Extended Abstracts of the 2019 International Conference on Solid State Devices and Materials, Nagoya, 2019, pp595-596

# FinFET Extension towards N3 and Beyond – Local Fin Trimming

T. Miyashita, S. Mittal, K. Suen, M.S. Kim, A. Pal, A. Sachid, and M. Cogorno
Applied Materials Inc., 974 East Arques Avenue, Sunnyvale, CA, USA, Email: Toshihiko_Miyashita@amat.com

***Abstract*** — Selective fin trimming in replacement metal gate (RMG) module is proposed as the local fin width scaling approach for further FinFET extension towards 3nm node and beyond. Channel fins are selectively trimmed after dummy gate stack removal in RMG module, where source/drain (S/D) region epi's and S/D extension fins are not affected by the trimming. This selective trimming approach is preferable from the view point of lower S/D parasitic resistance, while enjoying electrostatics improvement as original motivation of fin width scaling. Well-calibrated TCAD simulation clarified local fin trimming pros and cons, as well as overall 5% gate delay improvement with optimal fin trimming condition. Furthermore, this local fin trimming concept was performed by using precisely controlled selective silicon etching process for 14nm node compatible FinFETs, demonstrating more than 10% $I_{on}$-$I_{off}$ improvement with consistent electrostatics improvement. Local fin trimming is one of the promising approaches for FinFET extension towards 3nm node and beyond.

## 1. Introduction

Density-driven pitch scaling has been continuing for further CMOS scaling even after introducing FinFET technology in 22nm node [1], where the fin pitch (FP), contacted poly pitch (CPP), and minimum metal pitch (MMP) are the most critical dimensions. Figure 1 shows the FP and CPP scaling trend for different foundry technologies down to 3nm node. Because recent "technology node" is not intuitive for area scaling discussion, practical technology node metric, standard node [2], is also shown in Fig. 1. It is clearly shown that both FP and CPP are being continuously scaled, however, the scaling speed is actually slowing down, becoming larger gap between technology node and standard node. In order to further scale the CPP, gate length ($L_g$) scaling is essential, but one of the challenges of $L_g$ scaling is degraded short-channel control, and new device architectures, such as horizontal gate-all-around (hGAA), have been extensively developed [3]. However, it requires much more complicated processes than those of FinFET, continuous FinFET extension is, therefore, still the candidate in CMOS scaling due to its highly matured technology. In FinFET devices, fin profile is the most critical factor, and its key physical parameters are shown in Fig. 2, in which fin height ($H_{fin}$) and fin width ($W_{fin}$) are especially important for current drivability and electrostatics, respectively. Figure 3 shows the fin profile evolution [4] in previous FinFET technologies, and it has been migrated from shorter/wider/tapered to higher/narrower/vertical fins.

## 2. Local fin trimming approach

There are several ways to achieve narrow fin width, and fin trimming is one of the candidates. Figure 4 shows two different fin trimming approaches, namely (a) global and (b) local fin trimming [5]. As for global fin trimming, both source/drain (S/D) and channel regions are simultaneously trimmed because it is performed after fin reveal. This leads to the significant increase in S/D resistance, $R_{sd}$, although gate electrostatics are improved with narrow channel fins. On the other hand, local fin trimming is performed after dummy gate removal in replacement metal gate (RMG) module, and it enables selective channel fin trimming while maintaining the original fin width in the S/D region as shown in Fig. 5. Therefore, we propose this local fin trimming as the fin width scaling approach for next generation FinFETs.

## 3. Results and discussions

Figure 6 shows the local fin trimming impact on device DC performance. Three major impacts can be considered, where $R_{sd}$ reduction contributes to drive current increase while electrostatic degradation due to capacitance coupling and fin height reduction lead to current decrease as shown in the figure. Each factor shown in Fig. 6 was examined in the following simulations. Figure 7 shows the simulated local fin trimming impact on $I_{on}$-$I_{off}$ performance for different fin width cases. For both N and PMOS, $I_{on}$-$I_{off}$ performance is improved with fin trimming, and a narrower fin shows more performance gain. Figure 8 shows the device resistance as a function of local fin trimming. It can be seen that $R_{sd}$ decreases with fin trimming, while channel resistance, $R_{ch}$, keeps constant or rather increases due to fin height reduction. Gate-to-S/D parasitic capacitance degradation can be expected, and Figure 9 shows the simulated effective capacitance, $C_{eff}$, as a function of fin trimming. Due to increased S/D extension and epi volumes, $C_{eff}$ increase with fin trimming is clearly observed. Figure 10 shows the overall gate delay as functions of both fin width and fin trimming. It is shown that the optimal fin trimming amount is around 2nm with fin width between 5 and 6nm. If we take a look at fin width of 5nm, 2nm local fin trimming brings about 5% delay improvement.

The local fin trimming concept was demonstrated experimentally by using highly selective silicon fin etching and an internal FinFET test vehicle with 14nm node compatible design rule. As shown in Fig. 11, TEM cross-sections after local fin trimming demonstrated precise fin trimming without fin damages. The process controllability was also confirmed in Fig. 12, where the target and actual fin trimming amounts show good correlation up to 5nm fin trimming. One of the concerns of this process is fin line-edge roughness (LER) degradation, however, no fin LER degradation or rather improved LER was observed, which enables further fin width scaling for next technology node.

Electrical device characterization was also performed by implementing local fin trimming in FinFET devices. $L_g$ is 30nm and FP is 44nm, that correspond to 14nm node design rule, and highly-doped S/D epi and W-filled RMG were implemented. Local fin trimming impact on NMOS electrostatics is shown in Fig. 13. It is shown that gate control is actually improved by fin trimming, where both DIBL and sub-threshold swing are improved with narrow fin width achieved by local fin trimming. Figure 14 shows the fin trimming impact on NMOS $I_{on}$-$I_{off}$ performance. As a result of 2nm trimming, $I_{on}$-$I_{off}$ performance is improved by more than 10%, demonstrating this technology as a promising candidate for further FinFET extension.

## 4. Conclusions

Local channel fin trimming after dummy gate removal is proposed for further fin width scaling. The calibrated TCAD simulation shows that local fin trimming produces a DC performance benefit with lower $R_{sd}$. Even considering parasitic capacitance increase, overall gate delay shows the benefit in terms of strong $I_{on}$ improvement. We have also demonstrated well controlled local fin trimming process without fin damage, showing more than 10% $I_{on}$-$I_{off}$ performance improvement. This local fin trimming is promising for 3nm node to further extend FinFET technology.

## References

[1] C. Auth *et al*., Symp. VLSI Tech. Dig., p. 131, 2012. [2] G. Dicker *et al*., Proc. SPIE 9661, p. 96610F, 2015. [3] R. Ritzenthaler *et al*., IEDM Tech. Dig., p. 508, 2018. [4] K. Mistry, "10 nm Technology Leadership," Intel Technology and Manufacturing Day, 2017. [5] T. Miyashita *et al*., IEDM Tech. Dig., p. 827, 2018.

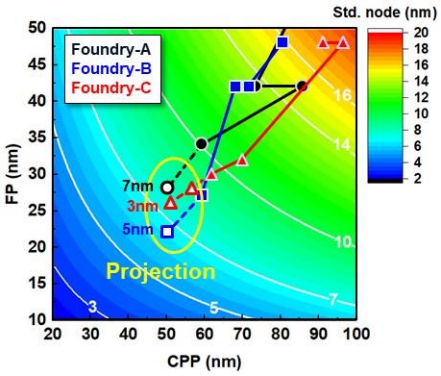


**Fig.1** Contacted poly pitch (CPP) and fin pitch (FP) scaling trend along with practical technology node metric, standard node [2], for different foundry technologies.

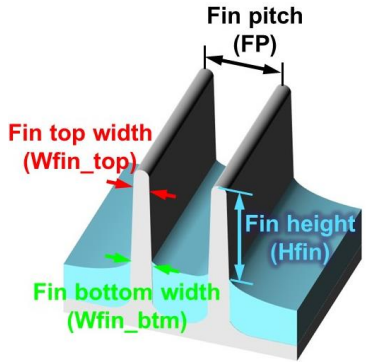


**Fig.2** Schematic diagram of channel fin profile and key physical parameters in FinFET device. The difference between fin top width and fin bottom width defines fin taper angle, both of which are correlated with on current and electrostatics, respectively.

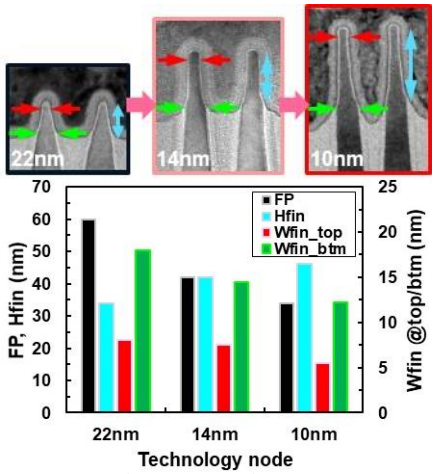


**Fig.3** Fin profile evolution from 22nm to 10nm node [4]. Higher and narrower fin profile has been introduced in advanced technology for higher current drivability and better electrostatics.

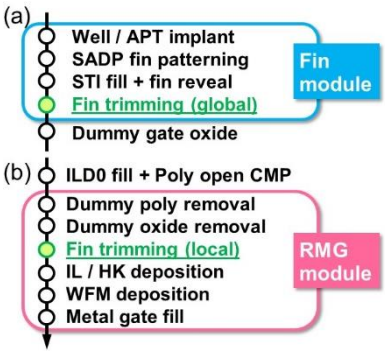


**Fig.4** Process flow for two different fin trimming approaches; (a) global trimming after fin reveal in fin module and (b) local trimming after dummy gate stack removal in RMG module.

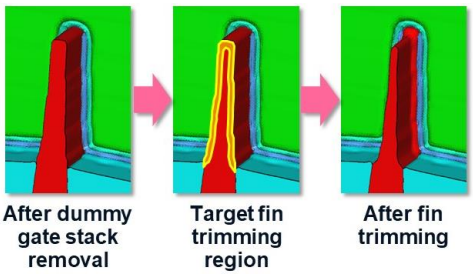


**Fig.5** Schematic concept of local fin trimming approach at RMG module. Fin channel is selectively trimmed to scale down channel fin width while keeping S/D region as original fin width for embedded S/D epi.

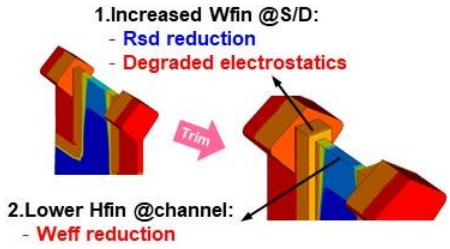


**Fig.6** Local fin trimming impact on device DC performance. Parasitic resistance reduction contributes current drive increase, while electrostatics degradation and fin height reduction lead to current drive decrease.

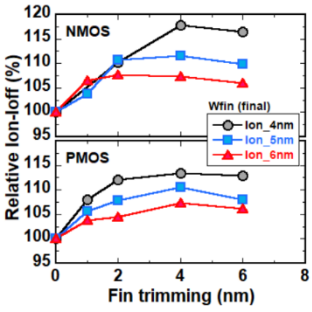


**Fig.7** TCAD simulated local fin trimming impact on $I_{on}$-$I_{off}$ performance for different final fin widths. In this figure, increase in fin trimming corresponds to increase in S/D fin width.

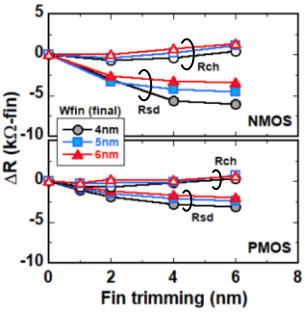


**Fig.8** TCAD simulated local fin trimming impact on device resistance. $R_{sd}$ decreases with fin trimming, while $R_{ch}$ keeps constant or rather increases due to fin height reduction.

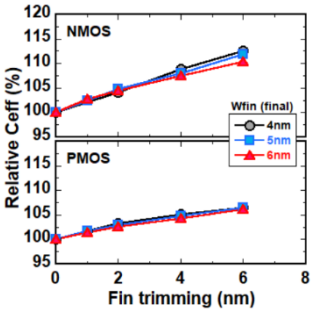


**Fig.9** TCAD simulated effective capacitance ($C_{eff}$) as a function of local fin trimming. With larger S/D extension and epi volumes, $C_{eff}$ increases with fin trimming. This penalty is still smaller than $I_{on}$ improvement.

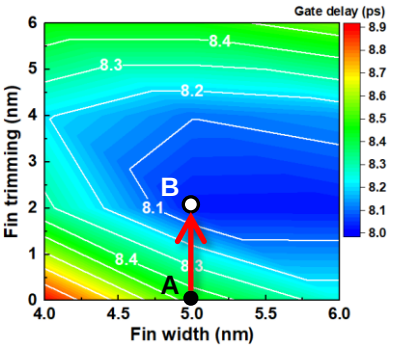


**Fig.10** Gate delay (both N and PMOS) contour as functions of fin width and fin trimming. Optimal fin trimming amount is around 2nm with final fin width of 5 ~ 6nm, leading to 5% performance gain (from A to B).

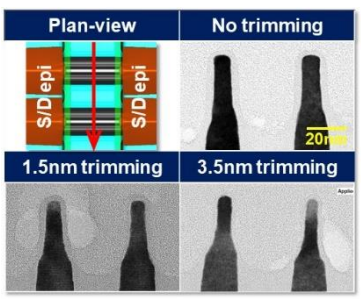


**Fig.11** TEM X-sections of fin channel region corresponding to red arrow in plan-view layout for no trimming, 1.5nm trimming, and 3.5nm trimming cases.

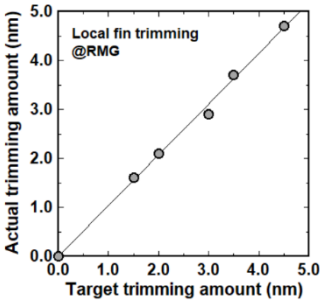


**Fig.12** Target vs. actual fin trimming amount showing well controlled selective fin trimming by up to 5nm.

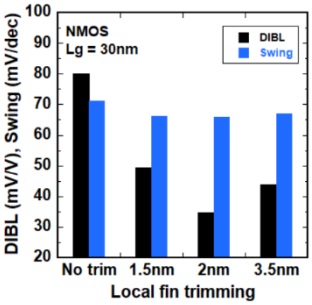


**Fig.13** Local fin trimming impact on NMOS electrostatics. Both DIBL and swing are improved with narrower fins by local fin trimming.

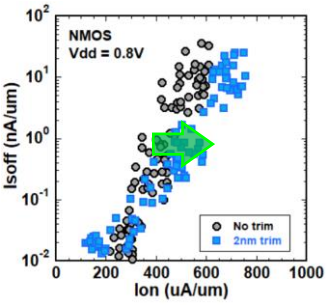


**Fig.14** $I_{on}$-$I_{off}$ performance with and without local fin trimming. More than 10% improvement is demonstrated by 2nm fin trimming.