Improving Performance, Power, and Area by Optimizing Gear Ratio of Gate-Metal Pitches in Sub-10nm Node CMOS Designs

Yongchan Ban, Xuelian Zhu, Jan Petykiewicz, Jia Zeng
GLOBALFOUNDRIES, 2600 Great America Way, Santa Clara, California 95054, USA
yongchan.ban@globalfoundries.com

Abstract

This paper presents improvements in performance, power, and area (PPA) obtained by optimizing the gear ratio (GR) between the Gate and vertical metal layer pitches in standard cells in sub-10nm node CMOS SoC designs. Changing the GR from 1:1 to 3:2 leads to better pin accessibility, routability, and higher cell density. This in turn enables a gate pitch relaxation and associated improvements in cell delay. Implementation of 3:2 GR ultra-dense cells in an SoC CPU block results in up to 17% higher performance, 4% smaller logic size, and 8% lower dynamic power at typical PVT conditions.

Introduction

The metal 1 (M1) layer is typically used as the first routing layer and gives input/output pins for connecting standard cells with back-end interconnect metals. With the advent of the unidirectional patterning, the bidirectional M1 is decomposed into the horizontal M0 layer and the vertical M1 layer [1-2] as shown in Fig.1. Cell dimensions are driven by the minimal contacted poly pitch (CPP, ~cell width) and horizontal metal pitch (~cell height). Given expected sub-10nm node pitches, a design-technology co-optimization (DTCO) effort is necessary to improve pin accessibility and routability for better PPA.

The ratio of pitches between Gate and M1 is commonly referred to as the gear ratio. It has traditionally been 1:1 (the same pitch for Gate and M1) in unidirectional designs due to aggressive shrinking of those layers. While the minimal CPP is dictated by process impacts on parametric variations, the M1 pitch (M1P) does not suffer from similar limitations and can be readily decreased to match other critical metal pitches. Reducing the M1 pitch to achieve a 3:2 M1P:CPP GR, as in Fig.2, can improve routability [3]. This paper demonstrates the PPA benefits and examines the manufacturability tradeoffs of an optimized GR for our advanced node CMOS technology.

Optimization of Gate-to-Metal Gear Ratio

Two high-symmetry flavors of 3:2 GR configuration are possible, as shown in Fig.3; "M1 on-top-of PC" where some M1 polygons can be placed directly over gate polygons (PC), and "M1 in-between PC" where some M1 polygons are located directly between two PCs. In addition, each symmetry option requires two versions of each cell with different M1 offset (Fig.4) in order to avoid an M1 coloring problem in sequential locations. Thus, a 3:2 GR library must have 2 times as many cells as an 1:1 library.

The two different M1 offset cell types can have varying cell delays. Fig. 5 reports an average delay gap of 0.04% across all the cells in an "M1 in-between PC" library. In sequential cells, up to 0.50% variation in the rising arc is observed. The maximum delay difference in combinational cells in all timing arcs is about 0.25%. Yet, all delay differences are within the cell characterization criteria.

We investigate the best symmetry option for 3:2 GR cells with considerations of manufacturing difficulty and design benefits. Examination of a complex cell (AOI22x1) reveals that the "M1 in-between PC" option has relatively smaller pin accessibility in between M0 and M1, but shows the best manufacturability; meanwhile, the "M1on-top-of PC" option has more pin accessibility in routing, at the cost of manufacturing challenges, e.g. V0 (M0 to M1) enclosure rules.

Enhancing PPA in SoC with the Best Gear Ratio

We applied the 3:2 GR cells to an SoC CPU block, OR1200, and compared PPA with ultra-dense design kits. 45 cells with 1:1 GR (1x M1P and CPP) and 90 (45x2 M1 shift) cells with 3:2 GR (0.71x M1P and 1.07x CPP) were used in Fig.7, and all PPA were measured at the TT/0.75v/25c corner. We choose the "M1 in-between PC" option for 3:2 GR cells since it is more compatible with current patterning capabilities. 13 metal stacks with 3 1x metal layers and a dual Vdd/Vss power rail architecture were applied. Due to CPP relaxation, the 3:2 GR library is about 7% bigger in cell size, but 1.6% faster in delay.

We first measured the maximum achievable frequency in OR1200. The 1:1 GR library achieves up to 2.4GHz, while the 3:2 GR library reaches up to 2.8GHz, shown in Fig.8. The key components of the performance benefits with 3:2 GR are 1) slight better cell delay (around 1.6%), 2) much better routability and utilization; 3:2 GR has a less logic depth at the critical path as shown in Fig.9. 3) smaller wire loads; the 3:2 GR shows smaller interconnect net delay portion for critical path groups in Fig.9. The 3:2 GR has a similar or lower total wire length, despite its bigger cell size in Fig.10. M1 metal usage in the 3:2 GR is around 7% higher at the same frequency due to its superior pin accessibility and routability. About 16% smaller via count is reported with the 3:2 GR in Fig.11. The numbers of V1 and V2 are very similar, but those of V3 and above are reduced with the 3:2 GR. This implies that the 3:2 GR has better routability, fewer detours and lower wire loads.

The utilization (cell density) in the 3:2 GR library is about 8% higher in Fig.12. Although 3:2 GR cells are about 7% larger, the overall logic area for the chip becomes about 4% smaller (block level scaling) at 2GHz target (Fig.13).

The total dissipated power is similar in the entire frequency range without consideration of chip area, as displayed in Fig.14. At the same performance (2GHz) and the same area, the 3:2 GR design shows about 8% less in total dynamic power, as shown in Fig.15. Even though the pin capacitances of 3:2 cells are slightly (1%) larger than those of 1:1 cells, the effective dynamic power dissipation capacitance in 3:2 cells is smaller due to faster switching as well as the shorter wire load.

Conclusion

We present the benefit to PPA from a 3:2 M1-to-PC gear ratio in sub-10nm node CMOS designs. Experimental results with an SoC block and ultra-dense library show that the 3:2 GR cell library promises sizeable improvements to power and performance. For future nodes, it is important to develop technology architecture definitions that have a good M1-to-PC gear ratio and allow M1 to PC offset.

References

[1] L. Liebmann et al., in Proc. VLSI Tech. Digest, pp.112, 2016.
[2] L. Clark et al., in *Microelectronics Journal*, 53, pp.105-115, 2016.
[3] L. Lu, in Proc. ACM ISPD, pp. 63, 2017.

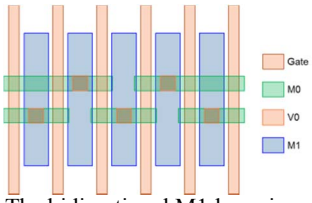

Fig. 1. The bidirectional M1 layer is decomposed into the unidirectional M0 and M1 layers in sub-10nm node technology.

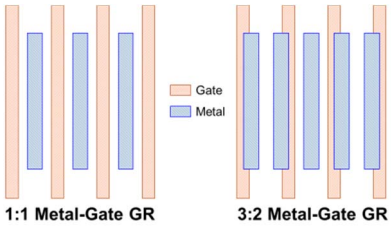

Fig. 2. Metal to Poly (Gate) gear ratio; M1 pitch in 3:2 GR is 1.5x smaller than that of PC.

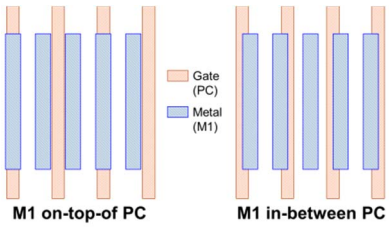

Fig. 3. 3:2 GR configuration has two high symmetry options for metal shift.

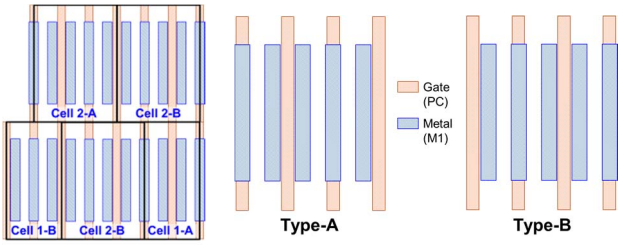

Fig. 4. For 3:2 GR, each cell needs two versions of M1-offset for cell abutting in placement.

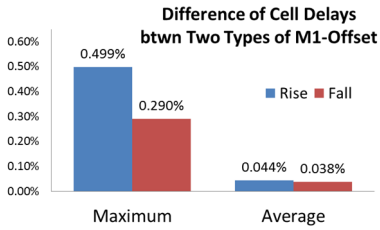

Fig. 5. The delay difference of the two types of 3:2 GR cells shows around 0.04% on average and up to 0.50% for the overall cells.

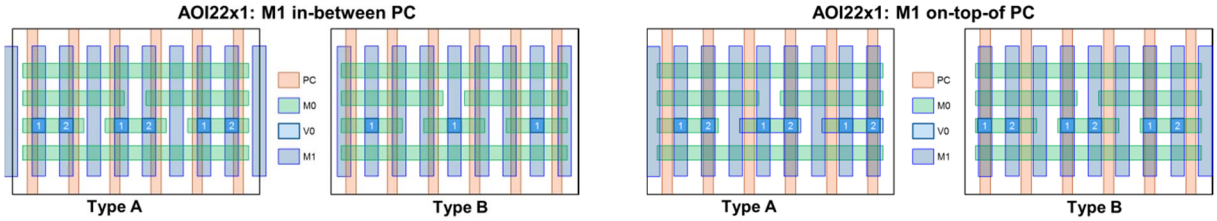

Fig. 6. The best 3:2 GR option can be found by considering DTCO. "M1 in-between PC" shows the best manufacturability yet relatively smaller number of pin accessibility in between M0 and M1, while "M1on-top-of PC" has the better pin accessibility with challenging Via enclosures.

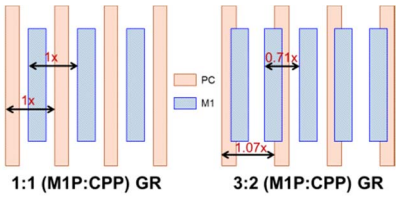

Fig. 7. 3:2 GR library shows around 7.1% larger cell size, yet about 1.6% faster in delay due to CPP expansion.

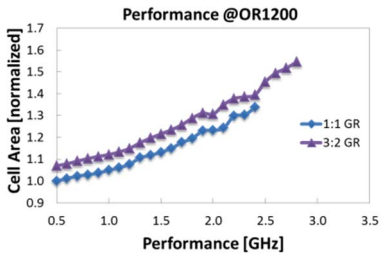

Fig. 8. 1:1 GR achieves up to 2.4GHz, while 3:2 GR reaches up to 2.8GHz.

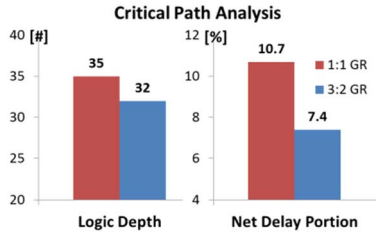

Fig. 9. 3:2 GR has less logic depth and smaller net delay portion in the critical path.

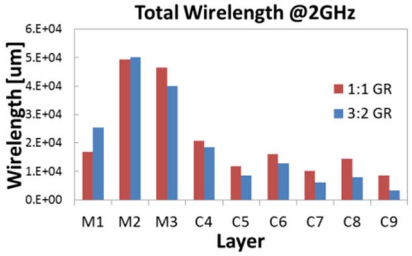

Fig. 10. 3:2 GR has higher M1 usage and less total wire length despite its bigger cell size.

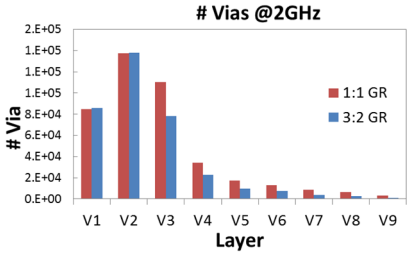

Fig. 11. About 16% smaller Via count is reported with a 3:2 GR.

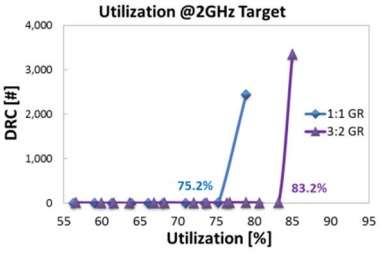

Fig. 12. 3:2 GR library shows about 8% better utilization (cell density) at the same frequency.

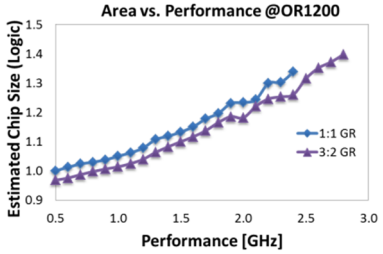

Fig. 13. The overall logic area in chip is about 4% smaller in 3:2 GR at a 2GHz target.

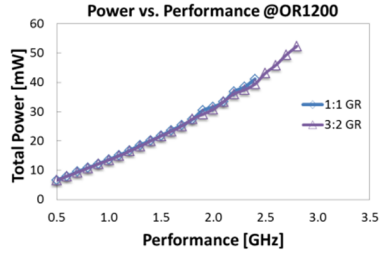

Fig. 14. The total powers are similar (slightly higher in 1:1 cells) in the entire freq. range.

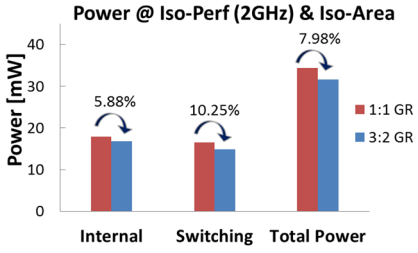

Fig. 15. At the same performance and area, 3:2 GR shows about 8% less total power.