The Trade-off between STI Stress and Gate Resistance in RF MOSFETs Design for High Frequency Performance and RF Noise

Chih-You Ku, Kuo-Ling Yeh, and *Jyh-Chyurn Guo
Institute of Electronic Engineering, National Chiao Tung University, Hsinchu, Taiwan
Tel: +886-3-5131368, Fax: +886-3-5724361, E-mail: jeguo@mail.nctu.edu.tw

Abstract

STI stress induced mobility and transconductance ($g_m$) degradation appear as a penalty of multi-finger devices for RF and analog design. Donut device layout is proposed to eliminate the STI transverse stress and achieve higher $g_m$. Both NMOS and PMOS can benefit from the donut layout, with higher cut-off frequency ($f_T$). However, the trade-off between $g_m$ and gate resistance ($R_g$) may impose undesired impact on high frequency performance other than $f_T$, such as maximum oscillation frequency ($f_{om}$) and RF noise. In this paper, a comparison between multi-finger and donut MOSFETs in terms of $f_T$, $f_{max}$, and $NF_{min}$ can provide a useful guideline of device layout for RF design using nanoscale CMOS technology.

I. Introduction

Multi-finger layouts with smaller finger width ($W_f$) and larger finger number ($N_f$) have been extensively used in RF devices to reduce $R_g$ and improve $f_{max}$ as well as RF noise [1-3]. Unfortunately, the multi-finger layout introduces undesired penalty of lower $g_m$ and larger parasitic capacitances. The former one comes from mobility degradation due to STI stress and the latter one stems from finger-end fringing capacitances [4]. Both factors become worse with $W_f$ scaling and the resulted impact on $f_T$ may offset the benefit in $f_{max}$ from lower $R_g$. A ring type device was proposed, trying to minimize the STI transverse stress ($\sigma_t$) and verify the influence on flicker noise [5]. However, the impact on high frequency performance and RF noise is unknown. In our previous work, donut devices were created to eliminate $\sigma_t$ and proven effective to increase effective mobility ($\mu_{ef}$) and reduce flicker noise for both NMOS and PMOS [6]. In this paper, further investigation is performed on multi-finger and donut MOSFETs on the same chip to look into the trade-off between STI stress and $R_g$ in different layouts and the impact on $f_T$, $f_{max}$, and RF noise.

II. Experimental

90nm RF CMOS process was employed for device fabrication. Fig.1(a)-(c) illustrate the layout for a multi-finger device with $W_{fd}=2\mu m x 32$ (W2N32) and two donut devices with $W_{ind}=16\mu m x 4$ but different gate-to-STI edge spaces, 0.3μm (min. rule) and 3μm, namely D1S1 and D10S10. Note that both D1S1 and D10S10 are free from $\sigma_t$ but with maximum and relaxed $\sigma_o$, respectively. S-parameters were measured by Agilent network analyzer E8364B up to 40GHz. Open and short deembedding to the bottom metal, i.e. M1, were performed to remove the parasitic capacitances from the pads as well as interconnection lines [4]. RF noise was measured by using ATN-NP5B from 1GHz to 18GHz.

III. Results and Discussion

A comparison of $g_m$ vs. $V_{GS}$ between multi-finger (W2N32) and donut (D1S1, D10S10) NMOS shown in Fig.2(a) indicates that $g_{m,max}$ of D10S10 is enhanced by 7.5% but that of D1S1 is degraded by 9.7% compared to W2N32. It suggests that the compressive $\sigma_o$ from STI, maximized in D1S1 is the key factor responsible for $g_m$ degradation. As for D10S10, the much lower $\sigma_o$ due to 10 times larger space and eliminated $\sigma_t$ from donut layout contributes to the $g_m$ improvement. $f_T$ vs. $V_{GS}$ shown in Fig.2(b) indicates a consistent result that D10S10 with the highest $g_m$ can offer the best record of $f_T$. Note that $f_T$ is determined by the unit current gain, i.e. $f_T=\pi|H_2|^{1/2}$. Also, $f_T$ can be calculated by an analytical model of (1) [7], which predicts a linear increase of $f_T$ from $g_m$ enhancement, considering nearly the same $C_g$ and $C_m$ under fixed $W_{ind}$. As for PMOS shown in Fig.3(a), D1S1 and D10S10 demonstrate 12.2% and 7.6% higher $g_{m,max}$ than W2N32. Referring to [6], D1S1 free from $\sigma_t$, but with max. $\sigma_o$, at the gate to STI-edge space can benefit the most in hole mobility. The $f_T$ improvement from donut layout becomes particularly larger for PMOS. As shown in Fig.3(b), D1S1 can offer the highest $f_T$, which is around 28% enhancement on the max. $f_T$ compared to W2N32.

Besides $f_T$, $f_{max}$ and minimum noise figure ($NF_{min}$) are two more important high frequency parameters for RF circuits like PA and LNA. Fig.4(a) and (b) present $f_{max}$ vs. $|V_{GS}|$ measured from NMOS and PMOS, respectively. Note that $f_{max}$ is determined by the unilateral gain (U) method defined by $f_{max}=r(U=1)$. Unfortunately, the donut devices reveal around 4-6 times lower $f_{max}$ and the degradation can be up to beyond 6 times for D10S10, even though it can offer around 11% higher $f_T$ than W2N32 for PMOS. To explore the mechanism underlying $f_{max}$ degradation revealed in donut devices even with higher $f_T$, an analytical model for $f_{max}$ given by (2) suggests that increase of $R_g$ may be a key factor responsible for the lower $f_{max}$ [7]. Then, an accurate $R_g$ extraction method becomes the prerequisite. Through a small signal equivalent circuit analysis, $R_g$ can be determined by improved $Z$-method written in (3) and $Y$-method given by (4). Note that the second term of (3), i.e. $A_4/(\omega^2+\beta^2)$ makes major difference from a simple $Z$-method and improve the accuracy of extracted $R_g$ in both magnitude and frequency dependence. Taking D10S10 NMOS as an example, Fig.5(a) shows that both $Re(Z(12))$ and $A_4/(\omega^2+\beta^2)$ are a decreasing function of frequency and $R_g$ calculated by (3), shown in Fig.5(b) can be approximated by $Re(Z(12))$. Then, an accurate $R_g$ extraction method becomes the prerequisite. Through a small signal equivalent circuit analysis, $R_g$ can be determined by improved $Z$-method written in (3) and $Y$-method given by (4). Note that the second term of (3), i.e. $A_4/(\omega^2+\beta^2)$ makes major difference from a simple $Z$-method and improve the accuracy of extracted $R_g$ in both magnitude and frequency dependence. Taking D10S10 NMOS as an example, Fig.5(a) shows that the comparison between improved $Z$-method and $Y$-method, Fig.5(a) and (b) presents the extracted $R_g$ from W2N32, D1S1, and D10S10 (NMOS) using improved $Z$-method and $Y$-method, respectively. Both methods achieve consistent results and reveal much larger $R_g$ extracted from D10S10 and D1S1, which are 147Ω and 242Ω, i.e. 17 and 27 times larger than that of W2N32. This analysis proves that the extraordinarily low $f_{max}$ suffered by donut devices comes from the large $R_g$, according to (2).

As it has been well know that $R_g$ plays an important role in high frequency noise and the layout dependence may impose further impact on the noise parameters. Fig.8 (a) and (b) show $NF_{min}$ measured from W2N32, D1S1, and D10S10 for NMOS and PMOS. Indeed, donut devices reveal significantly higher $NF_{min}$ compared with W2N32 and the steeper slope accelerates the increase of $NF_{min}$ with frequency. According to the noise parameters model derived from noisy two-port network, given by
(5) and (6) for \( N_{\text{min}} \) and \( R_c \), the increase of \( R_c \) introduces larger \( R_b \) and the increase of \( R_b \) leads to higher \( N_{\text{min}} \) and also steeper rise with increasing frequency. This model incorporating layout dependent \( R_b \) (Fig. 7) can explain the drastic increase of \( N_{\text{min}} \), in donut devices and that D10S10 suffers the highest \( N_{\text{min}} \) due to 27 times larger \( R_b \) than W2N32. PMOS follow the same trend but reveal higher \( N_{\text{min}} \), owing to larger \( g_m \) and larger \( R_b \) (not shown).

\[ f_f = \frac{g_m}{2 \pi \sqrt{C_{g_s} C_{gd}}}, \]

\[ f_{\text{max}} = \frac{2 \pi f_f}{g_m (g_m + 2 \pi f_f C_{gd} C_{gs})}, \]

\[ R_b = \text{Re}(Z_{in} - Z_{out} - A \cdot \omega^2 B\), \]

\[ \frac{1}{g_m} = \left[ \frac{g_m (g_m + 2 \pi f_f C_{gd} C_{gs})}{C_{gd} C_{gs} + g_m C_{gs} + g_m C_{gd}} \right] \]

\[ f_{\infty} = 1 + 2R_\infty \text{Re}(Y_{in}) (1 + R_\infty \text{Re}(Y_{out})), \]

\[ N_{\min} = 10 \log f_{\infty} \]

\[ R_c = f_f R_g \quad (\gamma > 1 \text{ for short channel devices}) \]

IV. Conclusion

Donut devices can offer higher \( g_m \) and \( f_f \) due to higher \( \mu_m \) from eliminated STI \( \sigma_f \). The gain of \( f_f \) can be up to 28% for PMOS D1S1 with zero \( \sigma_f \) and max. \( \sigma_f \). However, the donut layout leads to 17–27 times larger \( R_b \) and the penalty of lower \( f_{\text{max}} \) and higher \( N_{\text{min}} \). This critical trade-offs between \( f_f \), \( f_{\text{max}} \), and \( N_{\min} \) suggests that multi-finger layout remains the better choice for RF circuits design. Multi-ring layout with multiple donut devices in parallel may effectively reduce \( R_b \) and offer the optimized solution, yielding the highest \( f_f \) and \( f_{\text{max}} \) and minimum \( N_{\min} \).

Acknowledgement

This work is supported in part by the National Science Council under Grants NSC101-2221-E-009-115. Besides, the authors acknowledge the support from NDL for noise measurement and CiC for device fabrication.

References


Fig. 1 Layout of multi-finger and donut MOSFETs with STI \( \sigma_f \) and \( \sigma_d \). (a) multi-finger device : \( W2N32 = 2 \mu m \times 32 \) (W2N32), donut devices 16 \( \mu m \times 48\), (b) D1S1 gate-to-STI space=0.3um (c) D10S10: gate-to-STI space=3um.

Fig. 2 Comparison of multi-finger NMOS (W2N32) and donut NMOS (D1S1,D10S10) \( g_A \) vs. \( V_{GSS} \) (b) \( f_f \) vs. \( V_{CDS} \) after deembedding.

Fig. 3 Comparison of multi-finger PMOS (W2N32) and donut PMOS (D1S1, D10S10) (a) \( g_m \) vs. \( |V_{GSS}| \) (b) \( f_f \) vs. \( |V_{CDS}| \) (|V_{CDS}|=1.2V) after openMI and shortMI deembedding.

Fig. 4 Comparison of \( f_{\infty} \) vs. \( |V_{GSS}| \) between multi-finger and donut devices (D1S1, D10S10) (a) NMOS (b) PMOS, \( f_{\infty} = f(U=1), |V_{CDS}|=1.2V \).

Fig. 5 \( R_c \) extraction by improved Z-method (a) \( \text{Re}(Z_{in} - Z_{out}) \) and \( A/(\omega^2 + B) \) (b) \( R_c = \text{Re}(Z_{in} - Z_{out}) A/(\omega^2 + B), \) donut NMOS D10S10.

Fig. 6 \( R_c \) extraction by Y-method (a) \( \text{Re}(Y_{in}) \) and \( \text{Im}(Y_{in}) \) (b) \( R_c = \text{Re}(Y_{in})/|\text{Im}(Y_{in})|^2, \) donut NMOS D10S10.

Fig. 7 Comparison of multi-finger (W2N32) and donut NMOS (D1S1, D10S10) (a) \( \text{Re}(Z_{in} - Z_{out}) A/(\omega^2 + B) \) (b) \( R_c = \text{Re}(Y_{in})/|\text{Im}(Y_{in})|^2 \)

Fig. 8 Comparison of \( N_{\min} \) vs. freq. between multi-finger (W2N32) and donut devices(D1S1,D10S10)(a) NMOS (b) PMOS (|V_{CDS}|=1.2V, |V_{CDS}|=1.0)

- 69 -