BS-PDN-Last: Towards Optimal Power Delivery Network Design With Multifunctional Backside Metal Layers

Min Gyu Park, Amaan Rahman, Sung Kyu Lim
School of ECE, Georgia Institute of Technology, Atlanta, GA
mgpark@gatech.edu

***Abstract*—The increasing demand to maximize PPA gains with backside metal layers has driven their function beyond power delivery. This study introduces a novel BS-PDN-last flow, crucial for leveraging multifunctional backside, by deferring power routing to the post-signal routing stage. This approach addresses the IR-drop and performance trade-offs inherent in conventional PDN-first flows. Experimental results show that the BS-PDN-last flow achieves a 90% reduction in Total Negative Slack and a 12% performance gain with BS-CDN. Additionally, our work presents the first comprehensive comparison of FS-PDN, BS-PDN, and multifunctional backside designs, evaluated on both physical design and workload metrics leveraging accurate vector-based analysis.**

I. Introduction

As CMOS technology scales down, increasing resistance in the frontside metal layers has led to a substantial IR-drop. The Backside Power Delivery Network (BS-PDN) has emerged as a promising solution, incorporating thicker and less resistive backside metals beneath the silicon substrate. Furthermore, backside power delivery is motivated to decouple the PDN from frontside signal and clock routes, improving power integrity and system performance due to improved frontside routability.

However, backside process incurs additional costs due to advanced processing such as wafer thinning and nano-Through Silicon Vias (nTSVs) formation [4]. To offset these costs, there is an increasing demand to leverage backside metal resources for maximizing Power, Performance, and Area (PPA) gains. For instance, recent research efforts push towards the integration of Backside Clock Delivery Network (BS-CDN) [2] and Backside Signal Routing (BSS) [5]. These studies focus on maximizing the potential of backside metals by extending their function beyond power delivery. However, they adhere to the conventional PDN-first flow, wherein BS-PDN is implemented initially, leaving BS-CDN and BSS to be deployed in the remaining regions. This approach induces substantial disruption of cell placement due to design rule violations encountered during the conversion of frontside cells to the backside, shown in Fig. 1-(a).

For the first time, we investigate the optimal sequence of design steps to maximize the benefit of multifunctional use of backside metal layers. We introduce a novel BS-PDN-last flow, which postpones P/G routing after the completion of frontside clock and signal routing as well as backside clock routing. By positioning P/G routing in the later stages, the BS-PDN-last breaks down the inherent trade-off between IR-drop and performance gain on the conventional PDN-first flow, demonstrating up to an 88% reduction in Total Negative Slack. Furthermore, we offer a comprehensive comparison of frontside only, backside with PDN only, and multifunctional backside designs, assessed through both physical design and workload metrics using precise vector-based analysis. The BS-PDN-last flow with BS-CDN achieves a 16.1% improvement in performance and an 18.9% increase in energy efficiency on the ResNet-50 workload compared to the frontside-only design.

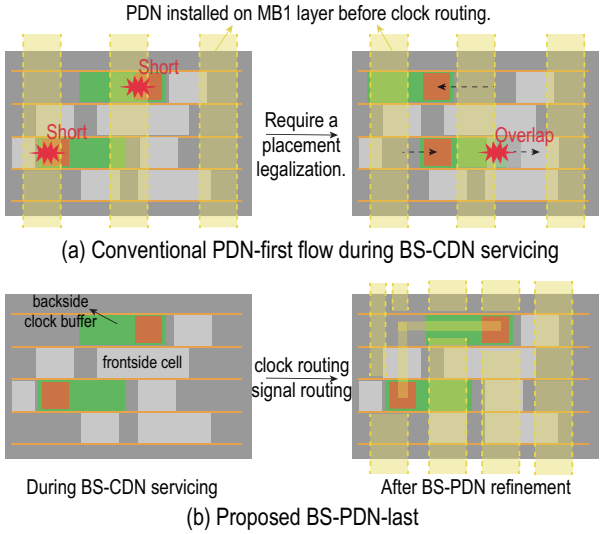


Fig. 1: Comparison between PDN-first and BS-PDN-last flows: The conventional PDN-first flow triggers substantial placement legalization to resolve short between PDN and nTSVs as well as cell overlap. However, BS-PDN-last maintains optimal cell positions with maximization of PDN.

II. Related Work and Shortcoming

Backside Clock Delivery Network was initially presented in [2], [8]. The approach involves building the clock tree on the frontside first, then converting selected clock buffers into backside buffers. This method employs a paired-buffer strategy, as illustrated in Fig. 2-(a): backside-out buffers transfer the clock signal from the frontside to the backside, while backside-in buffers route the clock from the backside to the frontside. To achieve this, nano-TSVs are employed on the output and input pins, respectively, within each cell. They revealed that combining BS-CDN with BS-PDN provides performance improvements of 8.7% in one benchmark. However, the performance gains were marginal in other benchmarks, achieving only 1.3% and 2.8% improvements, respectively. This failed to deliver improvements compelling enough to offset the increased costs of backside integration, raising concerns about whether successful improvements can be achieved in industrial-scale designs. We also face similar challenges when using the PDN-first flow with BS-CDN as shown in Table II.

III. Problem Formulation and Benchmark

A. Problem Formulation

The conventional PDN-first flow employed in prior work [2] follows the sequence of backside power routing, frontside clock construction, backside clock servicing, clock routing, and frontside

signal routing. In scenarios where backside metal layers are utilized for multiple purposes, our goal is to determine the optimal sequence of each step to maximize the benefits of backside integration for both performance and IR drop.

B. Benchmark Architectures

We evaluate the impact of BS-PDN-last flows using the Rocket Chip, an open-source processor, and the Gemmini SoC, a DNN accelerator system, along with a simple pure logic AES benchmark.

The Rocket Chip [1], a RISC-V CPU generator, provides a complete SoC RTL with support for CPU cores, a memory system interfacing with L2 cache, MMIO peripherals, and a DMA device. The specific configuration used in this study includes a dual-core setup, with each Rocket Core featuring a 5-stage in-order scalar processor, along with 16kB L1 data cache and 16kB L1 instruction cache.

The Gemmini project [3] offers a reconfigurable full-stack platform that integrates the DNN accelerator Gemmini, a CPU, and a L2 cache, enabling comprehensive system-level analysis. This platform not only generates hardware RTL for given configurations but also enables workload simulation, providing instance-level toggle rates and total cycle counts. For this study, we utilized a Gemmini accelerator configured with a 16x16 systolic array comprising a 4x4 tile and 4x4 mesh, a 4-bank 128kB scratchpad, a 2-bank 64kB accumulator, a Rocket single core with 16kB L1 data cache and 16kB L1 instruction cache, and a 128kB L2 cache.

IV. Backside PDN-last Flow

A. Motivation

With the conventional PDN-first flow, the conversion of frontside buffers to backside requires careful consideration of Power/Ground (P/G) straps on Backside Metal layer 1 (MB1), as shown in Fig. 2-(b). They can cause a P/G-to-clock short if any overlap happens between MB1 P/G straps and nano-TSV in the backside buffer. Additionally, backside buffers have a longer width than frontside buffers due to the presence of the nano-TSV. This can potentially lead to overlaps with adjacent cells if the back buffer is placed in its original front buffer position, as illustrated in Fig. 2-(c). To address potential design rule check (DRC) violations within a PDN-first flow, Algorithm 1 legalizes cell placement with minimal displacement. Recognizing that relocating adjacent cells facilitates DRC resolution, the algorithm allows the movement of both BS cells and surrounding cells. This is achieved by employing an advanced legalizer option with the *place* status set for all cells, excluding those on timing-critical paths.

However, such an approach can significantly disrupt the original clock tree structure. The impact becomes more pronounced with higher BS-PDN utilization, as shown in Table I: with a 70% BS-PDN utilization in a PDN-first flow, backside clock buffer displacement reaches up to 4.13 μm, equivalent to a vertical shift of approximately 35 rows. Such substantial alterations to the clock tree structure degrades the performance compared to the design without BS-CDN. Furthermore, even with 50% PDN utilization, the effectiveness of backside clock routing is reduced compared to a BS-CDN implementation without PDN. This result reveals the following insights: (1) The PDN-first flow obstructs backside buffer conversion, potentially leading to suboptimal performance or performance degradation. (2) The PDN-first flow introduces a trade-off between IR-drop reduction and performance gain, depending on PDN utilization. This trade-off arises because higher PDN utilization increases interference but reduces IR-drop.

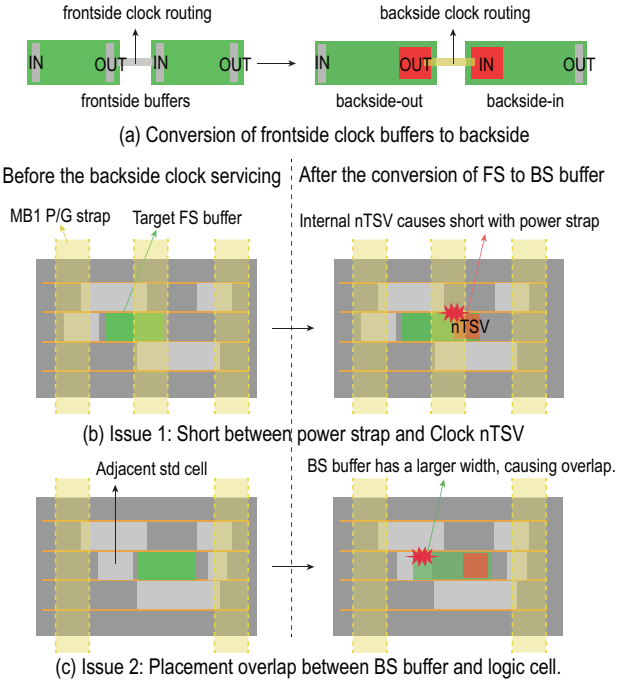


Fig. 2: Backside clock buffer overlap issues in [2]. (b) between a nTSV and a backside wire, (c) between a BS buffer and a frontside logic cell.

TABLE I: Quantification of the overlap issues with PDN-first flow described in Fig. 2. In this experiment, Gemmini SoC services 218 backside clock buffers for BS-CDN.

BS-CDN	disabled	enabled		
PDN util.	90%	None	50%	70%
Displacement of clock buf. (μm)	- -	- -	Mean: 0.27 Max: 0.49	Mean: 0.39 Max: 4.13
WNS (ps)	192.6	114.1 (-40.8%)	177.3 (-7.96%)	206.6 (+7.27%)
D. IR-drop (mV)	54.6	-	89.9	79.3

B. Overview of the Approach

To break down the inherent trade-off in the PDN-first flow, we proposed the BS-PDN-last flow, shown in Fig. 3. It is designed to fully leverage backside metal layers, supporting BS-CDN to maximize the performance gain. This methodology postpones PDN routing to the final stage of physical design, introducing an initial routing blockage on backside metal layers to pre-allocate essential regions. It accommodates various process constraints, including metal density limits, regular P/G pad placement rules, and spacing requirements between backside power and clock routing for signal integrity, through parameter adjustments in the Algorithm 2. Additionally, if the design fails to meet IR-drop targets, the BS-PDN-last flow can re-build PDN with minimal impact on overall physical design time.

C. Step 1: BS-PDN Planning

To defer PDN routing after the signal routing stage, it is crucial to ensure that BS-CDN does not obstruct power delivery to each cell and macro. Therefore, the BS-PDN planning step is implemented before utilizing backside resources for other purposes. This step pre-allocates routing blockages in critical regions to ensure proper P/G routing in subsequent steps.

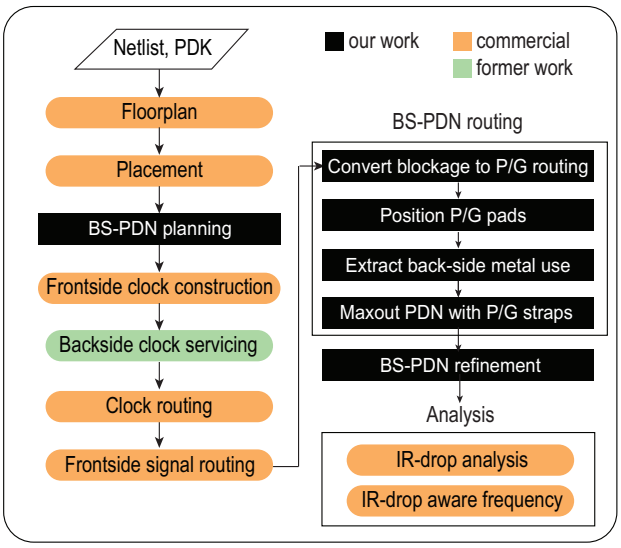


Fig. 3: Proposed BS-PDN-last flow: It preserves optimal backside clock buffer placement while maximizing PDN utilization to mitigate IR-drop.

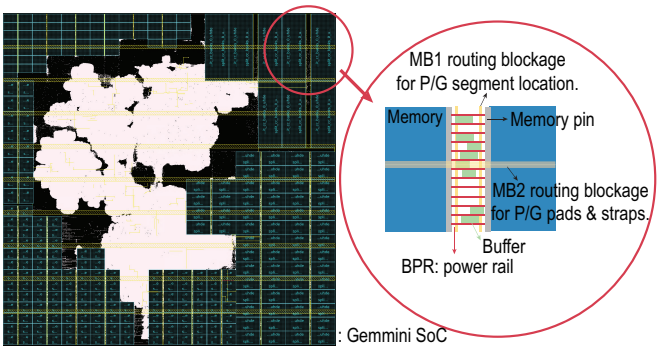


Fig. 4: BS-PDN planning: Routing blockages are added on MB1 and MB2 layers to ensure power delivery to every component.

In detail, for the design with memory macros, EDA tools often place buffers within the narrow spaces between memory macros to meet the timing. During multifunctional backside servicing, backside buffers can limit available space for power routing in these regions. To prevent such issues, Fig. 4 illustrates the placement of routing blockages on the vertical MB1 layer to reserve future P/G segment areas. Moreover, when the P/G pads need to follow a regular arrangement, routing blockages are established on the horizontal MB2 layer to facilitate subsequent assignment of MB2 P/G straps. These MB2 P/G straps will be aligned directly above the regular P/G pads. With these two strategies, essential power routing regions can be pre-allocated, facilitating seamless integration with BS-CDN while maintaining robust power delivery across all cell instances.

D. Step 2: BS-PDN-aware Backside Clock Servicing

Algorithm 1 is employed to identify candidates within the frontside clock tree and switch to backside buffers in both PDN-first and BS-PDN-last flows. Candidates are selected from buffers located on the path between the clock pin and the launch DFF of timing-critical paths. This effectively minimizes Worst Negative Slack (WNS) by reducing the delay from the clock pin to the launch DFF and providing additional timing margin for signal routing between the launch and capture DFFs. To ensure the paired scheme of backside

Algorithm 1 BS-PDN-aware Backside Clock Servicing

Input: $\mathbf{G_{clock}}$: Clock tree graph using only frontside buffers.
$\mathbf{Path_{crit}}$: List of timing critical paths in order of worst case.
$\mathbf{N}$: Target percentage.
Output: $\mathbf{G^*_{clock}}$: Clock tree graph with backside buffers.

```
Lock the location of logic cells (non-clock cells) on critical paths.
for path in {top N% of Path_crit} do
    ctp ← clock tree path from clock pin to launch DFF on path
    for buf in {reverse order of ctp} do
        if dests of buf are all frontside buffers then
            Convert buf to BS-OUT and dests to BS-IN.
            Legalize placement with minimum displacement.
```

buffers, the algorithm verifies that all destinations are frontside buffers, enabling their conversion to backside-in buffers. Furthermore, logic cells on timing-critical paths are fixed during legalization, ensuring no disruption to the placement quality. For the BS-PDN-last flow, as PDN is not installed in the initial phase, only a small amount of MB1 routing blockage and cell placement overlap causes legalization.

E. Step 3: BS-PDN Routing

Once the physical design is done alongside routing blockages on MB1 and MB2 layers, these blockages are converted into P/G routing, and P/G pads are positioned according to a specified pitch. Furthermore, backside metal regions used by BS-CDN and pre-installed BS-PDN are identified to ensure P/G routing compatibility in subsequent steps. For the MB1 layer, regions with existing MB1 metal usage, nano-TSV locations within backside clock buffers, and VB1 (via between MB1 and MB2) locations are extracted. On the other hand, for the MB2 layer, areas with existing MB2 metal usage and VB1 locations are extracted.

Unlike the conventional FS-PDN, which must balance routing resource allocation and IR-drop mitigation due to interference with frontside routing, BS-PDN focuses on maximizing the remaining backside metal resources. This strategy not only minimizes IR-drop but also enhances lateral thermal conduction [6]. By implementing the BS-PDN after BS-CDN servicing, it is possible to fully maximize PDN deployment without concerns of interference with BS-CDN.

Accordingly, first step of Algorithm 2 positions P/G straps between existing backside metal usage while avoiding overlap and spacing rule violations. For the maxout process with P/G straps, we can adjust the width and pitch of P/G straps to achieve the target PDN utilization, while also varying spacing to control the impact of P/G straps on backside clock routing. These adjustments help to manage key factors, including IR-drop, thermal conduction, coupling capacitance, and signal integrity.

F. Step 4: BS-PDN Refinement

In conventional power routing, P/G straps are used in a form that connects both ends of the core to ensure connectivity between different metal layers regardless of strap locations. However, when a sufficient number of P/G straps are installed, short P/G segments can also maintain reliable inter-layer connections. Additionally, as illustrated in Fig. 5-(a), in cases where backside clock routing is densely packed, broad areas without P/G straps happen, leading to significant IR-drop. Therefore, Algorithm 2 positions shorter P/G segments between existing P/G straps in step 2, avoiding any overlap with previously installed backside routing. By incorporating

Algorithm 2 BS-PDN Routing & Refinement.

Input: $\mathbf{GDSII_{in}}$: Layout with P/G planning and backside clock.
$\mathbf{MB1_x}, \mathbf{MB2_y}$: Lists of [*start*, *end*] intervals indicating regions where the respective metal layers are occupied.
$\mathbf{MB1_y}, \mathbf{MB2_x}$: Intervals in the y and x-direction representing the occupied ranges for each interval in $\mathbf{MB1_x}$ and $\mathbf{MB2_y}$.
$\mathbf{w1}, \mathbf{w2}, \mathbf{w3}, \mathbf{w4}$: Target widths of P/G.
$\mathbf{p1}, \mathbf{p2}, \mathbf{p3}, \mathbf{p4}$: Target pitch values.
$\mathbf{spacing}$: P/G spacing from backside clock routing.
Output: $\mathbf{GDSII_{out}}$: Layout with maximized P/G utilization.

Step 1: Maxout BS-PDN with P/G straps.
for i in $\{1, 2\}$ **do** ▷ loop over MB1 and MB2.
  $unit \leftarrow \mathbf{w}i + \mathbf{p}i$
  for j in $\{0, \ldots, \text{len}(\mathbf{MB}i_{\mathbf{x}|\mathbf{y}})\}$ **do**
    ▷ Insert P/G straps between consecutive occupied regions.
    $interval \leftarrow start_j - end_{j-1}$
    if $interval - 2 \times \mathbf{spacing} > unit$ **then**
      $N \leftarrow \text{CEIL}((interval - 2 \times \mathbf{spacing})/unit)$
      Integrates N P/G straps between end_{j-1} and $start_j$.
Step 2: Add P/G Segments
$unit \leftarrow \mathbf{w}(i+2) + \mathbf{p}(i+2)$ ▷ loop over MB1, MB2. ($i \in \{1, 2\}$)
for j in $\{0, \ldots, \text{len}(\mathbf{MB}i_{\mathbf{x}|\mathbf{y}})\}$ **do**
  $width \leftarrow end_j - start_j$
  if $width > unit$ **then**
    $M \leftarrow \text{CEIL}(width/unit)$
    for k in $\{0, \ldots, \text{len}(\mathbf{MB}i_{\mathbf{y}|\mathbf{x}}[\mathbf{j}])\}$ **do**
      $H \leftarrow start_k - end_{k-1}$
      Inserts M P/G segments of $\mathbf{w}i$ width and H height.

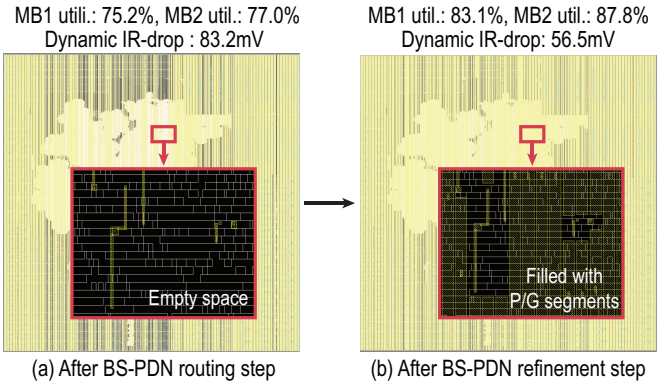


Fig. 5: Maximizing PDN utilization through the addition of P/G wire segments during the BS-PDN refinement step.

supplementary P/G segments, BS-PDN utilization is fully maximized, achieving reduced IR-drop, as demonstrated in Fig. 5-(b).

V. Results and Discussion

A. Experimental Setup

For the Gemmini SoC benchmark, we incorporate waveform-based analysis with the instance-wise switching activity throughout workload execution, whereas other benchmarks use common methods that assume uniform switching across all DFFs. Designs with backside integration leverage six frontside metal layers and two backside metal layers, along with buried power rails, across all three benchmarks.

The detailed methodology of waveform-based analysis is presented in Fig. 6. First, workload simulation using Synopsys VCS provides

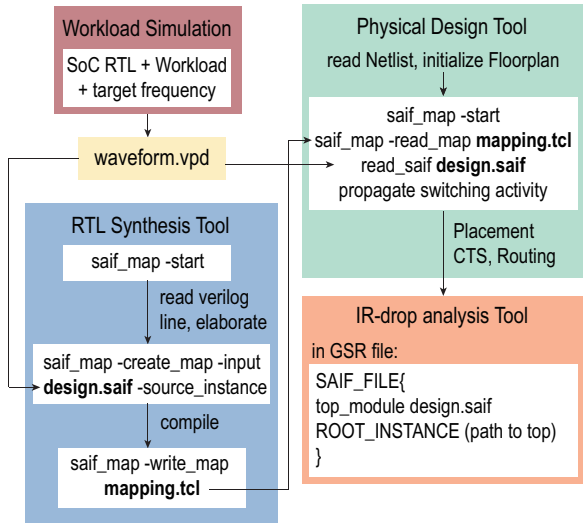


Fig. 6: Waveform-based analysis used in this study for Gemmini SoC benchmark. It enables a comprehensive evaluation of PDN-first and BS-PDN-last flows in real-world applications.

output in the Value Change Dump Plus (VPD) format, detailing value changes across each RTL component. However, component-wise VPD files can be extremely large; for instance, a 16x16 Gemmini SoC running ResNet-50 generates approximately 1TB of waveform data. To streamline power analysis, the waveform is converted into a Switching Activity Interchange Format (SAIF) file, summarizing average toggle counts across the duration of the workload.

Using the GTCAD 3nm PDK [9] at a supply voltage of 0.7V, Synopsys Design Compiler facilitates mapping between RTL instances in the SAIF file and gate-level cells in the synthesized netlist file by creating a naming correlation file. Using Synopsys ICC2, toggle counts from the SAIF file are then mapped into the corresponding cells, enabling precise vector-based power consumption estimates under workload conditions.

For IR-drop analysis, Ansys Redhawk incorporates SAIF files to yield a realistic IR-drop profile. This IR-drop information is then integrated into performance analysis through Synopsys Primetime, leveraging gate-level cell-specific voltage drops to improve the accuracy of timing assessments.

Using these comprehensive methodologies, workload runtime, energy consumption, and energy efficiency are calculated as shown below. Total cycles can be obtained from workload simulation. In this study, we use the inference of ResNet-50 workload with a batch size of 4 to analyze the impact of PDN-last flow on workload metrics.

$$\text{Runtime (ms)} = \frac{\text{Total Cycles}}{\text{Performance (GHz)} \times 10^6}$$

$$\text{Energy (mJ)} = \frac{\text{Vector-based Power (mW)} \times \text{Runtime (ms)}}{10^3}$$

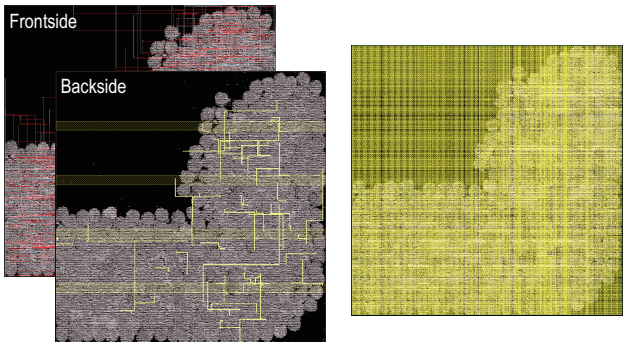
$$\text{Energy Eff. (TOPS/W)} = \frac{\text{Trillion Operations (TOP)}}{\text{Runtime (ms)} \times \text{V.-Power (mW)} \times 10^6}$$

B. Comparison between PDN-first and PDN-last Flow

Given the trade-off between IR-drop and performance in the PDN-first flow, it is important to select the PDN specifications carefully to enable a fair comparison between the PDN-first and PDN-last flows. In this work, we apply the same PDN utilization achieved from the PDN-last flow to ensure comparable IR-drop levels while demonstrating the performance advantages.

TABLE II: Iso-performance comparison between BS-PDN-first vs. PDN-last flow. BS-CDN is added in both cases (baseline: BS-PDN only.)

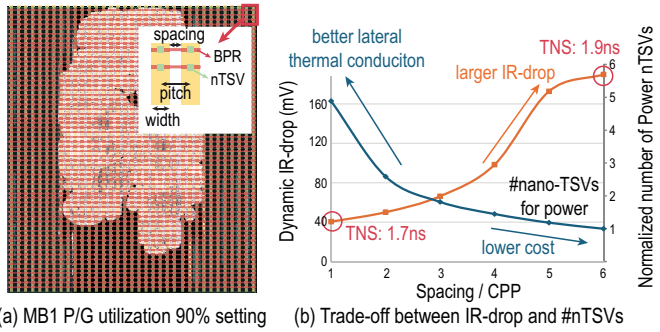
	AES (#cell=163K, 7.5GHz)			Rocket Chip (#cell=388K, 2.6GHz)			Gemmini SoC (#cell=1.17M, 1.5GHz)		
Backside PDN Backside CDN	PDN-first no	PDN-first yes	PDN-last yes	PDN-first no	PDN-first yes	PDN-last yes	PDN-first no	PDN-first yes	PDN-last yes
MB1 PDN util.	90%	80.6%	81.6%	90%	79.4%	79.1%	90%	82.7%	83.1%
MB2 PDN util.	90%	95.2%	95.8%	90%	85.6%	83.3%	90%	88.3%	87.8%
#Clock nTSV	-	127		-	306		-	217	
Max. disp. (μm)		2.94	0.41		3.61	0.42		37.4	0.43
WNS (ns)	26.0	13.2	9.2 (-65%)	46.8	36.7	21.9 (-53%)	192.6	235.85	135.3 (-30%)
TNS (ps)	5.1	0.66	0.47 (-90%)	7.5	6.7	1.7 (-77%)	35.3	203.96	24.4 (-31%)
Eff. freq (GHz)	6.27	6.83	7.02 (+12%)	2.32	2.37	2.46 (+6%)	1.16	1.11	1.25 (+8%)
Power (mW)	133.2	136.5	138.2	228.7	232.1	240.8	642.0	604.9	670.8
D. IR-drop (mV)	32.5	29.9	25.5	37.3	44.2	40.5	54.6	55.9	56.5



(a) Frontside and Backside clock routing with routing blockage on MB2 for BS-PDN planning

(b) Backside PDN structure with BS-PDN-last flow

Fig. 7: Final CDN and BS-PDN routing with BS-PDN-last flow on AES.



(a) MB1 P/G utilization 90% setting

(b) Trade-off between IR-drop and #nTSVs

Fig. 8: Impact of backside metal layer 1 (= MB1) P/G width & pitch on Rocket Chip.

Table II shows an iso-performance comparison across three configurations: BS-PDN only, BS-CDN with PDN-first flow, and BS-CDN with BS-PDN-last flow for each benchmark. As the BS-PDN-last flow defers PDN installation with minimal routing blockages, cell displacement during backside clock buffer conversion is significantly reduced compared to the PDN-first flow. Therefore, the BS-PDN-last flow successfully preserves the optimal logic cell and clock buffer positions, addressing the limitations of the PDN-first flow and achieving significant performance gains.

The results also indicate, as design size increases, legalizing placement for backside buffers within a PDN-first flow becomes significantly more challenging, leading to greater maximum displacements. This is because larger designs require stronger drive-strength buffers to meet clock specifications; however, these wider (=stronger) backside buffers are more difficult to legalize under a PDN-first approach. Consequently, the performance benefits gained from backside clock routing are further diminished. Given that the Gemmini SoC used in this study comprises 1.17 million cells—and recognizing that typical industrial designs are substantially larger—these observations highlight the critical importance of adopting a BS-PDN-last flow.

C. Impact of Backside Metal Width & Pitch

For a given PDN utilization, adjustments to P/G width and pitch are possible. For instance, a 90% target utilization can be maintained regardless of the absolute width and spacing values, as long as the width-to-spacing ratio remains 9:1. In this section, we examine the impact of varying the MB1 P/G spacing from 1x to 6x Contacted Poly Pitch (CPP) while maintaining a consistent 90% target utilization. Although Fig. 8-(a) illustrates a simplified layout with a regular BS-PDN arrangement, the actual design contains irregular pattern of P/G straps and segments similar to Fig. 7-(b). Under the current setting, MB1 utilization remained relatively stable, ranging from 79% to 81%, ensuring an equitable comparison across different spacing options.

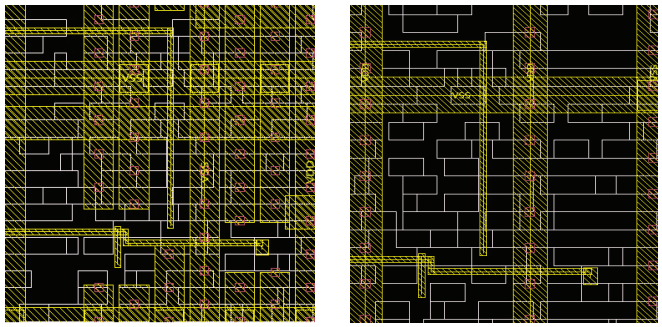
As shown in Fig. 8-(b), larger spacing increases the length of the high-resistance BPR paths, resulting in a larger dynamic IR-drop. As a result, timing metrics are negatively impacted, increasing the number of timing failure points from 428 to 480 and raising Total Negative Slack (TNS) from 1.7ns to 1.9ns. On the other hand, reducing the width and spacing of MB1 P/G straps increases the number of power nano-TSVs for MB1-BPR connection. Consequently, it improves lateral thermal conduction [6]; however, manufacturing costs increase. In summary, the MB1 width and pitch settings have a direct impact on the number of nano-TSVs and the power delivery path, creating a trade-off between IR-drop, TNS, thermal conduction, and cost.

D. Impact of Spacing between Backside Power and Clock Wires

The application of BS-CDN necessitates the PDN to share metal layers with clock routing, which introduces potential signal integrity issues due to coupling capacitance. To evaluate this, we compute the coupling capacitance by comparing the total capacitance of the backside net before and after PDN implementation. The spacing between P/G lines and clock/signal routing are varied from 1x-20x and 40x of the minimum pitch, respectively. According to Table III, increasing the spacing reduces total coupling capacitance by 2.9% and 3.1%, respectively. However, with increased spacing, as illustrated in Figure 9, the length and number of P/G straps decreased. Therefore, substantial spacing can hinder the installation of P/G straps, which

TABLE III: Effect of spacing between backside P/G and backside clock routing. The spacing in the x and y directions is set as a multiple of the minimum pitch for the MB1 and MB2 layers, respectively.

Spacing	1x	20x	40x
Total coupling cap. (fF)	49.4	47.9 (-2.9%)	47.8 (-3.1%)
Average PDN util. (%)	83.4	83.1	82.9
Dynamic IR-drop (mV)	37.3	38.2 (+2.4%)	39.9 (+7.0%)



(a) spacing = 1x min_pitch (b) spacing = 40x min_pitch

Fig. 9: Different spacing between power and clock net on Rocket Chip.

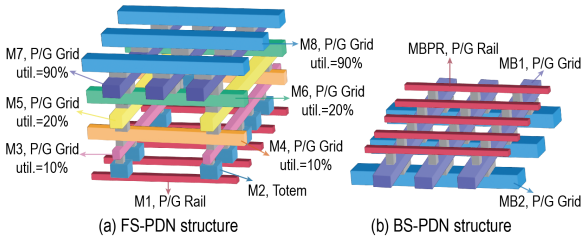
exacerbates IR-drop. Moreover, increasing P/G strap spacing beyond a specific threshold incurs minimal impact on coupling capacitance. Thus, determining an optimal spacing level is crucial to balance coupling capacitance minimization and efficient P/G planning.

E. Workload Simulation Results

This section presents a comparative analysis of FS-PDN, BS-PDN, and BS-CDN + BS-PDN-last flow on the Gemmini SoC, evaluating both physical design-related and workload-related metrics. Designs utilizing backside metal layers implement power grids on MB1 and MB2 with buried power rails, whereas FS-PDN designs construct power grids from M3 to M8 with distinct utilization, with power rails located on M1 as shown in Fig. 10-(a). The FS-PDN specification was configured to approach a worst-case dynamic IR-drop as close to 10% of the supply voltage as possible, while avoiding significant interference with frontside routing.

The GTCAD 3nm PDK [9] used in this study provides cells with consistent height across both BPR-enabled and BPR-free configurations, resulting in an equivalent footprint for all designs. Rather than achieving footprint savings as seen in conventional FS-PDN vs. BS-PDN analysis [7], our BS-PDN designs achieve power benefits due to reduced parasitic capacitance at cell pins as well as smaller load capacitance from better frontside routability.

As shown in Table IV, the design integrating BS-CDN with the BS-PDN-last flow achieves a 16.1% enhancement in performance, along with a 13.7% reduction in ResNet-50 runtime, a 17.4% improvement in energy consumption, and an 18.9% increase in energy efficiency. Additionally, BS-CDN with BS-PDN-last flow demonstrates substantial gains over the design with BS-PDN alone, especially in terms of performance and runtime, by 7.8% and 7.3% respectively. Given the relatively low cost of extending existing backside metal layers from single-function BS-PDN to multi-functional BS-PDN + CDN designs—compared to the higher cost of converting a frontside-only design to a backside-integrated —the importance of leveraging backside metal layers for diverse functionalities becomes evident.



(a) FS-PDN structure (b) BS-PDN structure

Fig. 10: PDN structure used in this work: the baseline FS-PDN uses P/G rails on M1 and grids on M3–M8, while the BS-PDN is implemented on the BPR and backside.

TABLE IV: Comparison between FS-PDN, BS-PDN, BS-CDN + PDN-last with 3nm PDK (supply voltage=0.7V). Workload metrics (runtime, energy, E-eff) are based on ResNet-50 with a batch size of 4.

Gemmini SoC	FS-PDN	BS-PDN	BS-PDN + CDN
PDN flow	PDN-first	PDN-first	BS-PDN-last
PDN pattern	regular	regular	irregular
PDN util. (%)	based on Fig. 10-(a)	MB1: 90 MB2: 90	MB1: 83.1 MB2: 87.8
# metals (f+b)	8+0	6+2	6+2
Footprint	$452\mu m \times 452\mu m$		
Vector-based Power (mW)	690.4 -	642.0 (-7.0%)	670.8 (-2.8%)
Perf. (GHz)	1.08	1.16 (+7.7%)	1.25 (+16.1%)
D. IR-drop (mV)	89.8	54.6	56.5
Runtime (ms)	115.5	107.5 (-6.9%)	99.7 (-13.7%)
Energy (mJ)	79.7	69.0 (-13.4%)	65.8 (-17.4%)
Energy-eff. (TOPS/W)	0.185 -	0.214 (+15.7%)	0.220 (+18.9%)

VI. Conclusion and Future Work

We present a BS-PDN-last flow that enhances performance and energy efficiency while maintaining IR-drop levels in multifunctional backside metal layer applications. By deferring PDN routing to the final design phase, this approach achieves up to a 90% reduction in Total Negative Slack and up to a 12% increase in performance when integrated with BS-CDN, effectively eliminating the trade-off between IR-drop and performance gain inherent to PDN-first flows. Notably, the benefits of the BS-PDN-last flow scale significantly with increasing design complexity, highlighting its relevance for industrial-scale chip implementations.

While our current focus is on backside clock and power routing, the proposed approach is readily extendable to backside signal routing (BSS). Since BSS involves relocating critical frontside signal nets to the backside, it necessitates the use of nano-TSVs and the MB1 layer, resulting in a competitive allocation with the BS-PDN—similar to the conflict observed with BS-CDN. Especially, the MB1 routing blockages introduced during our BS-PDN planning stage are effective in mitigating issues associated with memory-to-register frontside net conversions to backside.

Acknowledgment

This work was supported by the Semiconductor Research Corporation under the JUMP 2.0 Center Program (CHIMES), Samsung Advanced Institute of Technology (SAIT) under the AI for Semiconductors Program, Samsung Semiconductor Logic Path Finding Lab, and the Ministry of Trade, Industry & Energy of South Korea (1415187652, RS-2023-00234159).

REFERENCES

[1] K. Asanovic, R. Avizienis, J. Bachrach, S. Beamer, D. Biancolin, C. Celio, H. Cook, D. Dabbelt, J. Hauser, A. Izraelevitz, et al. The Rocket Chip Generator. *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2016-17*, 4:6–2, 2016.

[2] N. E. Bethur, P. Vanna-Iampikul, O. Zografos, L. Zhu, G. Sisto, D. Milojevic, A. García-Ortiz, G. Hellings, J. Ryckaert, F. Catthoor, et al. GNN-assisted Back-side Clock Routing Methodology for Advance Technologies. In *61st ACM/IEEE Design Automation Conference (DAC)*. ACM, 2024.

[3] H. Genc, S. Kim, A. Amid, A. Haj-Ali, V. Iyer, P. Prakash, J. Zhao, D. Grubb, H. Liew, H. Mao, et al. Gemmini: Enabling Systematic Deep-Learning Architecture Evaluation via Full-Stack Integration. In *58th ACM/IEEE Design Automation Conference (DAC)*, pages 769–774. IEEE, 2021.

[4] A. Jourdain, M. Stucchi, G. Van der Plas, G. Beyer, and E. Beyne. Buried Power Rails and Nano-Scale TSV: Technology Boosters for Backside Power Delivery Network and 3D Heterogeneous Integration. In *IEEE 72nd Electronic Components and Technology Conference (ECTC)*, pages 1531–1538. IEEE, 2022.

[5] B.-S. Kim, S. Choi, J. H. Lee, K. Lee, J. Park, J. Kwon, S. Park, K. Chun, H. Simka, A. Kumar, et al. Expanding Design Technology Co-Optimization Potentials with Back-Side Interconnect Innovation. In *IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)*, pages 1–2. IEEE, 2024.

[6] H. Oprins, J. L. R. Bohorquez, B. Vermeersch, G. Van der Plas, and E. Beyne. Package level thermal analysis of backside power delivery network (BS-PDN) configurations. In *21st IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (iTherm)*, pages 1–7. IEEE, 2022.

[7] D. Prasad, S. T. Nibhanupudi, S. Das, O. Zografos, B. Chehab, S. Sarkar, R. Baert, A. Robinson, A. Gupta, A. Spessot, et al. Buried Power Rails and Back-side Power Grids: Arm® CPU Power Delivery Network Design Beyond 5nm. In *IEEE International Electron Devices Meeting (IEDM)*, pages 19–1. IEEE, 2019.

[8] P. Vanna-iampikul, H. Yang, J. Kwak, J. X. Hu, A. Rahman, N. E. Bethur, C. Hao, S. Yu, and S. K. Lim. A Design Methodology for Back-side Power and Clock Routing Co-Optimization (Invited). In *IEEE Symposium on VLSI Technology Circuits (VLSI Technology and Circuits)*. IEEE, 2024.

[9] J.-S. Yoon, J. Jeong, S. Lee, J. Lee, S. Lee, R.-H. Baek, and S. K. Lim. Performance, Power, and Area of Standard Cells in Sub 3 nm Node Using Buried Power Rail. *IEEE Transactions on Electron Devices (T-ED)*, 69(3):894–899, 2022.