N_F. nMOS and pMOS indicate similar result in layout and $V_{\rm GS}$ dependence. Fig. 7 shows that the extracted R_g versus W_F can be approached by a simple analytical model given by (16), in which the first term from poly sheet resistance $(R_{\rm SH(poly)})$ is proportional to W_F/N_F and the second term from the contact resistance of M1 to poly gate $(R_{\rm CT(poly)})$ is proportional to $1/N_F$. The good match between the extracted and calculated R_g suggests that (16) can serve as a scalable model to predict the layout dependence of R_g in multifinger devices with various

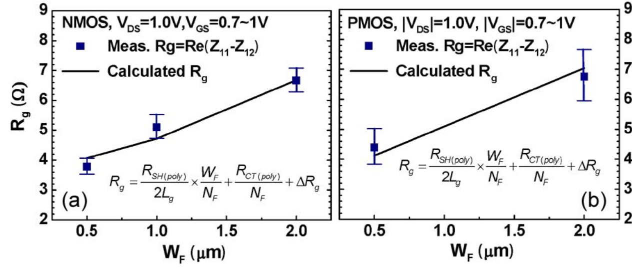

Fig. 7. Extracted and calculated R_g-versus-W_F for multifinger devices. (a) nMOS (W2N32, W1N64, and W05N128) and (b) pMOS (W2N32 and W05N128).

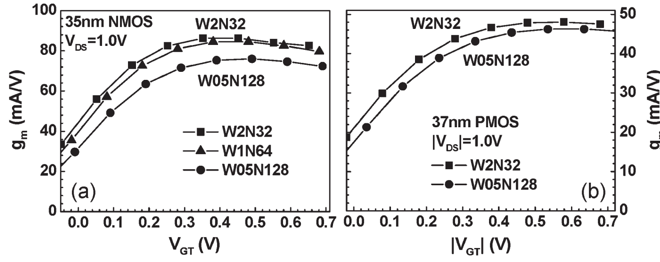

Fig. 8. $g_{m,\mathrm{sat}}$-versus-V_{GT} in saturation region, measured from (a) nMOS with W2N32, W1N64, and W05N128 and (b) pMOS with W2N32 and W1N64. $|V_{\mathrm{DS}}| = 1.0$ V and $V_{\mathrm{GT}} = V_{\mathrm{GS}} - V_T$.

W_F and N_F. In the following section, the influence of R_g on f_{MAX} and RF noise parameters like R_n and NF_{min} will be investigated

$$R_g = \frac{R_{\mathrm{SH(poly)}}}{2L_g} \times \frac{W_F}{N_F} + \frac{R_{\mathrm{CT(poly)}}}{N_F} + \Delta R_g. \tag{16}$$

IV. Narrow-Width Effect on High-Frequency Performance and RF Noise

The influence on μ_{eff}, g_m, C_{gg}, and R_g from narrow-width effects (STI $\sigma_\perp$, TCR-induced ΔW, C_{of}, and $C_{f(\mathrm{poly\ end})}$) suggests an extended impact on f_T, f_{MAX}, and NF_{min}, which are key performance parameters for RF circuit design. The experimental results and detailed analysis will be described as follows.

A. Narrow-Width Effect on f_T in Multifinger nMOS and pMOS

First, Fig. 8(a) shows $g_{m,\mathrm{sat}}$-versus-V_{GT} in saturation region ($V_{\mathrm{DS}} = 1.0$ V) measured from nMOS with standard and narrow-OD layouts. The result reveals a monotonic degradation of $g_{m,\mathrm{sat}}$ with W_F scaling, which follows the same trend as that of linear g_m [Fig. 2(a)] and reflects the extended impact from STI $\sigma_\perp$ to saturation velocity. The $g_{m,\mathrm{sat}}$ degradation reaches 12.1% for W05N128 compared to W2N32 (standard). As for pMOS shown in Fig. 8(b), the $g_{m,\mathrm{sat}}$ degradation of W05N128 becomes much smaller to around 3%. Again, it indicates that $g_{m,\mathrm{sat}}$ degradation from compressive $\sigma_\perp$ can be compensated by much larger ΔW (77.1 nm) in pMOS, which results in a

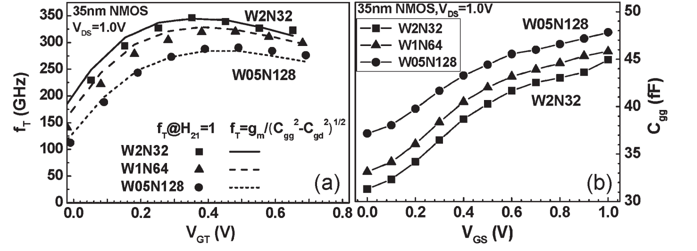

Fig. 9. Multifinger nMOS with W2N32, W1N64, and W05N128. (a) Measured and calculated f_T-versus-V_{GT} ($V_{\mathrm{DS}} = 1.0$ V) and (b) C_{gg}-versus-V_{GS} extracted from $\mathrm{Im}(Y_{11})$.

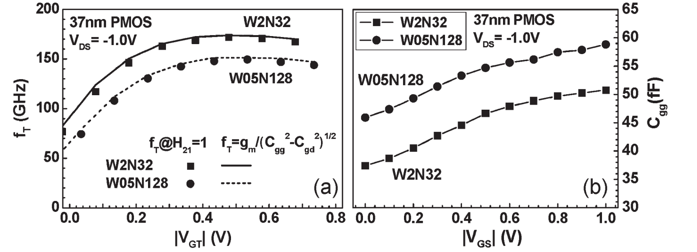

Fig. 10. Multifinger pMOS with W2N32 and W05N128. (a) Measured and calculated f_T-versus-$|V_{\mathrm{GT}}|$ ($|V_{\mathrm{DS}}| = 1.0$ V) and (b) C_{gg}-versus-$|V_{\mathrm{GS}}|$ extracted from $\mathrm{Im}(Y_{11})$.

significant increase of W_{eff} and then certain compensation to $g_{m,\mathrm{sat}}$.

Fig. 9(a) shows the measured and calculated f_T-versus-V_{GT} for standard and narrow-OD nMOS devices. Note that f_T is determined by the extrapolation of $|H_{21}|$ to unity gain and H-parameters can be converted from S-parameters after openM1 and shortM1 deembedding. The experimental result indicates that f_T can approach 350 GHz for the standard nMOS (W2N32), attributed to the aggressive L_g scaling to 35 nm. However, the W_F scaling leads to a monotonic degradation of f_T. As compared to W2N32, the maximum f_T of W1N64 is reduced by 5.8%, and the degradation becomes even larger to 15.1% for W05N128, resulting in f_T below 290 GHz. Referring to the analytical model given by (17) and (18) for calculating f_T [8], it is predicted that f_T degradation is originated from the degradation of $g_{m,\mathrm{sat}}$ and/or the increase of C_{gg}. For narrow-OD nMOS, the smallest $g_{m,\mathrm{sat}}$ appearing in W05N128 [Fig. 8(a)] suggests to be one of the factors responsible for the lowest f_T. Furthermore, C_{gg} measured from narrow-OD nMOS shown in Fig. 9(b) indicates 8.3% larger C_{gg} in W05N128. Note that the increase of C_{gg} with smaller W_F and larger N_F reveals the impact from $C_{f(\mathrm{poly\ end})}$ [6], [7]. The combined effect from lower $g_{m,\mathrm{sat}}$ and larger C_{gg} can explain f_T degradation in narrow-OD devices. A good match between the measured and calculated f_T shown in Fig. 9(a) for nMOS with various W_F and N_F justifies the accuracy of the proposed f_T model. As for pMOS shown in Fig. 10(a), W05N128 (narrow OD) reveals more than 15% degradation in maximum f_T than W2N32 (standard), even though the extraordinarily large ΔW (77.1 nm) can effectively suppress $g_{m,\mathrm{sat}}$ degradation [Fig. 8(b)]. Again, the result can be ascribed to the increase of C_{gg} by around 15.2% shown in Fig. 9(b). Also, the proposed f_T model can

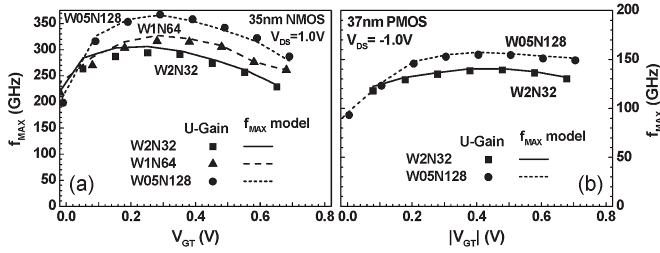

Fig. 11. Measured and calculated $f_{\rm MAX}$-versus-$|V_{\rm GT}|$. (a) nMOS with W2N32, W1N64, and W05N128 and (b) pMOS with W2N32 and W1N64. $|V_{\rm DS}| = 1.0$ V and $V_{\rm GT} = V_{\rm GS} - V_T$.

accurately predict f_T degradation suffered by the narrow-OD pMOS

$$f_T = \frac{g_m}{2\pi\sqrt{C_{\rm gg}^2 - C_{\rm gd}^2}} \tag{17}$$

$$C_{\rm gg} = \frac{{\rm Im}(Y_{11})}{\omega} \quad C_{\rm gd} = -\frac{{\rm Im}(Y_{12})}{\omega}. \tag{18}$$

B. Narrow-Width Effect on $f_{\rm MAX}$ in Multifinger nMOS and pMOS

The maximum oscillation frequency $f_{\rm MAX}$ is another important performance parameter for RF circuit design, particularly for power amplifiers. In this paper, $f_{\rm MAX}$ is determined by conventionally used unilateral gain (U) method in which the frequency corresponding to the unit power gain is defined as $f_{\rm MAX}$, i.e., $f(U = 1) = f_{\rm MAX}$. Fig. 11(a) shows $f_{\rm MAX}$ extracted from nMOS and reveals an interesting result that the narrow-OD layout can lead to higher $f_{\rm MAX}$ and the maximum $f_{\rm MAX}$ of W05N128 can achieve 366 GHz, which is 24.8% improvement over W2N32. Through an equivalent circuit analysis on unilateral gain (U), $f_{\rm MAX}$ can be calculated by (19) in which $g_{\rm ds}$ is the output conductance, $C_{\rm gd}$ is the gate–drain capacitances, R_i is the real part of the input impedance, and R_s is the source series resistance [9], [10]. This model predicts that the higher f_T and lower R_g can enhance $f_{\rm MAX}$. Referring to Fig. 9(a), W05N128 suffers the lowest f_T and more than 15% degradation in the maximum $f_{\rm T}$ compared to W2N32. However, the smaller W_F and larger N_F in narrow-OD nMOS can effectively reduce R_g as shown in Fig. 6(a). The 40% lower R_g realized in W05N128 can overcompensate f_T degradation and contribute higher $f_{\rm MAX}$. A good agreement between the measured and calculated $f_{\rm MAX}$ shown in Fig. 11(a) justifies the accuracy of the proposed $f_{\rm MAX}$ model in (19) [9], [10]. Fig. 11(b) shows the $f_{\rm MAX}$ measured from pMOS, which indicates W_F and N_F effect similar to NMOS, i.e., the narrower W_F and larger N_F can yield higher $f_{\rm MAX}$. The narrow-OD pMOS W05N128 can offer 13.5% improvement in the maximum $f_{\rm MAX}$ than W2N32 even though it suffers 15% lower f_T [Fig. 10(a)]. Again, Fig. 6(b) shows around 30% lower R_g in W05N128 than in W2N32 and explains the origin responsible for $f_{\rm MAX}$ enhancement. The proposed $f_{\rm MAX}$ model showing a good agreement with measured $f_{\rm MAX}$ can consistently predict the layout dependence and guide device optimization design

$$f_{\rm MAX} = \frac{f_T}{2\sqrt{R_g(g_{\rm ds} + 2\pi f_T C_{\rm gd}) + g_{\rm ds}(R_i + R_s)}}. \tag{19}$$

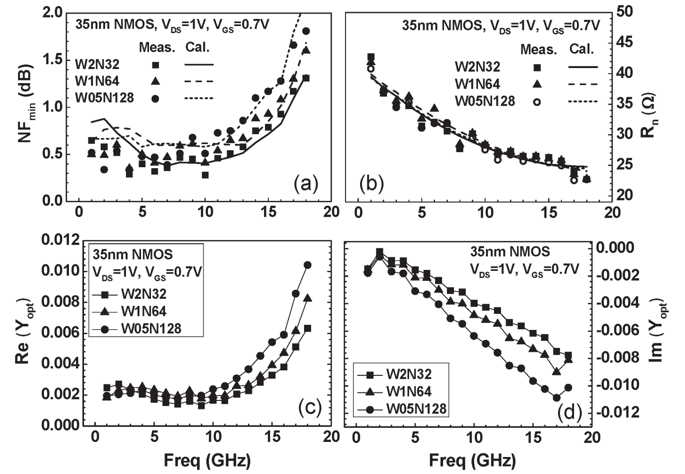

Fig. 12. Noise parameters of nMOS (standard: W2N32; narrow OD: W1N64 and W05N128). (a) Measured and calculated $NF_{\rm min}$, (b) measured and calculated R_n, (c) ${\rm Re}(Y_{\rm opt})$, and (d) ${\rm Im}(Y_{\rm opt})$ measured under $V_{\rm GS} = 0.7$ V and $V_{\rm DS} = 1.0$ V.

C. Narrow-Width Effect on RF Noise in Multifinger nMOS and pMOS

The proliferated impact from layout-dependent stress, parasitic capacitances, and, most importantly, R_g on high-frequency noise parameters appears as one more key topic for RF circuit design, particularly for low-noise amplifiers using nanoscale CMOS devices. Note that high-frequency noise measurement was carried out using ATN-NP5B system with NP5 controller, a noise figure meter (HP8970B), a remote receiver module, a mismatch noise source, a noise source, a network analyzer (HP 8510), and a dc power supply (HP4142).

Fig. 12(a)–(d) shows four noise parameters, such as $NF_{\rm min}$, R_n, ${\rm Re}(Y_{\rm opt})$, and ${\rm Im}(Y_{\rm opt})$ measured from nMOS under $V_{\rm DS} = 1.0$ V and $V_{\rm GS} = 0.7$ V corresponding to maximum $g_{\rm m,sat}$. Note that $NF_{\rm min}$ is the minimum noise figure, R_n is the equivalent noise resistance, and $Y_{\rm opt}$ is the optimum source admittance achieving $NF_{\rm min}$. Unfortunately, W05N128 with the smallest W_F and R_g suffers 0.2–0.5 dB higher $NF_{\rm min}$ in 9-18 GHz. R_n shown in Fig. 12(b) indicates very minor difference among three layouts. However, ${\rm Re}(Y_{\rm opt})$ and ${\rm Im}(Y_{\rm opt})$ shown in Fig. 12(c) and (d) reveal significant increase (absolute value) in W05N128 at frequency above 9 GHz. An analytical model for noise parameters derived from noisy two-port network given by (20)–(22) [20] is employed to explain the layout-dependent effect on $NF_{\rm min}$ and R_n. According to (20) and (21), the increase of either R_n or ${\rm Re}(Y_{\rm opt})$ will lead to higher $NF_{\rm min}$ and the calculated $NF_{\rm min}$ match with measured data in terms of frequency and layout dependence, i.e., $W05N128 > W1N64 > W2N32$, as shown in Fig. 12(a). This proven model combined with Fig. 12(b) and (c) for measured R_n and ${\rm Re}(Y_{\rm opt})$ indicates that the increase of ${\rm Re}(Y_{\rm opt})$ is the primary factor responsible for higher $NF_{\rm min}$ in narrow-OD nMOS. The aforementioned analysis is applied to measured noise parameters before deembedding, and the increase of ${\rm Re}(Y_{\rm opt})$ and $|{\rm Im}(Y_{\rm opt})|$ in narrow-OD device is originated from larger $C_{\rm gg}$ due to more parasitic capacitances. Regarding the W_F scaling effect on R_n, further analysis on three key elements given by (22), shown in Fig. 13(a), reveals that the benefit of smaller

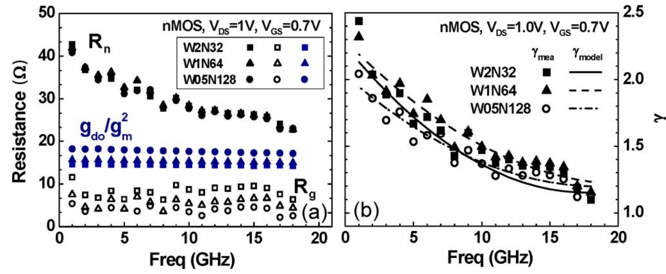

Fig. 13. Multifinger nMOS with W2N32, W1N64, and W05N128. (a) Measured R_n, R_g, and $g_{\rm do}/g_m^2$ versus frequency. (b) Measured $\gamma_{\rm mea}$ and $\gamma_{\rm model}$ for best fitting to frequency dependence of $\gamma_{\rm mea}$. $V_{\rm DS} = 1.0$ V and $V_{\rm GS} = 0.7$ V.

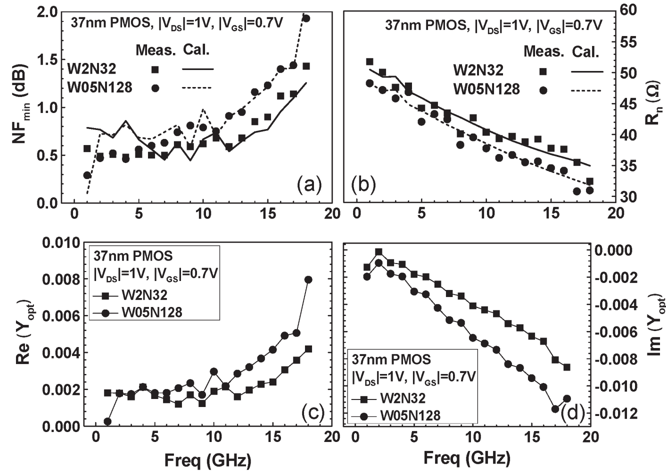

Fig. 14. Noise parameters of pMOS (standard: W2N32; narrow OD: W05N128). (a) Measured and calculated $NF_{\rm min}$, (b) measured and calculated R_n, (c) $\mathrm{Re}(Y_{\rm opt})$, and (d) $\mathrm{Im}(Y_{\rm opt})$ measured under $V_{\rm GS} = -0.7$ V and $V_{\rm DS} = -1.0$ V.

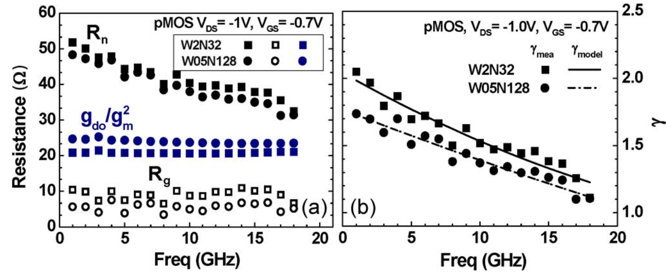

Fig. 15. Multifinger pMOS with W2N32 and W05N128. (a) Measured R_n, R_g, and $g_{\rm do}/g_m^2$ versus frequency. (b) Measured $\gamma_{\rm mea}$ and $\gamma_{\rm model}$ for best fitting to frequency dependence of $\gamma_{\rm mea}$. $V_{\rm DS} = -1.0$ V and $V_{\rm GS} = -0.7$ V.

R_g in narrow-OD devices happens to be cancelled out by the increase of the second term, i.e., $\gamma \cdot g_{\rm do}/g_m^2$, due to lower g_m from compressive $\sigma_\perp$ and $\gamma > 1$ shown in Fig. 13(b). Note that γ represents excess noise factor and $\gamma > 1$ comes from short-channel effect. The increase of γ at lower frequency can be ascribed to significant substrate potential variation [21]. Fig. 14(a)–(d) shows four noise parameters of pMOS with lower R_n but significantly higher $\mathrm{Re}(Y_{\rm opt})$ for W05N128. The lower R_n is attributed to smaller R_g [Fig. 6(b)] and lower γ. However, the higher $\mathrm{Re}(Y_{\rm opt})$ still plays as the dominant factor and leads to 0.2–0.5 dB higher $NF_{\rm min}$ than W2N32. Again, the proposed model (20)–(22) and R_n analysis made in Fig. 15 can consistently predict the measured $NF_{\rm min}$ and R_n and explain the narrow-width effect

$$F_{\rm min} = 1 + 2R_n\mathrm{Re}(Y_{\rm opt})\left[1 + R_n\mathrm{Re}(Y_{\rm opt})\right] \quad (20)$$

$$NF_{\rm min} = 10 \cdot \log F_{\rm min} \quad (21)$$

$$R_n \approx R_g + \gamma\frac{g_{\rm do}}{g_m^2}(\gamma > 1 \text{ for short channel}). \quad (22)$$

V. Conclusion

Narrow-OD MOSFET W05N128 with four times smaller W_F/larger N_F than the standard multifinger device W2N32 can achieve 40%/30% lower R_g and 24.8%/13.5% higher $f_{\rm MAX}$ for nMOSFET/pMOSFET. The maximum $f_{\rm MAX}$ can reach 366 GHz for 35-nm nMOS and 155 GHz for 37-nm pMOS. However, these narrow-OD devices, even with the advantage of lower R_g, suffer lower f_T and higher $NF_{\rm min}$. The g_m degradation caused by STI compressive $\sigma_\perp$ and the increase of $C_{\rm gg}$ due to finger-end fringing capacitance are identified as two layout-dependent factors responsible for f_T degradation in narrow-OD devices. The increase of measured $NF_{\rm min}$ before deembedding can be ascribed to the increase of $\mathrm{Re}(Y_{\rm opt})$ also due to larger $C_{\rm gg}$. The lower R_g cannot guarantee lower R_n due to a competing factor from g_m degradation. The narrow-width effects on high-frequency performance and RF noise and the mechanism underlying the tradeoff between different parameters provide an important guideline of multifinger MOSFET layout for RF circuit design using nanoscale CMOS technology.

Acknowledgment

The authors would like to thank the support from the National Device Laboratory for noise measurement and the Chip Implementation Center for test key tape-out and device fabrication.

References

[1] L. F. Tiemeijer, R. J. Havens, R. de Kort, A. J. Scholten, R. van Langevelde, D. B. M. Klaassen, G. T. Sasse, Y. Bouttement, C. Petot, S. Bardy, D. Gloria, P. Scheer, S. Boret, B. Van Haaren, C. Clement, J.-F. Larchanche, I.-S. Lim, A. Duvallet, and A. Zlotnicka, “Record RF performance of standard 90 nm CMOS technology,” in *IEDM Tech. Dig.*, 2004, p. 441-444.

[2] I. Post, M. Akbar, G. Curello, S. Gannavaram, W. Hafez, U. Jalan, K. Komeyii, J. Lin, N. Lindert, J. Park, J. Rizk, G. Sacks, C. Tsai, D. Yeh, P. Bai, and C.-H. Jan, “A 65 nm CMOS SOC technology featuring strained silicon transistors for RF applications,” in *Proc. IEDM*, 2006, pp. 1013–1015.

[3] H. Li, B. Jagannathan, J. Wang, T.-C. Su, S. Sweeney, J. J. Pekarik, Y. Shi, D. Greenberg, Z. Jing, R. Groves, L. Wagner, and S. Csutak, “Technology scaling and device design for 350 GHz RF performance in 45 nm bulk CMOS process,” in *Proc. VLSI Symp. Technol.*, 2007, pp. 56–57.

[4] P. VanDerVoorn, M. Agostinelli, S. Choi, G. Curello, H. Deshpande, M. A. El-Tanani, W. Hafez, U. Jalan, L. Janbay, M. Kang, K. Koh, K. Komeyli, H. Lakdawala, J. Lin, N. Lindert, S. Mudanai, J. Park, K. Phoa, A. Rahman, J. Rizk, L. Rockford, G. Sacks, K. Soumyanath, H. Tashiro, S. Taylor, C. Tsai, H. Xu, J. Xu, L. Yang, I. Young, J. Yeh, J. Yip, P. Bai, and C. Jan, “A 32 nm low power RF CMOS SOC technology featuring high-k/metal gate,” in *Proc. VLSI Symp. Technol.*, 2010, pp. 137–138.

[5] B. Heydari, M. Bohsali, E. Adadi, and A. M. Niknejad, “Millimeter-wave devices and circuit blocks up to 104 GHz in 90 nm CMOS,” *IEEE J. Solid State Circuits*, vol. 42, no. 12, pp. 2893–2903, Dec. 2007.

[6] 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.

[7] K.-L. Yeh and J.-C. Guo, “A new method for layout-dependent parasitic capacitance analysis and effective mobility extraction in nanoscale multifinger MOSFETs,” *IEEE Trans. Electron Devices*, vol. 58, no. 9, pp. 2838–2846, Sep. 2011.

[8] T. Manku, “Microwave CMOS—Device physics and design,” *IEEE J. Solid-State Circuits*, vol. 34, no. 3, pp. 277–285, Mar. 1999.

[9] P. H. Woerlee, M. J. Knitel, R. V. Langevelde, D. B. M. Klaassen, L. F. Tiemeijer, A. J. Scholten, and A. T. A. Z. Duijnhoven, “RF CMOS performance trends,” *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1776–1782, Aug. 2001.

[10] J. N. Burghartz, “Tailoring logic CMOS for RF applications,” in *IEDM Tech. Dig.*, 2001, pp. 150–153.

[11] J. C. Guo, “Low-K/Cu CMOS-based SoC technology with 115 GHz f_{T}, 100 GHz $f_{\max}$, low noise 80-nm RF CMOS, high-Q MiM capacitor and spiral Cu inductor,” *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 3, pp. 331–338, Aug. 2006.

[12] Y. Luo and D. K. Nayak, “Enhancement of CMOS performance by process-induced stress,” *IEEE Trans. Semicond. Manuf.*, vol. 18, no. 1, pp. 63–68, Feb. 2005.

[13] 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.

[14] 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.

[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] R. Li, L. Yu, H. Xin, Y. Dong, K. Tao, and C. Wang, “A comprehensive study of reducing the STI mechanical stress effect on channel-width-dependent I_{dsat},” *Semicond. Sci. Technol.*, vol. 22, no. 12, pp. 1292–1297, Dec. 2007.

[17] X. Jin, J.-J. Ou, C. H. Chen, W. Liu, J. Deen, P. R. Gray, and C. Hu, “An effective gate resistance model for CMOS RF and noise modeling,” in *IEDM Tech. Dig.*, 1998, pp. 961–964.

[18] S. Lee, H. K. Yu, C. S. Kim, J. G. Koo, and K. S. Nam, “A novel approach to extracting small-signal model parameters of silicon MOSFETs,” *IEEE Microw. Guided Wave Lett.*, vol. 7, no. 3, pp. 75–77, May 1997.

[19] I. Kwon, M. Je, K. Lee, and H. Shin, “A simple and analytical parameter-extraction method of a microwave MOSFET,” *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 6, pp. 1503–1509, Jun. 2002.

[20] H. Hillbrand and P. H. Russer, “An efficient method for computer aided noise analysis of linear networks,” *IEEE Trans. Circuits Syst.*, vol. 23, no. 4, pp. 235–238, Apr. 1976.

[21] J.-C. Guo and Y.-M. Lin, “A compact RF CMOS modeling for accurate high frequency noise simulation in sub-100-nm MOSFETs,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 27, no. 9, pp. 1684–1688, Sep. 2008.

Kuo-Liang Yeh (M’09) is currently working toward the Ph.D. degree in electronics engineering at National Chiao Tung University, Hsinchu, Taiwan.

He is currently a Senior Manager with Silicon Motion Technology Corporation, Hsinchu.

Jyh-Chyurn Guo (M’06–SM’07) received the Ph.D. degree in electronics engineering from the National Chiao-Tung University (NCTU), Taiwan.

She is currently a Professor in the department of electronics engineering, NCTU. She has published more than 70 technical papers and is the holder of 19 U.S. patents