



Invited Paper: Opportunities of Chip Power Integrity and Performance Improvement Through Wafer Backside (BS) Connection

Rongmei Chen, Giuliano Sisto, Odysseas Zografos,
Dragomir Milojevic, Pieter Weckx, Geert Van der Plas and
Eric Beyne

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

November 4, 2022

Invited Paper:

Opportunities of Chip Power Integrity and Performance Improvement through Wafer Backside (BS) Connection

Rongmei Chen, Giuliano Sisto, Odysseas Zografos, Dragomir Milojevic,
Pieter Weckx, Geert Van der Plas, Eric Beyne
IMEC, Leuven, Belgium
{Rongmei.chen, Giuliano.sisto, Odysseas.zografos, Dragomir.milojevic,
Pieter.weckx, Geert.vanderplas, Eric.beyne}@imec.be

ABSTRACT

Technology node scaling is driven by the need to increase system performance, but it also leads to a significant power integrity bottleneck, due to the associated back-end-of-line (BEOL) scaling. Power integrity degradation induced by on-chip Power Delivery Network (PDN) IR drop is a result of increased power density and number of metal layers in the BEOL and their resistivity. Meanwhile, signal routing limits the SoC performance improvements due to increased routing congestion and delays. To conquer these issues, we introduce a disruptive technology: wafer backside (BS) connection to realize chip BS PDN (BSPDN) and BS signal routing. We first provide some key wafer processes features that were developed at imec to enable this technology. Further, we show benefits of this technology by demonstrating a large improvement in chip power integrity and performance after applying this technology to BSPDN and BS routing with a sub-2nm technology node design rule. Challenges and outlook of the BS technology are also discussed before conclusion of this paper.

KEYWORDS

BEOL, BSPDN, IR drop, BS-signal Routing, SRAM, Logic

1 Introduction

The advancement of semiconductor technology requires both FEOL (front-end-of-line) and BEOL process improvement to target both density and performance/power efficiency in a system. For FEOL scaling, research has focused (among others) on alternative device materials such as carbon nanotube FET (CNTFET), 2D device etc. [1] and alternative device structures such as FinFET, Nanosheet, Forksheet [2], or 3D stacking device like complementary FET (CFET) [3] which support scaling in various ways. Nevertheless, there have not been proposed disruptive technology solutions for BEOL optimization. Most research focus has gone into proposing new BEOL materials such as Graphene, CNT [4], and other metal materials or introduce some changes in the current/compatible BEOL scenario such as Airgap, hybrid height [5] of BEOL etc. Recently, we proposed and have been continuously working on a disruptive BEOL technique to tackle

routing congestion in the chip front side. Additional to routing congestion, this technique has potential in improving chip power integrity and performance/power efficiency. The technique is based on exploiting the wafer backside (BS) to build complementary metal routing space to the that in the chip frontside (FS). In this way, we have two spatially independent BEOL resources to support chip design. The BS metal can be customized and designed without interrupting the FS BEOL which will still be the main resource of signal routing at advanced technology nodes, such as sub-5nm ones. Instead of using the BS routing for the numerous small-distance signal routing, we are making use of the wafer backside for power delivery and for global signal routing such as clock, memory IO and SoC level block to block connections. The power delivery network has been first completely moved to the chip backside, forming the BSPDN, and then the chip global routing signals are partly/selectively put in the chip/wafer backside to share the BS metal resource with the BSPDN.

2 Key Technology Enabler

As shown in Fig. 1, the flow of wafer BS connection is summarized with various Transmission Electron Microscopy (TEM) pictures and illustrations presented [6]. In sub-figure (a), device (here finFET used) up to FSM1 (Front-side Metal 1) is built on Si/SiGe-ESL epi sacks. Source/Drain are contacted by metal lines to the device active layer (M0A). M0A is connected to BPR (Buried Power Rail) via VBPR; (b) connection between Gate and the BPR is presented with similar VBPR and M0A; (c) shows a whole cross-section view of the BS connection in one of the vertical directions cut. The wafer is thinned before nTSV is processed and landed on the top of BPR, followed by BSM1 (Back-Side Metal 1). There is no cost of area in FEOL (Front-End-of-Line) due to the great alignment and direct contact between BPR and nTSV. Contact resistance between BPR and nTSV is also significantly improved to under 20 Ohms. Except the BPR used as contact bridge between FS and BS signal routing, direct TSV connection between FS BEOL and BS metals has also been demonstrated and shown [7] where sufficiently low parasitic capacitance of nTSV was shown. Details of the process flow were also presented and available in [7].

3 BSPDN Design and IR Drop Evaluation

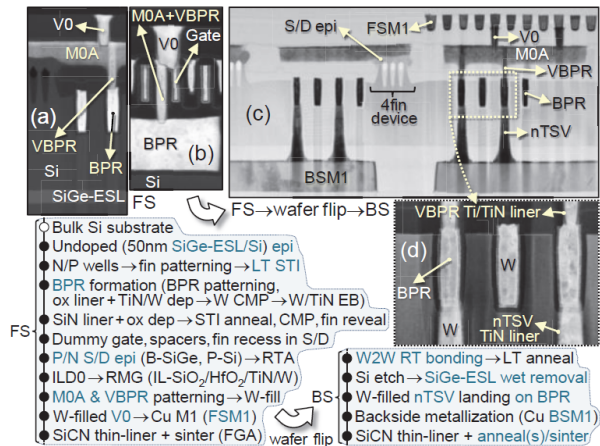


Figure 1: A demonstrated wafer BS connection integration flow using nTSV at imec [6]. Shown in the figure are the key process steps required in this flow.

3.1 FSPDN & BSPDN design

For a conventional chip design, the global PDN starts from the top of the BEOL layers. All the other metals and via layers are used to connect the power from the global to the local PDN (such as the M0 or BPR) for FSPDN. The connection between the global and local PDNs is usually highly resistive. Main reasons for this are: i) as shown in Fig. 2(a), many metal layers of small dimensions and pitches (hence large resistivity) are used for signal and power routing; ii) different pitches of consecutive metal layers require additional metal routing to connect orthogonal (pitch walking). This results in redundant and detrimental resistance involved in the IR drop path, leading to larger IR drop produced.

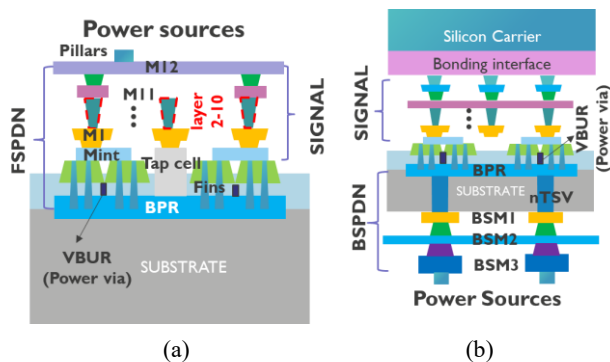


Figure 2: (a) and (b) are FSPDN and BSPDN schematics respectively. 2.5D/2D Mimcap can be integrated on the top of the FSPDN or bottom of the BSPDN between the pillars and on-chip PDN part. The pillars presented here contain parasitic capacitance and resistance with skin-effect considered.

For the BSPDN case, we move completely the PDN from the chip front-side (used for the signal routing layer in conventional BEOL) to the chip/wafer back-side [8] as shown in the schematic of Fig. 2 (b). As introduced in the technology section of this paper, the wafer is thinned to enable connection of the local power rails in BPR to

BS metals through nTSV. Thanks to the fewer (can be minimum 3 BS metal layers) and reserved metal layers for PDN and direct/high conductivity contact to the BPR, we can further customize the BS metal layers for optimized power delivery integrity by achieving much smaller PDN resistance and improved IR drop than the FSPDN counterpart. Meanwhile, the FS BEOL is saved for only signal routing improving the chip performance.

BSPDN can also be a perfect fit for developing stacking chips, i.e. 3D IC. We hereby showcase a possibility of 3D chip stacking using a face-to-face and wafer-on-wafer hybrid-bonding technique developed at imec [9]. The bonding technique enables high-density while lower resistivity connections for signal/power between two wafers and chips. As shown in Fig. 3, the power is delivered at the beginning from the BSPDN of the bottom chip. Further on, the power is shared with the top chip through the BEOL in the bottom chip, hybrid bonding and then the BEOL of the top chip. This long journey of power delivery for the top die leads to significant IR drop in the top die. For balancing IR drop in the top die, a lower-power hungry chip (such as SRAM cache) is more preferred than a logic chip (CPU, as studied in the following part of this paper).

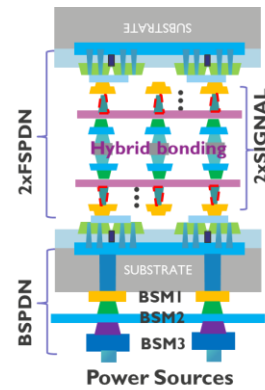


Figure 3: 3D chip stacking based on a face-to-face and wafer-on-wafer technique. BSPDN is applied to the bottom die by default. With additional BSPDN on the top die and 2.5D Mimcap integrated in both dies, the IR drop of the top die can be significantly reduced.

3.2 IR Drop Evaluation

We have shown in [7] that by using 2.5D Mimcap (Metal-Insulator-Metal) capacitor integrated in the BEOL, an improvement of 6x Mimcap density can be achieved compared with a conventional 2D Mimcap. The 2.5D Mimcap plus the BSPDN combination was demonstrated to be a significant booster in reducing the chip IR drop. As shown in Fig. 4, the heatmaps of IR drop for a lower-power CPU of sub-2nm node are presented based on various PDN structures. The BSPDN+2.5D Mimcap combination can achieve 32.1%/23.5% smaller 95th percentile IR-drop than no Mimcap/2D Mimcap counterparts respectively (Fig. 5). Meanwhile, it has 36.3% reduced 95th percentile IR drop than the FSPDN+2.5D Mimcap counterpart. It is worth mentioning that the BSPDN without any Mimcap integration can achieve ~10% smaller 95th percentile IR drop than the FSPDN with 2.5D Mimcap counterpart. It should be noted that we assume the same power density map

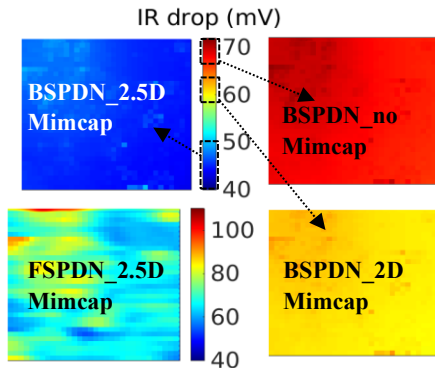


Figure 4: Heatmaps of IR drop of a lower power CPU based on various PDN and Mimcap structures. Different color bars are used for the FSPDN and BSPDN. Much higher IR drop is observed for FSPDN than BSPDN. 2.5D Mimcap also helps further improve the BSPDN [7].

generated from physical design for use in the IR drop simulation in the chip level. Hence, the IR drop results can fairly reflect the capability of the different power delivery network in reducing IR drop. It is true that different PDN may need different physical implementations and hence different power density maps [10] but this is not discussed here for simplicity without impacting conclusion significantly.

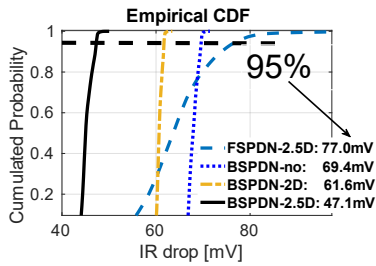


Figure 5: Empirical cumulative density function (CDF) curves of IR drop for various 2D design PDN structures [7].

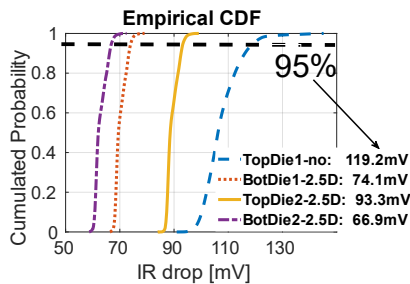


Figure 6: Empirical cumulative density function curves of IR drop for various 3D CPU-on-CPU PDN structures [7].

For a 3D CPU-on-CPU case study, we also assume an identical power density map (the same to 2D CPU) used for the top and bottom CPUs. As shown in Fig. 6, the top CPU IR drop is found as expected to have a much larger IR drop than the bottom die CPU. As analyzed earlier in this paper, this is due to the much larger IR

drop path for the top die (including 2x FSPDN + 1x BSPDN involved). Thanks to the utilization of BSPDN+2.5D Mimcap in the top die, the 95th percentile IR drop is improved by 21.7% than top die with BSPDN but without 2.5D Mimcap.

4 BS Signal Routing and Performance/Power Evaluation

In the technology section, we have shown two possible signal connection scenarios to wafer BS through nTSV. One is achieved through a VBPR+BPR+nTSV option and the other is through a direct nTSV connection from FS to BS. Fig. 7 shows various signal paths based on FS, mainly the FSM3+FSM4 and on the BS (BSM1+BSM2) using the second BS connection scenario [11]. Compared with the FS signal routing, the BS signal routing can avoid the first a few fine-pitch and high resistivity metal layers which are particularly detrimental for long-distance signals routing. With this advantage in mind, we first design a SRAM macro using the BS for global signal routing and then showcase the potential benefit of signal routing for generic logic gates driving long wirelengths.

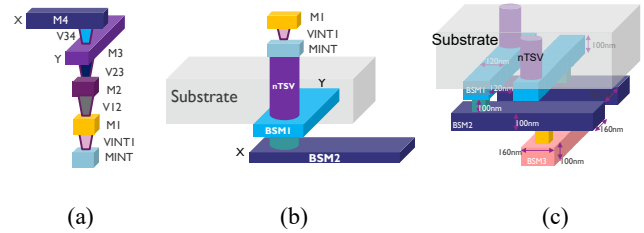


Figure 7: Schematic of different types of signal routing paths based on FS, here the FSM3+FSM4 (a), and BS only metals, here BSM1+BSM2 (b). (c) gives the complete view of BS signal routing layers, down to BSM3 [11]. The details of the BS metal dimensions are illustrated in (c) as well, which are consistent with the table electrical parameters in Tab. I.

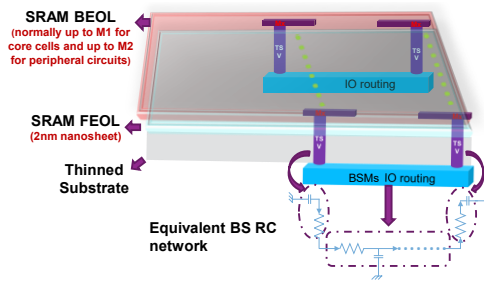


Figure 8: SRAM macro global routing with BS connections [11].

The SRAM macro using BS for global routing is designed and shown in Fig. 8. Global routing signals in SRAM macro mainly include the IO and the address ones. These signals spread across a single macro or even macro array depending on the required capacity and configuration of the caches. A larger capacity a cache

Tab. I The designed BS metals RC and comparisons with FS metals [11].

FS Metal layers (FS M2~M4)						nTSV	BS Metal layers (BS M1~M3)					
Metal layers	W/S (nm)	T (nm)	C _{tot} (fF/um)	R (Ohm/um)	FS vias (Ohm)	100nm(x)*100nm(y)*300nm(z) R=-20 Ohm C=-0.04fF Assuming keep out zone~10nm	Metal layers	W/S (um)	T (um)	C _{tot} (fF/um)	R (Ohm/um)	BS vias (Ohm)
FSM2 (Ru)	8	24	0.361	881	V ₁₂ =26.40		BSM1 (Cu)	0.12/0.1	0.10	0.2277	1.833	V ₁₂ =0.43
FSM3 (Ru)	14	28	0.259	383.3	V ₂₃ =37.38		BSM2 (Cu)	0.16/0.1	0.10	0.1867	1.374	V ₂₃ =0.25
FSM4 (Cu)	24	48	0.183	86.6	V ₃₄ =37.95		BSM3 (Cu)	0.16/0.1	0.10	0.1796	1.374	

is designed, longer global signal routing required. Hence, we expect higher benefit in power and performance for larger SRAM macro than small macro. Fig. 9 (a) shows various SRAM macro global routing delay improvement as a function of the macro physical columns and rows of SRAM cells (accumulating all the macro sub-arrays contributions). Ranging from 256 kbit to 4 Mbit macro sizes, the macro global routing performance improvement ranges from 28%~44%. Meanwhile, in Fig. 9 (b), the global routing power efficiency is improved from 20% to 32%. The delay reduction and power efficiency improvement come from both the BS metal R&C (Resistance & Capacitance) reduction as detailed in Tab. I. By further optimizing the BS metal R&C values (which is achievable as the BS metals are much more flexible than FS for configuration), these advantages over FS routing can be further increased.

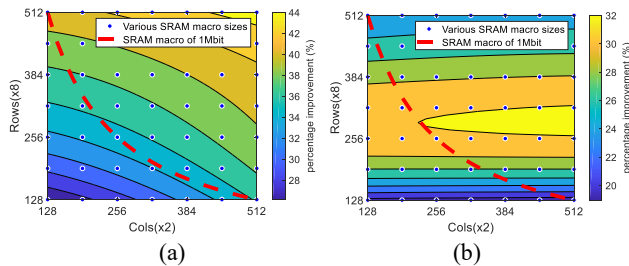


Figure 9: Performance (a) and power efficiency (b) advantages of BS metal routing-based SRAM Macro design over the conventional SRAM macro design counterpart [11]. They are of the same FEOL technology and macro configurations of various SRAM cell columns and rows with the exception of the different global routing schematics in Fig. 7.

Similarly, we found performance improvement of logic gates driving long wirelength using BS routing over the FS counterpart. The improvement is also overall increased with the BEOL wirelengths driven (in the table and the plot of Fig. 10). Hereby the performance of logic gate was characterized by doing ring-oscillator simulation with the parasitics of BEOL (also FEOL) extracted. This was also done after finishing a physical design of a lower power CPU at a sub-2nm node.

5 Challenges and Outlook

Logic Cells	Drive Strength /Fan out	Total Wirelength (um)
BUFFER	D1/F4	2.39
BUFFER	D2/F6	7.18
INVERTER	D1/F2	1.38
OR2	D2/F3	1.87
AOI-22	D1/F1	0.99
BUFFER	D4/F7	10.963
OAI-211	D1/F1	1.35
4-way NAND	D1/F1	1.86
4-way NOR	D1/F1	0.464

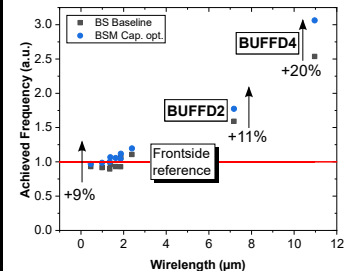


Figure 10: BS metal routing-based various logic gates driving a certain metal wirelength and their performance characterization using ring-oscillator. The BEOL loaded for each logic gate was extracted from a physical design of a low power CPU.

As shown and demonstrated in the paper, the wafer BS connection technology does bring a lot of new chances of chip power integrity and performance improvement by allowing PDN design and global routing for signals in the chip/wafer BS. In addition to this, the BPR (one case study of BSPDN local power rail) allows us to reduce the standard cell height by removing the local PDN metal to the chip BS, saving 1 or 2 tracks of metals in the frontside of standard cells [12]. This makes the scaling of technology node down to sub-3nm and further easier. Innovation of replacing BPR by direct contact technique [13] can further enable standard cell height reduction and technology node scaling. As for the challenge of this technology massive application, new EDA tools should be developed before the BSPDN and BS signal routing can be enabled and optimized for chip design. In system level, where and how BS technology should be used in SoC is also a big question for the industrial community. It is not always necessary and beneficial to use this technology for every block in an SoC. Selectively using the BS for certain blocks of an SoC in a system level is what we should further investigate and research.

6 Conclusion

This paper presents the disruptive technology of wafer BS connection for enablement of BSPDN and BS signal routing. By replacing the FSPDN with BSPDN, significant IR drop reduction is realized for a low power CPU design of 2D and 3D CPU-on-CPU configurations. 2.5D Mimcap can further boost the benefit of using BSPDN for the 2D and 3D ICs. BS routing for global signals in

SRAM macro and logic design show not only performance improvement but also power reduction. For a future scaled technology node $< 2\text{nm}$, the BSPDN and BS signal routing will be an essence of chip technology and design for both mobile and high-performance computing applications.

REFERENCES

- [1] R. Chen et al., "Carbon Nanotube SRAM in 5-nm Technology Node Design, Optimization, and Performance Evaluation—Part I: CNFET Transistor Optimization," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 30, no. 4, pp. 432–439, Apr. 2022, doi: 10.1109/tvlsi.2022.3146125.
- [2] H. Mertens, et al., "Forksheets FETs for Advanced CMOS Scaling: Forksheet-Nanosheet Co-Integration and Dual Work Function Metal Gates at 17nm N-P Space" 2021 Symposium on VLSI Technology, Jun. 2021.
- [3] P. Schuddinck et al., "PPAC of sheet-based CFET configurations for 4 track design with 16nm metal pitch," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), Jun. 2022, doi: 10.1109/vlsitechnologyandcir46769.2022.9830492.
- [4] A. Farokhnejad et al., "Evaluation of BEOL scaling boosters for sub-2nm using enhanced-RO analysis," 2022 IEEE International Interconnect Technology Conference (IITC), Jun. 2022, doi: 10.1109/iitc52079.2022.9881286.
- [5] R. Chen et al., "Variability Study of MWCNT Local Interconnects Considering Defects and Contact Resistances—Part I: Pristine MWCNT," *IEEE Transactions on Electron Devices*, pp. 1–8, 2018, doi: 10.1109/ed.2018.2868421.
- [6] A. Veloso et al., "Scaled FinFETs Connected by Using Both Wafer Sides for Routing via Buried Power Rails," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), Jun. 2022, doi: 10.1109/vlsitechnologyandcir46769.2022.9830177.
- [7] R. Chen et al., "Backside PDN and 2.5D MIMCAP to Double Boost 2D and 3D ICs IR-Drop beyond 2nm Node," 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), Jun. 2022, doi: 10.1109/vlsitechnologyandcir46769.2022.9830328.
- [8] G. Sisto et al., "IR-Drop Analysis of Hybrid Bonded 3D-ICs with Backside Power Delivery and μ - & n- TSVs," 2021 IEEE International Interconnect Technology Conference (IITC), Jul. 2021, doi: 10.1109/iitc51362.2021.9537541.
- [9] R. Chen et al., "3D-optimized SRAM Macro Design and Application to Memory-on-Logic 3D-IC at Advanced Nodes," 2020 IEEE International Electron Devices Meeting (IEDM), Dec. 2020, doi: 10.1109/iedm13553.2020.9371905.
- [10] R. Chen et al., "Power, Performance, Area and Thermal Analysis of 2D and 3D ICs at A14 Node Designed with Back-side Power Delivery Network," 2022 IEEE International Electron Devices Meeting (IEDM), Dec. 2022.
- [11] R. Chen et al., "Design and Optimization of SRAM Macro and Logic Using Backside Interconnects at 2nm node," 2021 IEEE International Electron Devices Meeting (IEDM), Dec. 2021, doi: 10.1109/iedm19574.2021.9720528.
- [12] J. Ryckaert et al., "Extending the roadmap beyond 3nm through system scaling boosters: A case study on Buried Power Rail and Backside Power Delivery," 2019 Electron Devices Technology and Manufacturing Conference (EDTM), Mar. 2019, doi: 10.1109/edtm.2019.8731234.
- [13] A. Veloso et al., "Insights into Scaled Logic Devices Connected from Both Wafer Sides" 2022 IEEE International Electron Devices Meeting (IEDM), Dec. 2022.