Challenges and Opportunities for Stacked Transistor: DTCO and Device

J. Wang, S. D. Suk, A. Chu, T. Hook, A. Young, R. Krishnan, R. Bao, I. Seshadri, B. Senapati, V. Zalani, T. Philip, H. Zhou, K. Zhao, D. Dechene, B. Haran, D. Guo, H. Bu
IBM Research, 257 Fuller Rd, Albany, NY 12203; Email: junliwa@us.ibm.com

Abstract

Vertically stacked transistors are being explored as potential options for future technologies beyond GAA Nanosheet technology. In this paper, we report our investigation of challenges and opportunities for vertically stacked transistors, with a focus on block level scaling and device performance consideration. At the standard cell level, our DTCO innovation of splitting the power rails can free up a metal track for signal routing. Implementing two circuit rows for complex cells also increases available tracks. At the block level, we performed the routing study through PnR and overcome the shortage of pin access by DTCO innovations, achieving 0.55x area scaling vs non-stack technology. For device design, choices of device structures and materials for each layer are carefully evaluated. SiGe FinFET(p) on Nanosheet(n) is identified as a strong candidate for vertically stacked device architecture. MOL resistance with one-sided power rail is found to be a bottleneck limiting the device performance. Double-sided power rail (DSPR) is introduced for the first time in this work to effectively address the MOL resistance for stacked transistors.

Introduction

Area scaling has been one of the key motivations for advancing CMOS technology. For decades, this has been enabled by critical pitch scaling through patterning solutions and recently DTCO innovations. Meanwhile, the CMOS device architecture has evolved from planar to FinFET and lately GAA nanosheet structure to satisfy device scaling requirements [1]. Beyond GAA nanosheet, the semiconductor industry is eagerly looking for innovative options to keep the scaling going forward. As one of the promising candidates, vertically stacked transistor has been brought up during the past few years [2-5]. This paper reports our investigation of challenges and opportunities for stacked transistors, with a focus on area scaling at block level and device performance consideration.

Area scaling and DTCO

To continue the Moore`s law of area scaling, the standard cell area can be shrunk by reducing CPP (Contacted Poly Pitch) and shortening Cell Height. CPP scaling is limited by device design and process integration challenges, while Cell Height scaling is limited by device spacing rules and BEOL (Back End of Line) Mx Pitches. In terms of the recent scaling trend from IRDS, the area scaling for device CPP is slowing down more than Mx scaling. To overcome the CPP scaling limit, we explore vertical stacking of transistors to improve the cell height reduction.

Ideally, the vertical stacking transistor architecture provides 50% area scaling without change of critical pitches, such as CPP and Mx pitches (Fig. 1). In a typical logic standard design (Fig. 2), a 6-Track Cell Height (a) can be reduced to 3-Track as shown in (b). This shortened cell has only two signal tracks. In order to improve cell design efficiency and pin access, we move one of the power rails to the bottom to form DSPRs (c) to free up another signal track.

Due to the short cell height having a limited number of routing tracks, many complex cells require 2-rows high for intra-cell routing. Some of these 2-row cells scale less than the theoretical 0.5X. For a typical standard cell library, the usage weighted average area scaling is approximately 0.55X.

To assess the block area scaling, we use a commercial Place and Route flow to iteratively reduce the block area until the number of routing errors increases to a threshold of 100. Increasing the BEOL layers allowed for routing, we repeat the block-area-reduction iteration process to determine the route-able minimum block area vs. available routing layers. We found that we can achieve block area scaling of ~0.55x when we add one more metal layer and reduce the metal M0 and M2 pitches by ~25% to overcome reduced pin access (Fig. 3).

With the smaller block area for stacked transistors, the average wire length would be shorter. This reduced wire length, thus lower BEOL load, will translate into potential performance gain. Using a conventional non-stacked Nanosheet PDK, we estimated the potential performance benefit. For a 0.55× area reduction, ~11% performance gain due to shorter wires is estimated (Fig. 4).

Device performance

The vertically stacked transistor architecture offers opportunities for device design. As nFETs and pFETs are separately processed, they may be separately optimized. In Fig. 5 we enumerate the device architecture options that we have explored through simulation. Using a FinFET structure for pFET improves its intrinsic performance because of the hole transport benefit of conduction on the (110) plane. A strained SiGe pFET channel is advantageous in either a Nanosheet or FinFET configuration. Detailed transistor-level TCAD was performed to calculate the AC performance and power with a calibrated analytical CV/I model, including parasitic capacitances and resistances. The results are shown in Fig. 6, normalized to a nanosheet-over-nanosheet configuration. The best performance is realized by FinFET for pFET with a strained SiGe channel on top of Si Nanosheet for nFET. It is important – and perhaps surprising – to note that considerable (5%) benefit may be realized with a simple unstrained pFinFET. The advantage of the new architecture is not only in potential performance, but as shown in Fig. 7 lower power can be realized.

Further improvements on the order of 10% may be realized by optimizing the nFET contact structure where the epitaxy does not play a role in stress engineering (Fig. 8). SiGe channel for p-FinFET can provide opportunity for Tinv and Lg scaling with superior NBTI [6] for 3% AC performance benefit. If Tinv and Lg scaling can be realized for both nfet and pfet, as much as 5% performance may be obtained. (Fig. 9).

For vertically stacked transistors, excessive resistance from the bottom device to the interconnect layers would be detrimental. Innovations such as double-sided power rail (DSPR) can not only improves routing as described above, but also obviates the resistance concern (Fig. 10). DSPR could fully recover the performance loss by MOL resistance from stacked structure.

Conclusion

Stacked transistor is one of the promising candidates to continue logic CMOS technology scaling. Innovations in DTCO at block level PnR are necessary to propagate cell-level area scaling benefit to block-level area scaling benefit. Stacked transistor also provides unique opportunities to optimize device design for nFET and pFET respectively. Unique channel materials, gate dielectrics, spacer and MOL contacts for nFET and pFET respectively, as well as enablement of double-sided power rail (DSPR) can be strong boosters for performance, with a potential enhancement on the order of 20%.

References

[1] N. Loubet et al., VLSI 2017, pp230. [2] J. Ryckaert et al, VLSI 2018, pp141. [3] W. Rachmady et al, IEDM 2019, pp697. [4] C.-Y. Huang et al, IEDM 2020, pp425. [5] T.-Z. Hong, IEDM 2020 pp319. [6] D. Guo et al., VLSI 2016 pp14.

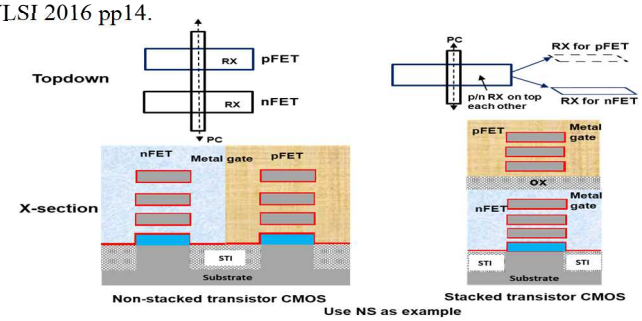

Fig. 1 (a) Typical CMOS: nFET and pFET are horizontally side by side. (b). Stacked transistor CMOS, nFET and pFET are vertically stacked on top of each other.

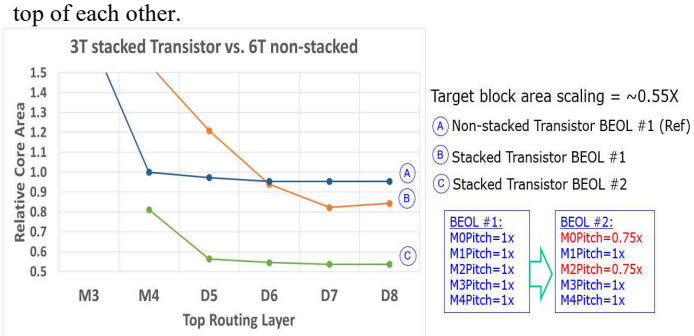

Fig. 3 Block Area vs. max. Routing Layer for various BEOL metal pitches.

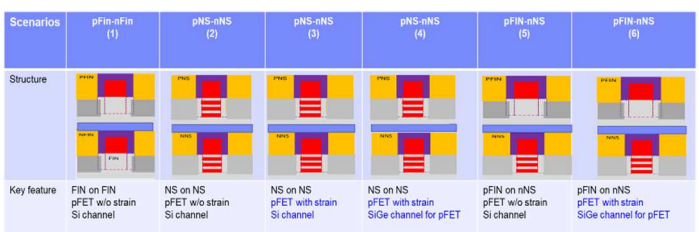

Fig. 5 Device architecture choices and channel material.

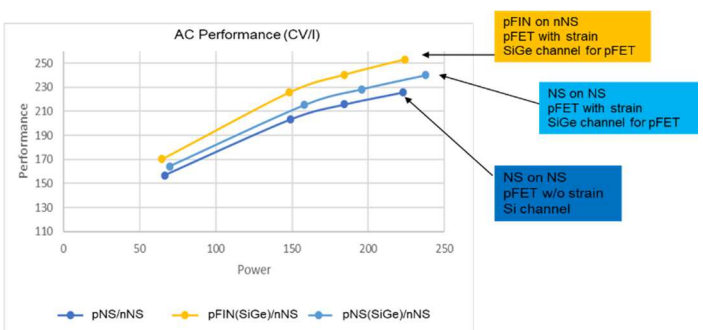

Fig. 7 Power-performance chart for different device architecture choices

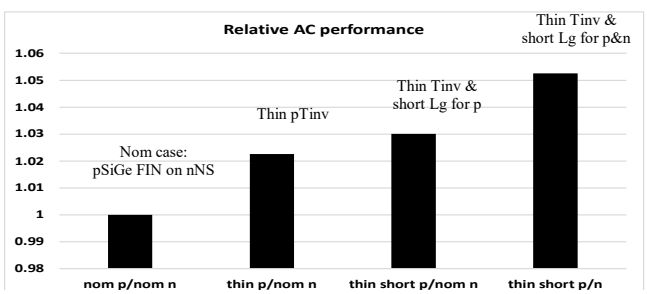

Fig. 9 Superior NBTI with SiGe FinFET provides for Tinv scaling for pFET for 2% performance. Lg scaling gives another 1% , and as much as 5% for if applied to both nfet and pfet.

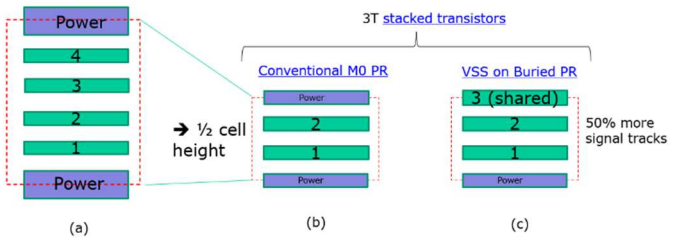

Fig. 2 In a typical logic standard design, (a) 6-Track Cell Height with 2 tracks for Powers and 4 tracks for signal. (b) allocate 1 track for power and 2 tracks for signal. (c) we move one of the power rails to the bottom to form DSPRs (not shown), freeing up a shared signal track.

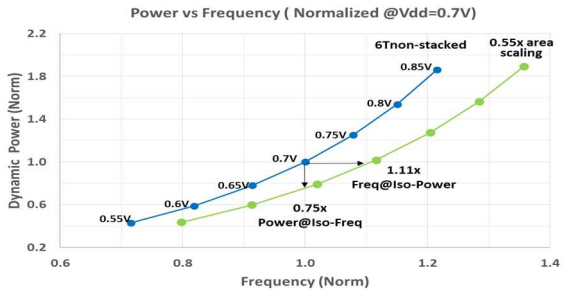

Fig. 4 Potential Power-Performance benefits from reduced wire length. For a 0.55x area scaling by stacked transistor, ~1.11x performance gain at Iso-Power or ~0.75x dynamic power reduction at Iso-Frequency due to shorter routing wire length.

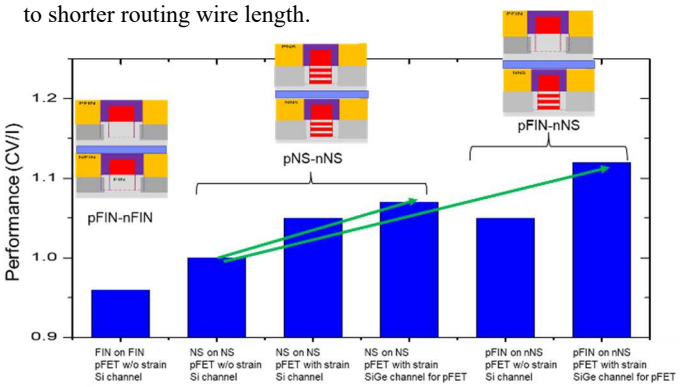

Fig. 6 With simplified metric (CV/I), two attractive candidates are found. a) combination of p-FinFET with strained SiGe channel and nFET with Nanosheet. b) combination of p-Nanosheet with strained SiGe channel and nFET with Nanosheet.

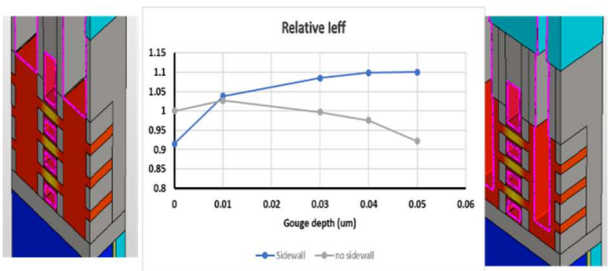

Fig. 8 Flexibility of device optimization separately for n and p. nFET S/D can be recessed without channel strain requirements.

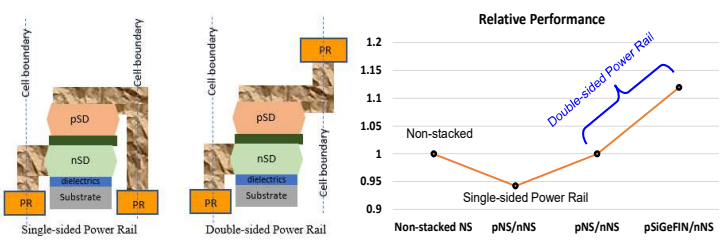

Fig. 10 MOL resistance challenge and double-sided power rail innovation to address the challenge.