

Enhanced Power Delivery Pathfinding for Emerging 3-D Integration Technology

Andrew B. Kahng, *Fellow, IEEE*, Seokhyeong Kang^{ID}, *Member, IEEE*,

Seungwon Kim, *Member, IEEE*, and Bangqi Xu^{ID}, *Graduate Student Member, IEEE*

Abstract—In advanced technology nodes, emerging 3-D integration technology is a promising “More Than Moore” lever for continued scaling of system capability and value. In the 3-D integrated circuit (3-D IC) implementation, the power delivery network (PDN) is crucial to meeting design specifications. However, determining the optimal PDN design is nontrivial. On the one hand, to meet the voltage (IR) drop requirement, a denser power mesh is desired. On the other hand, to meet the timing requirement, more routing resource is needed for signal routing. Moreover, additional competition between signal routing and power routing is caused by intertier vertical interconnects in 3-D IC. In this article, we propose a power delivery pathfinding methodology for emerging 3-D integration, which seeks to identify a “near-optimal” (or, very high quality) PDN for a given BEOL stack, vertical interconnection, and PDN specification. Compared with previous works, our methodology can explore richer solution spaces as it supports different PDN layer combinations and PDN layer configurations. We develop models for routability and worst IR drop to help reduce iterations between PDN design and circuit design in 3-D IC implementation. We present validations and demonstrate improvement in IR drop and routability with real design blocks in 28- and 14-nm foundry technology nodes.

Index Terms—3-D integration, voltage (IR) drop prediction, power delivery, routability analysis, system pathfinding.

I. INTRODUCTION

MODERN very-large-scale integration technology has enabled higher system performance and efficient power management based on advanced transistor technology.

Manuscript received August 1, 2020; revised October 26, 2020; accepted November 15, 2020. Date of publication December 21, 2020; date of current version April 1, 2021. This work was supported in part by Qualcomm, in part by Samsung, in part by NXP Semiconductors, in part by Mentor Graphics, in part by Defense Advanced Research Projects Agency (DARPA) under Grant HR0011-18-2-0032, in part by NSF under Grant CCF-1564302, and in part by the C-DEN Center. (Seungwon Kim and Bangqi Xu contributed equally to this work.) (Corresponding author: Bangqi Xu.)

Andrew B. Kahng is with the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093 USA, and also with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: abk@ucsd.edu).

Seokhyeong Kang is with the Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang 37673, South Korea (e-mail: shkang@postech.ac.kr).

Seungwon Kim is with the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: sek006@ucsd.edu).

Bangqi Xu is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: bangqixu@ucsd.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TVLSI.2020.3041665>.

Digital Object Identifier 10.1109/TVLSI.2020.3041665

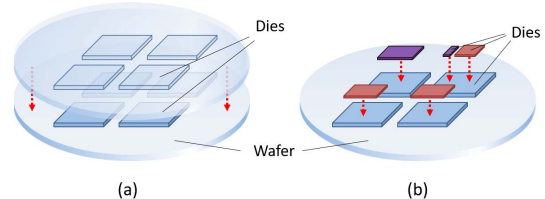


Fig. 1. Two integration technologies for foundry-driven 3-D IC. (a) WoW integration. (b) D2W integration.

With foundry 7-nm products reaching high-volume production, only a few feasible technology nodes remain to potentially deliver power, performance, area, and cost (PPAC) benefits from the transistor, cell architecture, and lateral scaling. 3-D integrated circuit (3-D IC) stacking techniques are receiving attention as a promising solution to continue Moore’s Law for future scaling of integration, area footprint, and design performance/power envelope.

3-D IC stacking technologies have historically been driven from two directions: the packaging industry and the foundry industry. Conventional packaging-driven 3-D IC technologies based on through-silicon vias (TSVs) have limitations to the high vertical integration density at the die level due to the size and pitch of the TSV structure [1]. Recent advanced intertier vertical interconnect (VI) technology has led to the emergence of multiple foundry-driven 3-D IC technologies to achieve significant PPAC benefits; these include high-precision face-to-face (F2F) wafer-on-wafer (WoW) and die-to-wafer (D2W) stacking, as shown in Fig. 1 [13], [34]. WoW technology focuses more on power, performance, and area improvements, while D2W technology seeks more cost-effective integration methods to enhance system-level power and performance, e.g., for memory-on-logic, single-chip solutions. WoW faces two key limitations compared with D2W technology: 1) the same area constraint for top and bottom dies limits partitioning scenarios and 2) overall low yield. On the other hand, D2W technology has achieved yield improvements with prebonding testing for existing 2-D IPs. Therefore, existing 2-D EDA tools can be used to perform realistic experiments for D2W stacking. In addition, D2W facilitates the integration of a heterogeneous 3-D IC into multiple dies (e.g., a large bottom die and variously sized smaller top dies). With this process-friendly approach coupled with relatively high integration density, D2W technology has become a practical solution to cope with 2-D scaling challenges.

A power delivery network (PDN) in the back end of line (BEOL) has a direct impact on the reliability and functionality of the product design. Determining high-quality PDNs with the increasing power density and design complexity is challenging even in 2-D ICs. The challenges are exacerbated

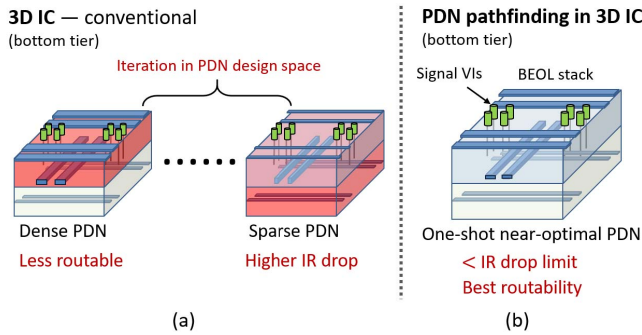


Fig. 2. Brief example of PDN design from the lowest layer to the top layer in BEOL. (a) In the 3-D IC, when using an optimal PDN structure of 2-D IC, routing congestion occurs due to lower routability in the higher metal layers. A design that is too pessimistic about routing congestion uses less PDN and, thus, has a large IR drop. (b) Our proposed “one-shot” pathfinding methodology obtains near-optimal PDN parameters (i.e., yielding best routability while meeting the IR drop requirement) with a one-time effort for each tier.

in 3-D ICs with additional resistance between the power supply and transistors in different tiers. In addition, feasible design solutions are limited because signal and power/ground routing must be passed through the intertier VIs. Smaller sizes of intertier VIs can make new integration feasible, but the higher resistance adversely affects the PDN quality [36], [37]. To achieve robust functionality, 3-D IC designs must mitigate and balance these PDN-related challenges. Fig. 2(a) illustrates a conventional PDN design flow for 3-D IC, where designers iteratively explore a large PDN design space to reach a balance between voltage (IR) drop and routability. However, this process takes a considerable portion of the design cycle. To reduce the turnaround time in 3-D IC design, ideally, a PDN pathfinding flow in 3-D IC would be capable of delivering a near-optimal PDN design without any iterations as illustrated in Fig. 2(b). This demands an efficient, accurate design space exploration (also known as pathfinding) methodology that—given various technology- and design-dependent parameters—can quickly provide quality of result (QoR) tradeoffs of various PDN solutions.

Our work builds on the power delivery pathfinding approach of [16], in which a fixed combination of BEOL layers is used for power delivery. Compared with [16], we propose an efficient two-stage pathfinding methodology for PDN design of emerging F2F 3-D designs that explore a larger design solution space, including various PDN layer combinations. In the literature and industry, finding good PDN designs is difficult largely due to the large solution space of PDN designs and the long turnaround time for PDN evaluation. With our solutions, near-optimal PDN designs can be found with models based on a relatively small set of data points from small artificial designs. In the first stage, we create sensitivity graphs that consider the IR drop and routability tradeoff and apply the shortest path algorithm to obtain a PDN layer combination with minimized cost. In the second stage, based on the layer combination obtained from the first stage, we build an IR drop model to predict the worst IR (WIR) drop of a given PDN configuration. To comprehend the effect of a given PDN solution on the overall design QoR, we also develop a routability model that predicts the routability of a design given a PDN configuration.

Putting the first- and second-stage elements together, our pathfinding methodology starts by identifying the best PDN

layer combination considering both WIR and routability. For the best PDN layer combination, we then filter out PDN configurations based on a given design’s prescribed WIR limits. Finally, the routability model is used to identify the WIR-feasible PDN configuration(s) that offer the best routability. We, thus, obtain a high-quality PDN solution that is “near-optimal” (to the extent that we have been able to make exhaustive enumeration-based experimental confirmations) in the sense of both predicted WIR and estimated routability within our modeled PDN design space. Our PDN solution approach offers direct benefits to design QoR and ease of implementation. The main contributions of our work are as follows.

- 1) We propose a novel interface to properly combine IR drop analysis of PDN configurations and the corresponding impact on routability.¹
- 2) We study the impact of VI density (VI_{density}) on design routability and build a VI-aware routability model.
- 3) We develop IR-drop and routability sensitivity graphs to obtain a best PDN layer combination given a set of IR-drop and routability weights.
- 4) On foundry 28-nm designs, we demonstrate that our pathfinding methodology identifies high-quality PDN designs compared with a reference industry PDN design.
- 5) We further confirm that our pathfinding methodology improves over an industrial PDN reference solution in foundry 14-nm technology.
- 6) To the best of our knowledge, we are the first to propose a pathfinding methodology that explores both PDN layer combination and per-layer PDN configuration to identify high-quality solutions for F2F 3-D designs.

The remainder of this article is organized as follows. Section II provides an overview of related works in the literature. Section III introduces our PDN pathfinding methodology. Section IV describes our PDN layer combination pathfinding (Stage 1), and Section V shows the experimental setup and results for this PDN layer combination pathfinding. Sections VI and VII, respectively, describe our PDN layer configuration pathfinding (Stage 2) and experimental results. We validate our overall two-stage PDN pathfinding flow in Section VIII. In Section IX, we extend our proposed methodology to 14-nm foundry technology. Section X gives conclusions and directions for ongoing works.

II. RELATED WORKS

In this section, we review previous works in the literature. We classify relevant previous works into three categories: 1) 3-D IC implementation methodology; 2) PDN design methodology; and 3) routability modeling.

A. 3-D IC Design Implementation Methodology

Several design methodologies using existing commercial 2-D CAD tools have been proposed for physical implementation of gate-level 3-D ICs [8], [19], [20], [22], and [24]. The Shrunk2D (S2D) flow [22], [24] performs gate-level 3-D IC implementation, while the subsequent Cascade2D flow

¹Note that we do not attempt PDN pathfinding that considers dynamic IR droop. This remains a “holy grail” that depends on the evolution of techniques, such as what we propose here and determining proper simulation vectors. Also, while our work aligns with goals, such as PDN pathfinding for large SoC designs, we focus on the 3-D IC context and its unique complexities, at a block scale (and we do not have access to large SoC designs and collateral data).

implements both gate- and block-level monolithic 3-D ICs [8]. Recently, a commercial-quality F2F-bonded 3-D IC implementation flow Compact-2D (C2D) has been proposed [20]. We note that these works on 3-D IC implementation leave open the issue of interactions between power delivery and routability.

B. PDN Design Methodology

Power delivery in gate-level 3-D ICs is considered in [23], which proposes a PDN-centered tier-partitioning technique that comprehends the IR drop versus thermal tradeoff in monolithic 3-D IC. Samal *et al.* [29] analyze full-chip impact of PDN designs in monolithic 3-D ICs. Optimized 3-D PDN design configurations (in six categories) are compared across power, performance, IR drop, and wirelength metrics in different technology nodes. However, design-specific PDN choices at the “Pareto frontier” of IR drop versus routability are not addressed, as this would require exploration of PDN structures with degrees of freedom on each metal layer. Chang *et al.* [9] develop a system-level PDN model, along with static and dynamic frequency- and time-domain analyses. The 2-D and 3-D ICs with extracted equivalent *RLC* parasitics are compared using a single-baseline PDN structure. The focus is on dynamic rail analysis with frequency-related environmental differences (e.g., decap insertion) rather than PDN optimization. Chang *et al.* [7], Chhabria *et al.* [12], and Kahng *et al.* [16] propose model-based power delivery pathfinding methodology that explores PDN design solution space for a given, fixed PDN layer combination (e.g., M2, M3, M4, M7, and M8). With these approaches, the exploration of different PDN layer combinations would require much more data for modeling and may encounter scalability challenges. Furthermore, vertical connections are not considered as these works are limited to the 2-D IC context.

C. Routability Modeling

Numerous techniques have been devised toward estimation of signal routing congestion in placement and global routing stages [11], [18], [21], [31]. Various methods, respectively, apply Rent’s rule to estimate the wirelength distribution of a region [32], estimate congestion by analysis of pin densities [3] or Steiner trees [28], or achieve bounding box-aware per-net wirelength estimation [4]. Machine learning-based routing congestion prediction models have also been proposed. Qi *et al.* [27] apply supervised learning to predict detailed routing violation and utilization in the global routing stage via multivariate adaptive regression splines (MARS). Zhou *et al.* [33] propose a machine learning model that predicts DRC violations from placement and global routing data. Chan *et al.* [5] extract hotspot features to identify gcells with DRC violations.

This work requires routability estimation that considers both BEOL stacks (including PDN and technology rules) along with given placement. Thus, in the following, we employ the routability characterization methodology of “PROBE” [15], which affords a ranking of BEOL stack options according to their intrinsic routing capacities.²

² Kahng *et al.* [15] note several challenges associated with studies of real design blocks, which can have: 1) large cell instance counts; 2) large variance in cell sizes; and 3) nonuniformity in net topologies. Our PDN pathfinding problem has these same challenges and additional impacts from TSV and VI connections. However, as noted, we confirm the robustness of our proposed methodology with real design blocks in two different foundry technologies.

Additional works have studied the issue of vertical cuts (interconnect demands) in gate-level 3-D IC implementation, e.g., attempting to maximize the benefits of 3-D ICs by increasing the number of monolithic intertier vias (MIVs) or F2F VIs [20], [22]. Peng *et al.* [26] note that, as the number of vertical cuts increases, interdie coupling capacitance increases, significantly affecting power and signal integrity in F2F bonded ICs.

D. Summary

From the above, we see that, while previous works on 3-D IC implementation have illuminated many aspects of partitioning, place-and-route, and power delivery, typically, only a very limited PDN solution space is considered. The need for PDN pathfinding in 3-D IC arises because power/ground (PG) delivery is far from “free”: in 3-D ICs, there are TSV and routability impacts, as well as a need for the PDN solution to support the delivery of PG and signal through intertier VIs. The number of VIs is a significant determinant of power and signal integrity, in light of routing congestion and IR drop. This is in contrast to PDNs in 2-D ICs that are generally less sensitive to signal routing congestion on upper metal layers.³ Our work attempts to close this gap by explicitly considering both IR drop and routability.

III. METHODOLOGY

In this section, we first generalize the PDN pathfinding problem presented in [16] and then describe our approach for PDN pathfinding methodology considering both PDN layer combination and PDN layer configuration.

- 1) *PDN Pathfinding Problem*: Given a mesh-like placement with VI locations, provide an optimized PDN design considering IR and routability.
- 2) *Inputs*: Mesh-like placement, VI locations, and BEOL stack.
- 3) *Output*: PDN with optimized IR and routability.
- 4) *Constraints*: Technology design rules.

A. Preliminaries

We divide the overall PDN pathfinding problem into two subproblems that are sequentially solved in the proposed two-stage PDN pathfinding methodology. We define the two stages for solving the two subproblems as follows.

- 1) PDN layer combination pathfinding focuses on the choices of metal layers in a BEOL stack that is used for PG metal stripes (e.g., M2-M3-M4-M7-M8).
- 2) PDN layer configuration pathfinding focuses on the detailed usage of routing resources (i.e., PDN density) of each metal layer for a given PDN layer combination (e.g., a set of PG stripe {width, spacing, pitch} configurations for all metal layers in a PDN layer combination).

For PDN layer combination pathfinding, in order to qualitatively provide guidance on the detailed usage of routing resources, we define a **usage corner** as a tuple of {width, spacing, pitch}.⁴ For PDN layer configuration pathfinding,

³If the total number of VIs is high relative to the total number of nets (i.e., a high #VIs-to-#nets ratio), this implies that the number of 3-D nets traversing through the VIs located on the top metal layer is also relatively high. The impact of these VIs (which are induced by the design’s partition across tiers) must be considered in the 3-D IC PDN design.

⁴In Section IV, we define three typical usage corners (i.e., min, base, and max). A finer granularity of usage corners can provide more detailed guidance for PDN layer configuration pathfinding.

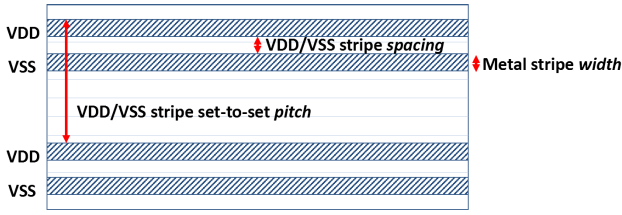


Fig. 3. Illustration of circuit design-independent PDN design knobs.

TABLE I
PDN DESIGN KNOBS

Circuit design-independent knobs	
Metal stripe width (w)	Width of PDN stripe for each layer
VDD/VSS stripe set-to-set pitch size (p)	Set-to-set distance of VDD/VSS PDN stripe for each layer
VDD/VSS stripe spacing (s)	Spacing between VDD/VSS PDN stripe for each layer
Circuit design-dependent knobs	
#Instances	Instances in one tier of 3D IC
Utilization	Row utilization of circuit
$VI_{density}$	Number of VIs / Number of nets

we define PDN design knobs to explore the PDN layer configuration solution space. Table I shows the PDN design knobs that we consider. Circuit design-independent knobs include width, space, and pitch size of metal stripe, as shown in Fig. 3. Combinations of these knobs must satisfy the design rule constraints of the given technology. For a given 3-D IC design, the circuit design-dependent knobs include the number of cell instances, row utilization, and $VI_{density}$.

B. Routability Measurement

While WIR can be directly measured by a power analysis tool, measurement of routability is less straightforward. In the context of 3-D IC, with consideration of TSV and VI effects, routability measurement becomes even more challenging. We apply the core technique of PROBE [15] to obtain an intrinsic measure of routability of a given PDN design, in terms of the so-called “ K threshold” (K_{th}) metric, based on a mesh-like placement. We construct a mesh-like placement, as illustrated in Fig. 4(a). A mesh-like placement uniformly arranges netlists in a $W_{die} \times H_{die}$ floorplan, where $H_{die} = M_r \times H_{gate}$ and $W_{die} = M_c \times W_{gate}/U$ according to the mesh topology. M_r is the number of rows indexed by p , and M_c is the number of columns indexed by q . W_{gate} and H_{gate} are the width and height of a given cell, respectively, and U is a predefined placement (row) utilization. For a mesh-like placement, as shown in Fig. 4(a), each pin of a given instance is initially connected to its neighboring instances, and originally, there are zero (or very few) design rule violations after routing.

The PROBE methodology iteratively swaps the placement locations of random pairs of neighboring instances. This progressive “tangling” gradually degrades the placement, increasing congestion until, eventually, the perturbed placement becomes unroutable (i.e., the number of postroute design rule violations exceeds a predefined threshold).⁵ The number of neighbor swaps (normalized to total instance count) before routing failure occurs is called the K threshold (K_{th}). We use a three-input AOI cell as the basis for the starting mesh-like placement, with inputs and output of each cell being connected, as shown in Fig. 4.⁶

⁵Following [15], we define routing failure as $\#DRVs > 150$.

⁶Note that, for training data collection, the mesh-like placement enables the fine-grained increase of routing difficulty with decent runtime scalability compared with placement perturbation and routing for real design blocks.

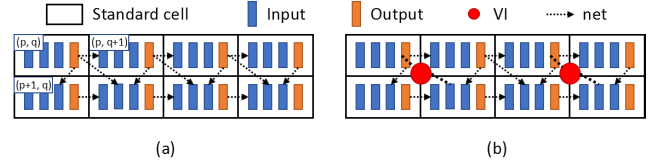


Fig. 4. Illustration of (a) mesh-like placement as in [15] and (b) our 3-D mesh-like placement with VIs (on top metal layer).

A robust PDN design can potentially consume considerable routing resources as it satisfies the WIR constraint; this, in turn, worsens routing congestion in surrounding areas. When comparing PDN designs, a higher K_{th} value implies that a given PDN design has better routability, i.e., more routing capacity. According to [15], the rank ordering of BEOL routing capacities, based on mesh-like placement, is stable and consistent across different designs. In this work, we make the key observation that different PDN designs implemented in a BEOL stack are equivalent to variants of the original BEOL stack with reduced routing capacities. Following this observation, we can measure the routability of a PDN for a given BEOL stack, and the rank ordering of routabilities across PDNs can be generalized and applied to different designs. To understand the impact of VIs on routability of a given PDN in the 3-D IC context, we extend the mesh-like placement with connections from cell pins to VI pins on the top metal layer, as shown in Fig. 4(b). We fix the locations of VI pins during the random swapping of neighboring cells. We use the parameter $VI_{density}$ (see Table I) to reflect that the impact of VI on relative routability (of a given BEOL stack plus PDN) is independent of the design size. The number of VIs is determined by $VI_{density} \times \#nets$. The VIs are placed on the top metal, and VIs do not overlap the PDN.⁷ Each VI is connected to the net of the nearest cell output pin.

C. F2F Mesh-Like Placement Setup

Unlike the 2-D mesh-like placement used in PROBE [15], routability measurement in 3-D IC must comprehend the unique aspects of vertical interconnection, including TSVs and intertier connections. Therefore, in our setup of mesh-like placement for the F2F case, we place TSVs on the bottom tier as both placement and routing blockages. Following the methodology of such works, as in [22], we place I/O ports on the top routing layer to capture the behavior of VIs for intertier connections. Fig. 5(a) illustrates the cross section view of F2F stacking with TSV, and Fig. 5(b) illustrates our implementation for routability and WIR experiments. In order to simulate the 3-D F2F stacking structure with available 2-D EDA tools, we replace the connections between PDN TSVs and top-layer PDN of the top tier with fictitious metal stripes and vias with very low resistance, such that the top-layer PDNs from both tiers are virtually shorted.⁸

Fig. 6(a) illustrates the top view of F2F stacking with TSV, and Fig. 6(b) illustrates our implementation for F2F mesh-like placement with PDN. We use a staggered TSV allocation with

⁷Note that, to implement routing by a commercial 2-D P&R tool in our experiments, the VIs in the routability model are placed as I/O pins. This technique has been used in previous works, such as [22].

⁸We recognize that the setup in Fig. 5(a) is not identical to the setup in Fig. 5(b), and that different current distributions will result. However, our separate studies confirm that such differences are small, do not affect the dominance of bottom-tier WIR, and do not qualitatively change our conclusions.

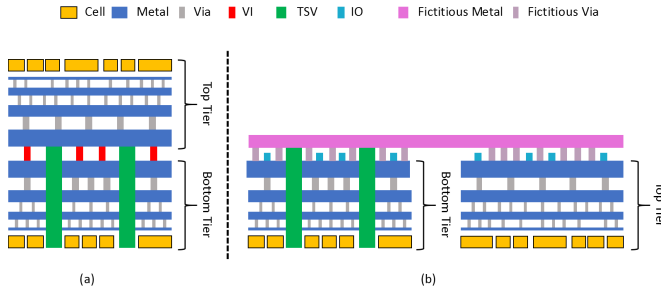


Fig. 5. Cross section view illustrations of (a) F2F stacking with TSV and (b) our experimental implementation. Physical connections from PDN TSVs to top-tier PDN are replaced by fictitious metal stripes and vias with very low resistance such that the top-most PDN layers from both tiers are virtually shorted. VIs are replaced with I/O ports to mimic intertier connection.

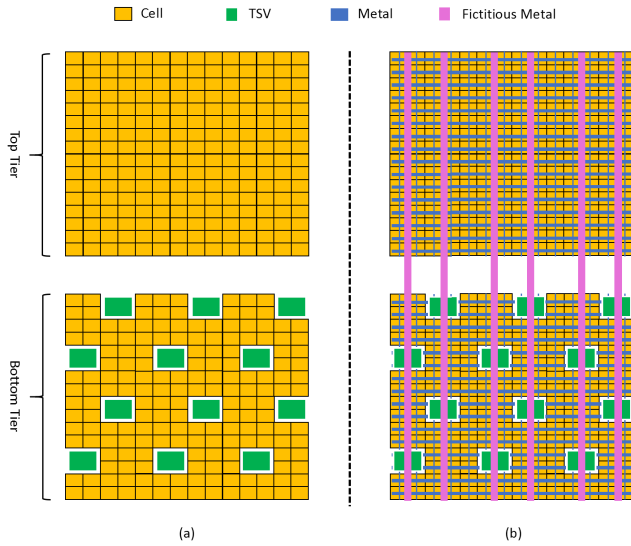


Fig. 6. Top view illustrations of (a) mesh-like placement with TSVs and (b) mesh-like placement with PDN overlay.

TSV size of $2.4 \times 2.4 \mu\text{m}^2$, an array pitch size of $40 \mu\text{m}$, and offset between VDD and VSS of $10 \mu\text{m}$ in this work. We perform routability and WIR experiments with the implementation described in Fig. 5(b) where both the routability and WIR characteristics of the design are preserved compared with the F2F stacking case. In the following, the routability studies are performed considering the TSVs unless otherwise specified.

D. Overall Flow

As mentioned in Section II, previous works have mainly focused on exploring PDN design solution space when the PDN layer combination is fixed due to scalability limitations. To achieve a power delivery pathfinding flow that explores a solution space, including both PDN layer combination and PDN layer configuration, we propose a two-stage PDN pathfinding methodology.

Fig. 7 illustrates the two stages of our methodology. In Stage 1, we formulate the PDN layer combination pathfinding problem as a shortest-path computation in a sensitivity graph, which determines the best layer combination. We introduce the routability-IR tradeoff factor α to modulate the balance of routing resource usage between PDN and signal routing. In Stage 2, based on the layer combination obtained from Stage 1, we develop and apply WIR and routability

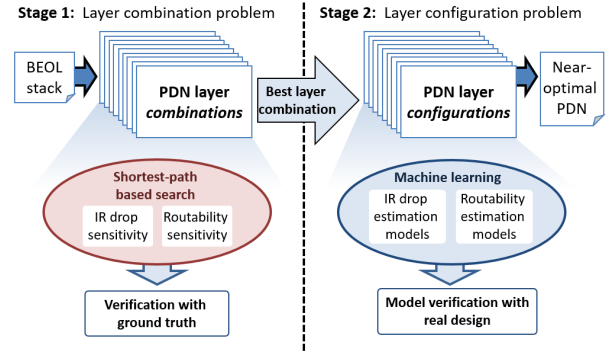


Fig. 7. Two-stage PDN pathfinding flow that gives the optimized PDN layer combination and per-layer PDN configuration considering both WIR requirement and routability requirement.

models to filter and rank possible PDN layer configurations, so as to obtain the most promising PDN design for given BEOL stack and WIR requirements. We validate our two-stage PDN pathfinding flow in Section VIII.

IV. STAGE 1: PDN LAYER COMBINATION PATHFINDING

In this section, we describe the problem statement and our shortest-path-based formulation for PDN layer combination pathfinding problem. As mentioned earlier, there is an obvious tradeoff between PDN quality and routability. Therefore, we introduce a routability-IR tradeoff factor α to modulate the balance between routability and WIR metrics. The following equation shows the cost function that we use in PDN layer combination pathfinding: a weighted sum of routability (i.e., K_{th}) and WIR (i.e., mV):

$$\text{cost} = \alpha \cdot \text{cost}_{K_{th}} + (1 - \alpha) \cdot \text{cost}_{IR}. \quad (1)$$

We validate our shortest-path-based PDN layer combination pathfinding flow in Section V-B.

- 1) *PDN Layer Combination Pathfinding Problem:* For a mesh-like placement, a BEOL stack, and routability-IR tradeoff factor, find the PDN layer combination that gives the minimum cost.
- 2) *Inputs:* Mesh-like placement, baseline PDN design, routability-IR tradeoff factor α , and PDN resource usage corners.
- 3) *Output:* PDN layer combination that gives the minimum cost along with resource usage guidance for each metal layer.

For a given BEOL stack with n metal layers and m PDN resource usage corners, our goal is to find the PDN layer combination that gives the minimum weighted sum of routability cost and IR cost from a sensitivity graph. The sensitivity graph, for which we require $\mathcal{O}(m \cdot n)$ experiments to obtain the sensitivity (i.e., cost) of each edge, is able to predict an overall solution space that requires $\mathcal{O}(m^n)$ experiments to obtain ground-truth data.

A. Sensitivity Graph

Similar in spirit to [2], we explore the possibility of leveraging a superposition assumption (i.e., empirical property) for each of routability and WIR of PDN for a given BEOL stack. Without loss of generality, we use Fig. 8 to illustrate the process of sensitivity graph construction with a BEOL stack with eight metal layers and three PDN resource usage corners.

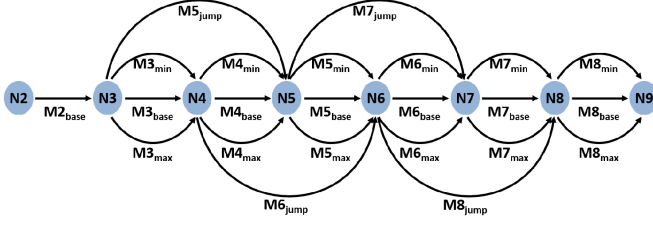


Fig. 8. Sensitivity graph of a BEOL stack of eight metal layers for PDN.

Each vertex in the sensitivity graph only serves for connectivity purposes and does not have physical meaning. Each edge in the sensitivity graph represents a usage corner (i.e., min, base, or max) of a metal layer if the corresponding metal layer is used for PDN. For example, $M3_{\min}$ indicates minimum routing resource usage for PDN on M3, and $M5_{\text{jump}}$ indicates that PDN does not use M3 or M4. Each directed path from node $N2$ to node $N9$ represents a valid PDN. For example, the path consisting of edges $\{M2_{\text{base}} - M3_{\text{base}} - M4_{\text{base}} - M5_{\text{base}} - M6_{\text{base}} - M7_{\text{base}} - M8_{\text{base}}\}$ represents our baseline PDN design. Note that the purpose of the baseline PDN design is to provide baseline K_{th} and WIR values for edge cost calibration.

B. Edge Cost Characterization and Shortest Path

From the baseline PDN design, we perform one of the following operations at a time to obtain a variant PDN layer combination and calculate the routability and WIR costs of each edge in the graph.

- 1) *Single-Layer PDN Resource Tuning*: Replace an edge that belongs to baseline PDN with min or max edge.
- 2) *PDN Layer Skipping*: Replace two consecutive edges in the baseline PDN with a skip edge.

For each variant PDN layer combination, we perform routability analysis (respectively, IR drop analysis) and calculate the difference from baseline PDN layer combination analysis result to obtain edge routability cost (respectively, WIR cost). We normalize the routability cost and WIR cost using

$$\bar{x} = (x - \mu) / \sigma \quad (2)$$

where μ represents the mean value of all raw data and σ represents the standard deviation of all data for routability and WIR, respectively. Note that, for routability cost, we take the negative normalized value of the difference in K_{th} since higher K_{th} indicates better routability, which should correspond to a lower cost in the sensitivity graph.

For a given routability-IR tradeoff factor α , we apply (1) to calculate the cost for each edge. We then apply the shortest-path algorithm to obtain the shortest path from node $N2$ to node $N9$, which represents the PDN layer combination that has the minimum weighted sum of routability and WIR costs.

V. EXPERIMENT AND VALIDATION OF PDN LAYER COMBINATION PATHFINDING

In this section, we describe our experimental setup and the results of the PDN layer combination pathfinding.

We perform experiments with an eight-track 28-nm FDSOI foundry enablement with a ten-metal-layer BEOL stack. The row utilization is determined by the number of available cell rows. For example, eight vertical tracks on a cell and

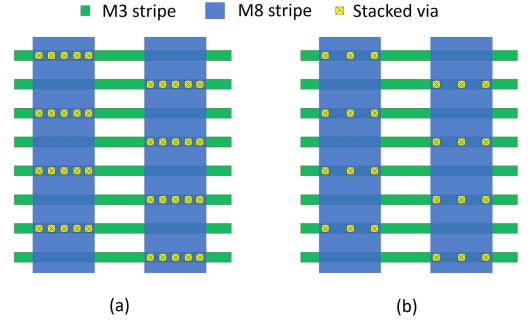


Fig. 9. Illustration of PG via array generation strategies. (a) Continuous via array. (b) Split via array.

TABLE II
REFERENCE DESIGN OF PDN FOR 28-nm FDSOI DESIGN AND PDN LAYER CONFIGURATIONS FOR SCALABILITY STUDY

PDN design										
Metal layer	Direction	width (μm)			spacing (μm)			pitch (μm)		
		small	ref.	big	small	ref.	big	small	ref.	big
M2	H	Standard cell power rails								
M3	V	0.3	0.4	0.7	7.5	10.0	17.5	15.0	20.0	35.0
M4	H	0.4	0.4	0.7	0.6	0.8	1.4	9.0	12.0	21.0
B1 (M7)	V	6.0	8.0	14.0	12.0	16.0	28.0	45.0	60.0	105.0
B2 (M8)	H	14.0	10.0	17.5	15.0	20.0	35.0	52.5	70.0	122.5
Circuit design										
#Instances	25000									
Utilization	0.7									
VI_{density}	0.05									

three vertical tracks of white space imply a row utilization of 0.727.⁹ For the PROBE-like routability study, we perform place-and-route using Cadence Innovus Implementation System v17.10 [39]. To comprehend the impact of TSVs on routability in 3-D IC implementation, we use TSVs as routing blockages during the PROBE-like routability study. Note that the PDN layer combinations considered in this work can have M3 PG stripes connected to M8 PG stripes directly using PG vias. In order to avoid blocking an excessive amount of routing resources between two nonneighboring routing layers, we split the generated PG vias, as shown in Fig. 9. Hence, the PDNs in this work are different from the PDNs in [16] even if the parameters are the same.

For the WIR study, we perform static IR analysis using ANSYS RedHawk v15.1.1 [42]. To capture the impact of TSVs on IR drop in 3-D IC implementation, TSVs are treated as blockages when we construct the PDNs, and the power is supplied through TSVs and then redistributed from the top metal layer, as illustrated in Fig. 5(b). Table II shows the reference design that we use for our experiments.¹⁰

A. Scalability Study

We study the scalability of our approach by varying design size as described. We perform routability analysis using variations of the reference PDN design.¹¹ We sweep the number of cells from 25k to 100k with a step size of 25k for a fixed utilization. A total of 24 distinct PDNs are enumerated by varying one parameter at a time, between 75% (small) and

⁹For ease of use, the values of the following utilizations are rounded to the first decimal place.

¹⁰For the feasibility of our determining the ground truth for all PDN combinations, we set up a reference with a low top-most layer.

¹¹WIR in 3-D IC depends on specific boundary conditions. We experimentally confirm that there is no obvious correlation between #Instances and WIR for a given utilization.

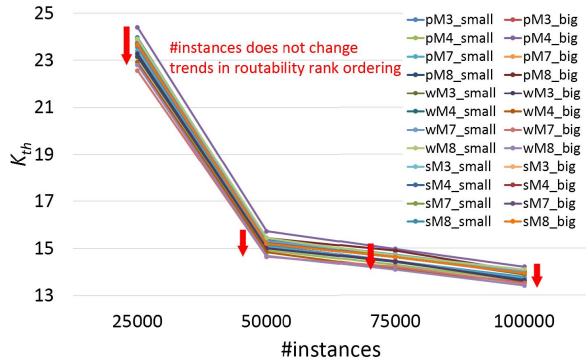


Fig. 10. Routability (K_{th}) versus #instances for 24 PDN variants derived from the reference PDN.

175% (big) of the corresponding value used in the reference PDN, as shown in Table II.

Fig. 10 shows the impact of design size (in terms of #instances) on routability (i.e., K_{th}). We can observe that routability decreases as we increase the design size. Note that, as expected, none of the 24 PDN variants from the reference PDN changes routability dramatically for a given design size. Although there is a change in the absolute value of K_{th} when design size changes (as explained in [15]), the routability rank ordering of PDN designs is independent of design size. Based on our empirical observation of this stability under scaling, we fix the number of instances at 25k for all the experiments reported in the following.

B. Sensitivity-Based PDN Layer Combination Pathfinding

To validate our approach, we perform experiments to obtain the ground-truth data of all PDN layer combinations defined by PDN layer usage corners for each metal layer. We assess the accuracy of our sensitivity-based PDN layer combination pathfinding approach by comparing the rank ordering of path-based cost and the rank ordering from ground-truth data. To assess the impact of routability-IR tradeoff factor α on our PDN layer combination pathfinding approach, we use different α values and perform the rank ordering comparison.

1) *Routability and WIR Sensitivity Graph Construction*: As mentioned in Section IV, we measure the impact of adding PG stripes on a certain metal layer to build a sensitivity graph comprehending both routability and WIR. We apply single-layer PDN resource tuning (i.e., switching between PDN layer usage corners) or 2) PDN layer skipping to obtain PDN layer combination variants. Compared with a baseline PDN layer combination (i.e., $\{M2_{base} - M3_{base} - M4_{base} - M5_{base} - M6_{base} - M7_{base} - M8_{base}\}$ path in Fig. 8), we measure the difference in routability (in terms of K_{th} value) and WIR between the baseline PDN layer combination and PDN layer combination variants. Table III shows the various PDN usage corners that we consider in this work. The minimum PDN is set to 75% of the width of the base PDN, and the spacing and the set-to-set pitch are set to 175%.

Tables IV and V show the raw/normalized routability and WIR cost, respectively. As mentioned in Section IV, we take the negative value of normalized routability cost for weighted sum edge cost calculation for a given routability-IR tradeoff factor α . We illustrate the ranges of WIR and routability sensitivity for corner cases (i.e., min and max for each layer) in Fig. 11. We can observe that, for each metal layer,

TABLE III
PDN CONFIGURATIONS FOR EACH METAL LAYER WHICH USE MIN, BASE, AND MAX RESOURCES. PDN DENSITY IS CALCULATED AS THE NUMBER OF BLOCKED TRACKS FROM THE TOTAL NUMBER OF TRACKS PER LAYER

Metal layer configuration		width (μm)	spacing (μm)	pitch (μm)	#Avail track	#Track blocked	PDN density
M3	Minimum	0.3	17.5	35	1660	60	0.035
	Base	0.4	10	20	1601	119	0.069
	Maximum	0.7	7.5	15	1490	230	0.134
M4	Minimum	0.3	1.4	21	1619	108	0.063
	Base	0.4	0.8	12	1517	210	0.122
	Maximum	0.7	0.6	9	1347	480	0.263
M5	Minimum	0.75	3.5	70	1654	66	0.038
	Base	1	2	40	1590	130	0.076
	Maximum	1.75	1.5	30	1384	336	0.195
M6	Minimum	0.75	3.5	70	1661	66	0.038
	Base	1	2	40	1597	130	0.075
	Maximum	1.75	1.5	30	1391	336	0.195
B1 (M7)	Minimum	6	28	105	728	132	0.153
	Base	8	16	60	602	258	0.300
	Maximum	14	12	45	349	511	0.594
B2 (M8)	Minimum	7.5	35	122.5	743	120	0.139
	Base	10	20	70	598	265	0.307
	Maximum	17.5	15	52.5	323	540	0.626

TABLE IV
RAW AND NORMALIZED ROUTABILITY EDGE SENSITIVITY COST

Edge	Min		Base		Max	
	raw	norm.	raw	norm.	raw	norm.
M3	23.0	-0.434	22.3	-0.113	22.0	0.368
M4	23.2	-0.594	22.6	-0.113	21.2	1.008
M5	22.8	-0.273	22.6	-0.113	19.8	2.130
M6	22.8	-0.273	22.6	-0.113	21.4	0.848
M7	22.8	-0.273	22.6	-0.113	20.8	1.329
M8	23.2	-0.273	22.6	-0.113	21.0	1.169
Skip_M3_M4			23.8	-1.074		
Skip_M4_M5			24.8	-1.876		
Skip_M5_M6			23.6	-0.914		
Skip_M6_M7			23.8	-0.914		

TABLE V
RAW AND NORMALIZED IR EDGE SENSITIVITY COST

Edge	Min		Base		Max	
	raw	norm.	raw	norm.	raw	norm.
M3	0.0154	0.940	0.0129	-0.053	0.0056	-2.952
M4	0.0142	0.463	0.0129	-0.053	0.0126	-0.172
M5	0.0142	0.463	0.0129	-0.053	0.0124	-0.252
M6	0.0131	0.026	0.0129	-0.053	0.0123	-0.291
M7	0.0134	0.145	0.0129	-0.053	0.0128	-0.093
M8	0.0126	-0.172	0.0129	-0.053	0.0126	-0.172
Skip_M3_M4			0.0182	2.052		
Skip_M4_M5			0.0152	0.860		
Skip_M5_M6			0.0128	-0.093		
Skip_M6_M7			0.0166	1.416		

the sensitivity values of the baseline PDN design for both WIR and routability lie between the corresponding values of min and max PDN, as expected.

2) *Sensitivity-Based PDN Layer Combination Pathfinding*: To assess the accuracy of our graph-based method, we measure all possible PDN layer combinations in the graph to obtain ground-truth WIR and routability values. For a given pair of WIR and routability weights, we compare the rank ordering of all layer combinations between the graph-based result and the ground truth. Fig. 12 shows the rank ordering comparison between the graph-based method and ground truth. We achieve a Spearman's coefficient of 0.96, which suggests that the graph-based method can accurately capture the tradeoff between WIR and routability for various layer combinations.

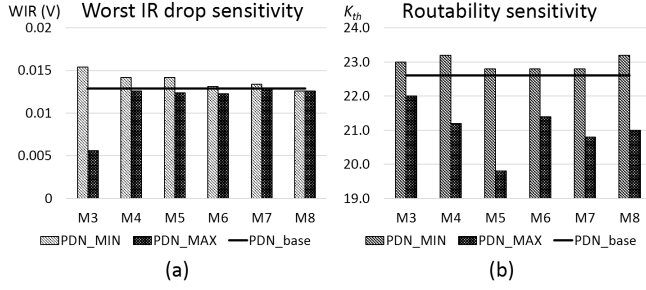


Fig. 11. IR and routability sensitivity analysis results for min/base/max configurations for each layer of the PDN. (a) WIR sensitivity. (b) Routability sensitivity.

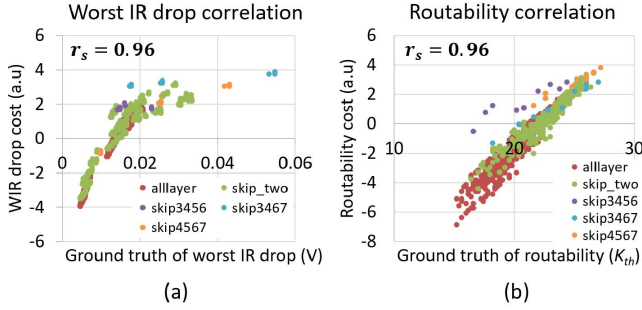


Fig. 12. Correlation results between graph-based method and ground truth for (a) WIR and (b) routability.

3) *Impact of α on Rank Ordering*: We use different routability-IR tradeoff factor α values to assess the impact of α on rank ordering. We use three α values {0.2, 0.5, 0.8} to represent different tradeoffs between routability and WIR. For each α value, we perform experiments with a total of 1080 PDNs. We summarize the breakdown of PDNs from different scenarios as follows.

- 1) *Using All Layers*: $3^{\text{#PDN Layers}} = 729$.
- 2) *Skipping Two Layers*: $3^{\text{#PDN Layers}-2} \times \text{#skipCase} = 3^4 \times 4 = 324$.
- 3) *Skipping Four Layers*: $3^{\text{#PDN Layers}-4} \times \text{#skipCase} = 3^2 \times 3 = 27$.

Fig. 13(a) and (b) shows the graph-based approach and ground truth of tradeoff between routability and WIR with the three α values, respectively. We highlight the Pareto curve from the graph-based approach in both Fig. 13(a) and (b). We can observe that the Pareto curve obtained from the graph-based approach fits the ground-truth tradeoff, which provides confirmation of the effectiveness of our approach.

VI. STAGE 2: PDN LAYER CONFIGURATION PATHFINDING

In this section, we describe the problem statement and our model-based flow for the PDN layer configuration pathfinding problem given a PDN layer combination. For a given PDN layer combination, we define the PDN layer configuration pathfinding problem as follows.

- 1) *PDN Layer Configuration Pathfinding Problem*: Given a mesh-like placement, VI locations, and a PDN layer combination, provide a PDN design that meets the WIR limit with best routability.
- 2) *Inputs*: Mesh-like placement, VI locations, and PDN layer combination.
- 3) *Output*: PDN with best routability meeting the WIR limit.
- 4) *Constraints*: WIR and technology design rules.

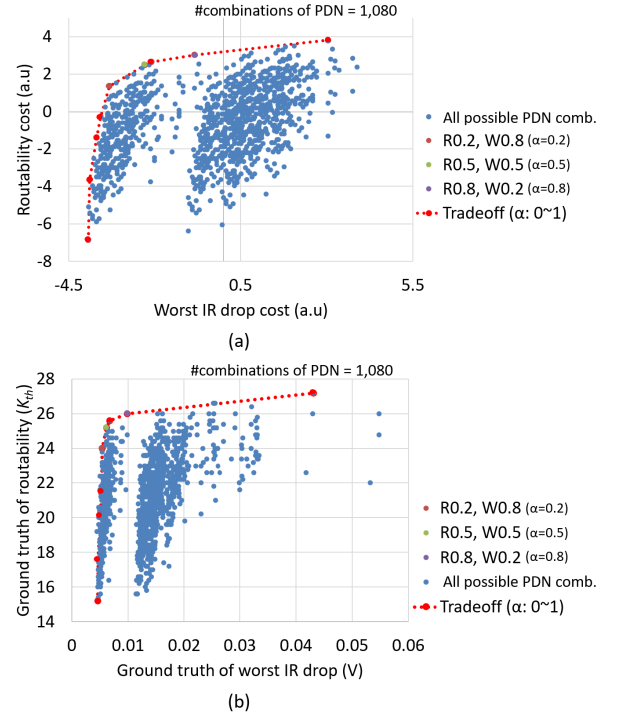


Fig. 13. Illustration of tradeoff between routability and WIR with $\alpha = \{0.2, 0.5, 0.8\}$ for (a) graph-based approach and (b) ground truth. The Pareto curve in (a) and (b) are both from the graph-based approach.

A. WIR and Routability Modeling

Fig. 14 illustrates the WIR and routability modeling flow. For a collection of PDN candidates, we perform static IR analysis on a mesh-like placement. We sweep the circuit design-independent knobs (i.e., width, spacing, and pitch of PG stripes) to generate a training data set of PDN layer configurations and obtain their corresponding WIR values. Based on the WIR values in the data set, we train a WIR model and use the model to predict the WIR for different PDN layer configurations. Fig. 14(a) illustrates the WIR modeling flow. Similar to WIR modeling flow, we build a routability model based on K_{th} values from the PDN layer configuration candidates. Fig. 14(b) illustrates the routability modeling flow. We perform PROBE-like routability analysis [15] to collect K_{th} data for various PDN layer configuration candidates. Besides the PDN variables including metal width, spacing, and pitch, we also consider utilization and $VI_{density}$ in the routability model, so as to comprehend the competition for routing resources between PDN and signal routing.

We use learning-based algorithms, such as the ordinary least-squares method (multivariable linear regression), and MARS [14], to build regression models for both WIR and routability. By combining several models (multivariable linear regression and MARS), we achieve a hybrid surrogate model to assess the WIR and routability of PDN layer configurations. Model validations are discussed in Section VII-C.

B. Model-Based PDN Layer Configuration Pathfinding

For a given PDN layer combination, PDN layer configurations are enumerated honoring technology constraints (width, space, and pitch) for all stripes on all layers. We use the WIR model to prune the PDN solution space according to the WIR requirement. For the enumerated PDN layer configurations, we apply the WIR model to predict their respective WIR

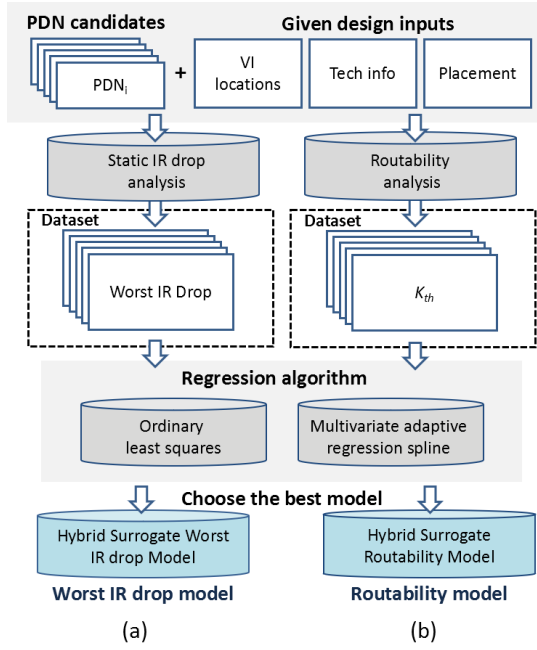


Fig. 14. (a) WIR modeling flow. (b) Routability modeling flow.

values and find PDN layer configurations that satisfy the WIR requirement. We then use our routability model to rank PDN layer configurations, which satisfies WIR constraint, based on their routability. Based on the WIR and routability models, our flow returns a PDN layer configuration that satisfies the WIR constraint and has the best routability. This PDN solution will in our experience provide the highest probability of a clean 3-D IC implementation.

C. Evaluation Metric

For the WIR model, we compare the measured WIR values and the predicted WIR values to assess our WIR model accuracy. For the routability model, we consider our goal to provide the most routable PDN layer configuration. We rank the relative routability by the K_{th} value over the absolute value of K_{th} predicted through regression. Thus, not only the linearity expressed by the Pearson correlation coefficient [25] but also the ranking comparison by each K_{th} is required. We use the Spearman's rank correlation coefficient [30] to compare the routability ranking of PDNs with predicted K_{th} values with the ranking of PDNs with real K_{th} values obtained experimentally from PROBE-like analyses. Spearman's coefficient of ≥ 0.9 between the two rankings may be taken as evidence of a strong correlation.

VII. EXPERIMENT AND VALIDATION OF PDN LAYER CONFIGURATION PATHFINDING

In this section, we describe our result of the experiment and validation of PDN layer configuration pathfinding. For each model in PDN layer configuration pathfinding, we use 67% of the overall data set for training and the remaining 33% of the data set for testing. We use a MARS implementation in Python3 from the Py-earth package [35]. Other aspects of the experimental setup are the same as in Section V.

A. PDN Layer Configuration Sensitivity Study

To assess the impact of each PDN and circuit design knob on WIR drop and routability for a given PDN layer

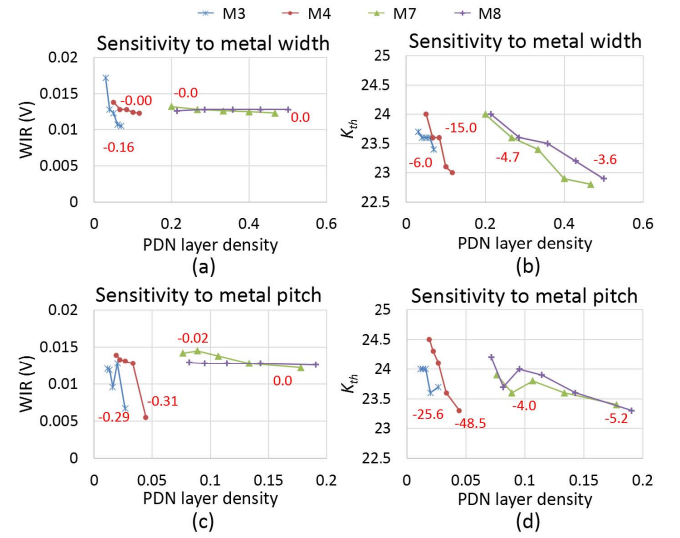


Fig. 15. WIR (left) and routability (right) sensitivity to circuit-independent knobs width (top) and set-to-set pitch (bottom). The red numbers indicate the slope of the K_{th} change with each knob.

combination, we investigate the sensitivities of WIR and routability to various design knobs discussed in Section VI. For PDN design knobs, all circuit-independent design knobs of width, spacing, and pitch for M3, M4, M7, and M8 are considered. For circuit-dependent design knobs, we consider utilization and $VI_{density}$. Only one knob is swept at a time, while all other knobs are fixed at their values in the reference design. Fig. 15 shows the sensitivity results between WIR/routability (y-axis) and PDN density (x-axis) by varying design knobs. The PDN density of each layer is calculated as $2 \times \text{width/pitch}$.

1) *Width*: We sweep width for M3, M4, M7, and M8 from 75% to 175% of the reference value. Fig. 15(a) shows the WIR as a function of width for M4, M7, and M8 separately. WIR decreases as we increase the width since VDD/VSS stripes become less resistive. Fig. 15(b) shows routability as a function of width. For all layers, routability decreases as width increases since less routing resource is available. Moreover, there is less sensitivity of routability to PDN layer density on higher layers.

2) *Spacing*: We sweep the VDD/VSS stripes spacing for M3, M4, M7, and M8 from 75% to 175% of the reference value. Spacing between VDD and VSS stripes is in practice mainly used to control dynamic IR drop, and it does not have a significant effect on static IR drop. The effect of spacing on routability is also negligible.

3) *Pitch*: We sweep the M4 VDD/VSS stripe pitch for M3, M4, M7, and M8 from 75% to 175% of the reference value. Fig. 15(c) shows that WIR increases as we increase pitch (i.e., sparser power mesh). Fig. 15(d) shows that routability decreases as PDN layer density increases. However, there is higher sensitivity to pitch than width, even with the same PDN layer density.

4) *Row Utilization*: In our routability model development, our use of mesh-like placement implies that current density is proportional to the row utilization of the placement. Fig. 16(a) shows WIR versus PDN layer density (determined by metal width), while Fig. 16(c) shows WIR versus PDN layer density (determined by metal pitch), on M3, M4, M7, and M8. Since IR drop is proportional to current density, which is, in turn,

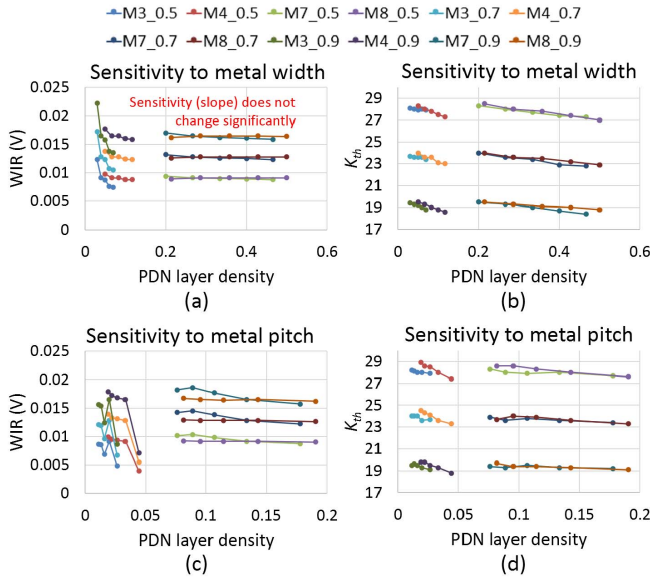


Fig. 16. WIR (left) and routability (right) sensitivity analysis results for circuit-independent knobs width (top) and set-to-set pitch (bottom) with various utilizations {0.5, 0.7, 0.9}.

proportional to row utilization in a uniform placement, we see that WIR is proportional to utilization.

Designs with higher row utilization in the placement tend to have DRVs on lower metal layers due to a lack of routing resources for pin access and/or promotion. Therefore, we simultaneously sweep design utilization and metal width (respectively, pitch) to study the routability impact of PDN design due to interactions between design utilization and stripe width (respectively, pitch). Fig. 16(b) shows the routability as a function of utilization and metal width, and Fig. 16(d) shows the routability as a function of utilization and metal pitch, on layers M3, M4, M7, and M8. We observe that routability decreases as we increase the utilization. We also observe that, for a given utilization, routability is more sensitive to changes in lower metal layers.

5) $VI_{density}$: We sweep the $VI_{density}$ from 0.025 to 0.25. Similar to the utilization sensitivity study, we simultaneously sweep metal width or pitch along with $VI_{density}$, as VI accessibility intuitively depends more on routing resources on the higher metal layer. Since signal VIs are circuit-dependent and affect only routing resources, only the routability analysis is performed.¹² $VI_{density}$ is given in Table VI. Fig. 17(a) shows the routability as a function of VI density and metal width, and Fig. 17(b) shows the routability as a function of $VI_{density}$ and metal pitch. We observe that routability suddenly decreases as we increase the $VI_{density}$. Moreover, for a given $VI_{density}$, routability is more sensitive to changes in higher metal layers, as we might expect.

B. WIR Model

To efficiently assess whether a PDN design satisfies the WIR requirement, we build a WIR model based on a data set that includes combinations of knob values from width, pitch, and utilization. In our experiment, we sweep the value of each knob from 75% to 125% of its reference value

¹²There is a slight difference between the target and actual $VI_{density}$ because the VI should be aligned to the cell grid in a mesh-like placement to guarantee the same distance between the VI and the connected net.

TABLE VI
SENSITIVITY TO VI DENSITIES (#NETS = 25 172)

Target $VI_{density}$	#VIs	$VI_{density}$
0.025	684	0.027
0.050	1242	0.049
0.075	2052	0.082
0.100	2736	0.110
0.250	6266	0.250

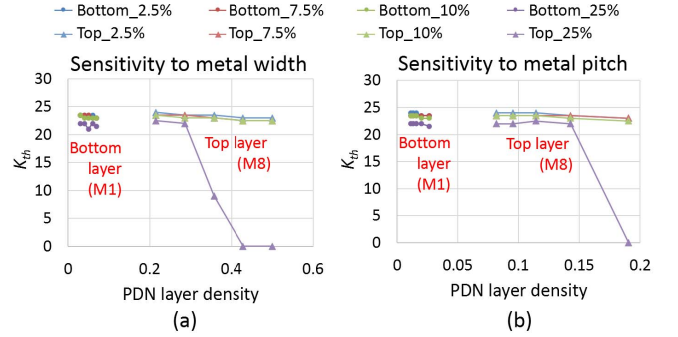


Fig. 17. Routability sensitivity analysis results for circuit-independent knobs (a) width and (b) set-to-set pitch, with various VI densities.

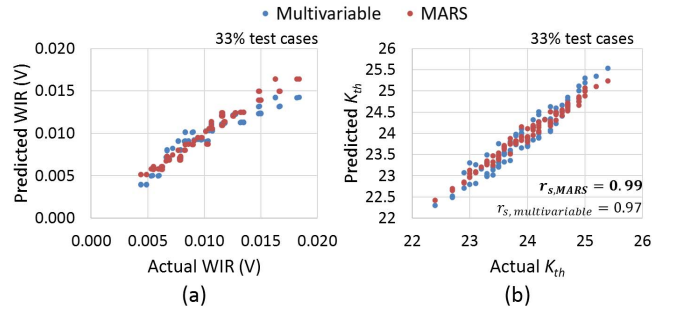


Fig. 18. Modeling results. (a) WIR model. (b) Routability model.

(e.g., 0.3–0.5 μm for M3 stripe width). Fig. 18(a) shows actual versus predicted WIR for various PDN designs with combinations of PDN design knob values. Our model achieves an absolute average error of 0.75 mV (respectively, 0.98 mV) for the training (respectively, testing) data set.

C. Routability Model

To find an optimal PDN, we must be able to rank PDN designs that satisfy the WIR requirement by routability. We use the same data set as in Section VII-B to build a routability model. The input of the model is a sequence of PDN design knobs for all metal layers in the BEOL stack, along with circuit design knobs. Fig. 18(b) illustrates correlation between the actual K_{th} and the predicted K_{th} by the routability model.

To assess the generality of our model, we also build another routability model based on a data set that is composed of routability data with knob values of {85%, 115%} of respective reference values (i.e., a “subset” of the original ({75%, 125%}) data set). We then test our model in an “Extrapolation” case (i.e., from the “subset” to the original data set) and in an “Interpolation” case (i.e., from the original data set to the “subset”). Fig. 19(a) shows that we achieve Spearman’s coefficient of 0.95 (respectively, 0.93) with multivariable linear regression (respectively, MARS) for

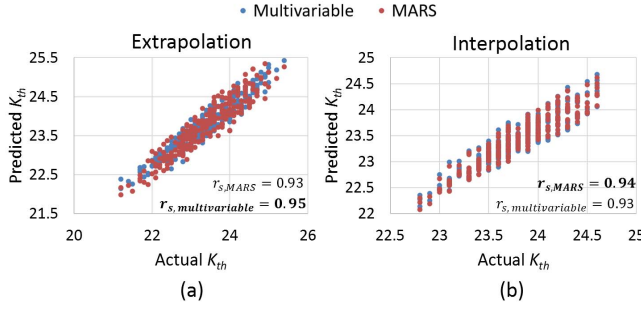


Fig. 19. Correlation of routability between the actual K_{th} and predicted K_{th} values of (a) extrapolation and (b) interpolation. The scatter points displayed in the graph represent a total of 256 #testing points and a total of 256 #PDNs training points.

TABLE VII

SIMULATION RESULTS WITH THE AES AND JPEG TESTCASES, WHERE THE K_{th} VALUES ARE AVERAGES OVER FIVE DENOISING RUNS

PDN	AES				JPEG			
	clk (ns)	#inst	K_{th}	WIR (V)	clk (ns)	#inst	K_{th}	WIR (V)
Best			7.94	0.0352			16.24	0.067
Reference	1.0	10k	5.68	0.0369	1.4	24k	15.62	0.0701
Worst			5.5	0.0206			14.74	0.0501

the Extrapolation case. Fig. 19(b) shows analogous values of 0.93 (respectively, 0.94) for the Interpolation case. This suggests that our model can be generalized and used for other testcases via interpolation and extrapolation.

D. Verification on Real Design Block

We verify our routability and WIR models by applying PDN layer configuration pathfinding methodology to real design testcases. We use the AES encryption and JPEG encoder cores from OpenCores [38]. Each design is synthesized with Synopsys Design Compiler L-2016.03-SP4-1 [43]. We perform experiments with eight-track standard cells from a 28-nm FDSOI foundry technology library. Since cells of real design blocks do not have uniform width as in a mesh-like placement, we perform legalization before routing to eliminate overlap caused by random swapping of neighboring cells. To apply the proposed routability model, we add VIs as I/O pins and then place the pins uniformly on the top metal at the $VI_{density}$ used in the model (5% of #VIs/#nets). The additional VIs are connected to the nearest different nets.

Without loss of generality, we use the WIR value of the reference PDN design as the WIR requirement for each testcase. The BEOL stack of the PDN is the same as that of the reference PDN of Table II. Based on the trained routability model, PDNs with a WIR greater than the WIR for the reference PDN are filtered; then, the design knobs that constitute the best PDN can be obtained through the predictive model. To validate the ranking of the routability model, we pick the best PDN, a reference PDN, and a worst-quality PDN for verification with real design blocks. Table VII shows the verification results with the AES cipher and JPEG encoder testcases. In actual designs, the cell placement is not uniform, so the denoising is performed through five different random seeds, and K_{th} of Table VII is the average value of five runs. Fig. 20 shows that, for design blocks AES and JPEG, superior PDNs, which have lower WIR and better routability, are found. Note that the placement in real design blocks is less uniform compared with mesh-like placement, which explains

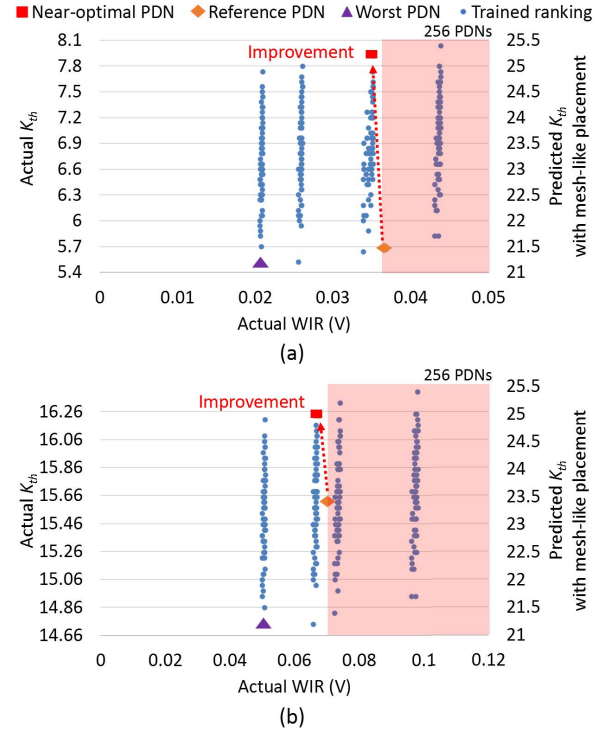


Fig. 20. Routability (K_{th}) versus WIR data for (a) AES encryption core and (b) JPEG encoder testcases. Blue dots denote the trained ranking of PDNs and are represented by the second y-axis as K_{th} values. Near-optimal, reference, and worst PDNs are verified by real design blocks. The red arrows indicate improvement from the reference PDN. The red regions indicate WIR greater than the WIR of the reference PDN.

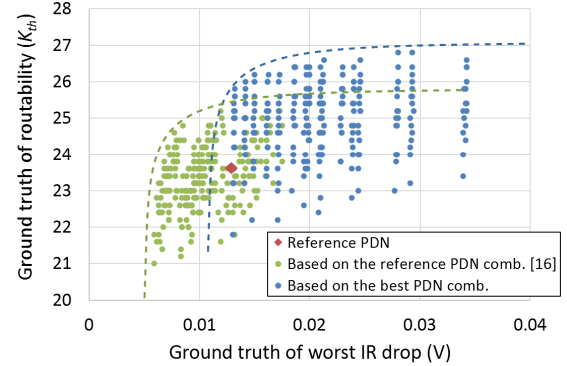


Fig. 21. WIR and K_{th} of PDN layer configurations using the best PDN layer combination from this work (blue dots) and the (human-designed) PDN layer combination in [16] (green dots). Toward the upper left corner is better.

the discrepancy between the actual K_{th} and the predicted K_{th} . However, our methodology is applicable as long as the rank ordering maintains, and we have experimentally verified that the rank orderings from mesh-like placement and real design blocks are the same.

VIII. VALIDATION OF TWO-STAGE PDN PATHFINDING

In this section, we validate our overall two-stage PDN pathfinding methodology. We verify our two-stage PDN pathfinding methodology using the PDN layer combination from [16], along with the best PDN layer combinations from PDN layer combination pathfinding. The best PDN layer combination that we obtain is $M_{3max} - M_{8min}$. Considering the same PDN layer configuration solution space as in

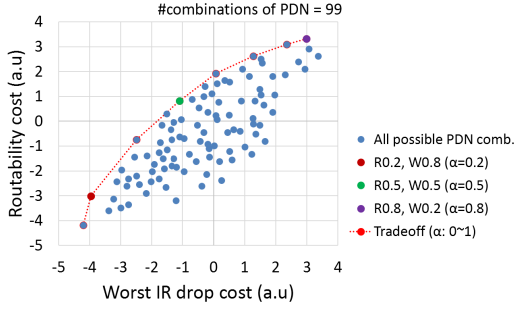


Fig. 22. Illustration of tradeoff between routability and WIR with $\alpha = \{0.2, 0.5, 0.8\}$ for graph-based approach in foundry 14-nm technology. The Pareto curve is from the graph-based approach.

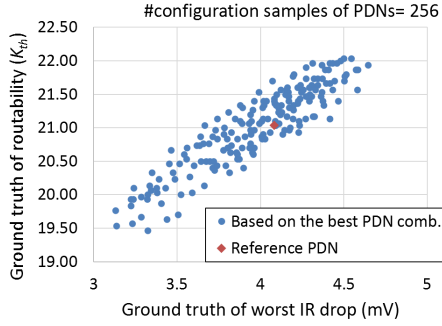


Fig. 23. WIR and K_{th} of PDN layer configurations using the best PDN layer combination in foundry 14-nm technology. Toward the upper-left corner is better.

TABLE VIII

REFERENCE DESIGN OF PDN FOR 14-nm FOUNDRY TECHNOLOGY

PDN design				
Metal layer	Direction	width (μm)	spacing (μm)	pitch (μm)
M2	H	Standard cell power rails		
C5 (M5)	V	0.5	13.5	28
C6 (M6)	H	2.01	11.99	28
Circuit design				
#Instances	25000			
Utilization	0.7			
$V_{I_{density}}$	0.05			

Section VII-B, we obtain the ground-truth WIR and routability values for the best PDN layer combinations. Fig. 21 shows the routability and WIR of: 1) the reference PDN; 2) PDNs based on the reference layer combination from [16]; and 3) PDNs based on the best layer combination from this work. Recall that our goal is to find the PDN that: 1) satisfies the given WIR constraint and 2) has the best routability. The best layer combination found by our approach is superior to [16], as our WIR-routability envelope contains solutions that have both lower WIR and better routability (i.e., higher K_{th}) than [16].¹³

Note that, for our two-stage methodology, execution of Stage 1 requires 17 runs with mesh-like placement to determine the best PDN layer combination. Execution of Stage 2 requires 256 runs to obtain WIR and K_{th} data for WIR and routability models. Each tool run, for both WIR and K_{th} data gathering, takes around 3 h. Thus, the overall runtime

¹³Note that although the average WIR value of all data points from [16] is lower than that in this work, our work achieves routability-dominant PDN solutions that satisfy the WIR constraint, as illustrated by the envelopes in Fig. 21.

TABLE IX

PDN CONFIGURATIONS FOR EACH METAL LAYER WHICH USE MIN, BASE, AND MAX RESOURCES IN 14-nm FOUNDRY TECHNOLOGY. PDN DENSITY IS CALCULATED AS THE NUMBER OF BLOCKED TRACKS FROM THE TOTAL NUMBER OF TRACKS PER LAYER

Metal layer configuration		width (μm)	spacing (μm)	pitch (μm)	#Avail track	#Track blocked	PDN density
M3	Minimum	0.15	17.5	35	2136	38	0.01
	Base	0.2	10	20	2089	85	0.02
	Maximum	0.35	7.5	15	1998	176	0.05
M4 (C4)	Minimum	0.225	17.5	35	1703	44	0.01
	Base	0.3	10	20	1659	88	0.03
	Maximum	0.525	7.5	15	1570	177	0.07
M5 (C5)	Minimum	0.375	23.625	49	1678	61	0.02
	Base	0.5	13.5	28	1634	105	0.04
	Maximum	0.875	10.125	21	1523	216	0.08
M6 (C6)	Minimum	1.5075	20.9825	49	1552	195	0.06
	Base	2.01	11.99	28	1351	396	0.14
	Maximum	3	11.99	28	1245	502	0.21

is approximately $(17+256) \text{ runs} \times 3 \text{ h} = 819 \text{ h}$. Using parallel execution with 20 processes, we are able to complete Stage 1 (respectively, Stage 2) within 3 h (respectively, 39 h) to obtain a high-quality PDN layer configuration based on the best PDN layer combination.

Moreover, to verify the overall two-stage PDN pathfinding flow on real design blocks, we compare the WIR and routability values of the two real design blocks using the following three PDN designs¹⁴:

- 1) industry reference PDN design in Section V;
- 2) the best PDN design in [16];
- 3) the best PDN design from this work.

Table X compares WIR and routability across the industry reference PDN design, the best PDN design in [16], and our best PDN design, on the AES and JPEG blocks. We observe that our best PDN design in this work has superiority over both the industry reference PDN design and the best PDN design in [16]. Our best PDN design in this work achieves up to 16% and 12% improvements in WIR compared with the industry reference PDN design and the best PDN design in [16], respectively. Our best PDN design in this work also achieves up to 35% and 10% improvements in routability compared with the industry reference PDN design and the best PDN design in [16], respectively. Furthermore, the average WNS and TNS of both real design blocks are improved over the previous [16] results and the reference results, while the routing resource usage difference is less than 1%.

IX. ADDITIONAL STUDY IN 14-nm FOUNDRY TECHNOLOGY

The abovementioned studies focus on finding near-optimal PDN for the FDSOI 28-nm technology library. However, in FinFET nodes and with tremendous pressure to maintain density scaling, the number of available routing tracks is further reduced, and the design rules become more complicated. To assess the general applicability of our pathfinding methodology in advanced technology, we perform further validations using a 10.5-track 14-nm foundry library and a nine-metal-layer BEOL stack. Since the 14-nm library collateral utilizes Cadence Quantus QRC [40] format, we use Cadence Voltus IC Power Integrity Solution [41] to measure WIR. Table VIII

¹⁴To compensate for potential modeling error in WIR, we apply a 10% margin for the WIR model. That is, we only consider PDN designs that have 90% of the required WIR, or better.

TABLE X
ROUTABILITY (K_{th}) AND WIR USING INDUSTRY REFERENCE PDN, BEST PDN IN [16], AND BEST PDN IN THIS
WORK WITH REAL DESIGN BLOCKS IN 28- AND 14-nm FOUNDRY TECHNOLOGY LIBRARIES

Technology library	PDN	AES						JPEG					
		WIR (mV)	K_{th}	Wirelength (μm)	#Vias	TNS (ns)	WNS (ns)	WIR (mV)	K_{th}	Wirelength (μm)	#Vias	TNS (ns)	WNS (ns)
FDSOI 28nm	Reference	36.9	5.7	113260	113115	0.0	0.011	70.1	15.6	323840	234629	-313.587	-0.591
	Best in [16]	35.2	7.9	113600	112858	0.0	0.005	67.0	16.2	324920	241521	-226.684	-0.524
	Ours	29.6	9.3	113380	112105	0.0	0.002	61.4	16.6	324620	237382	-294.706	-0.614
Foundry 14nm	Reference	12.0	12.2	73530	116224	0	0.398	27.0	22.7	187800	289775	0	0.536
	Ours	10.0	15.3	74350	117729	0	0.399	22.0	25.1	190000	294807	0	0.493

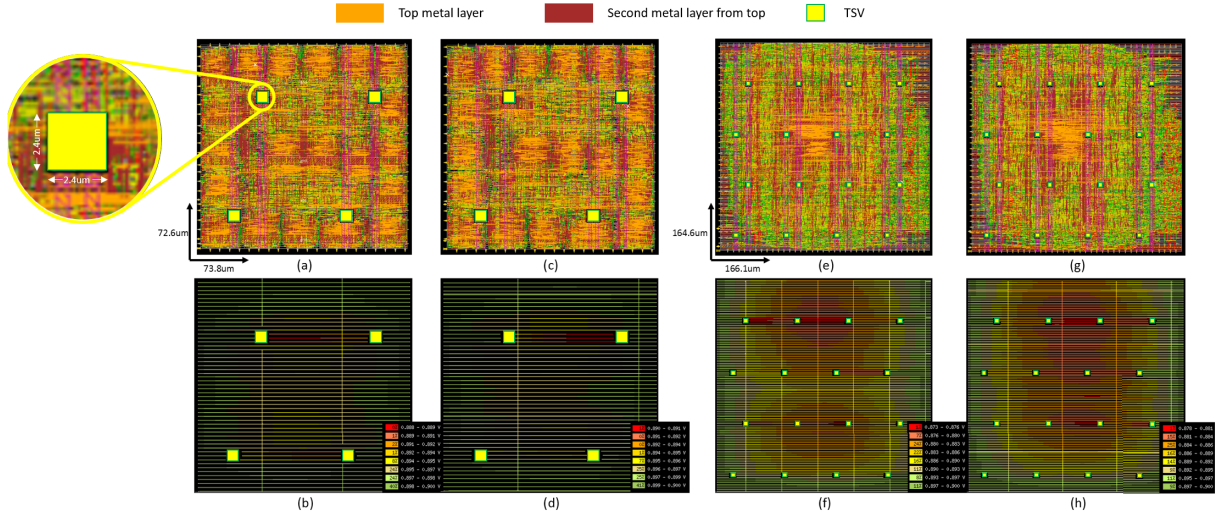


Fig. 24. Postrouting layout (top, with highlighted TSV allocation) and rail analysis (bottom) results of (a)–(d) AES and (e)–(h) JPEG. (a), (b), (e), and (f) Left-hand side of each design is the reference PDN and (c), (d), (g), and (h) right-hand side is our PDN in 14-nm foundry technology.

shows the industrial reference PDN design for the 14-nm foundry technology.

A. PDN Layer Combination Pathfinding for Foundry 14 nm

We identify high-quality layer combinations using the same methodology, as described in Section IV. The configurations of the PDN for the sensitivity graph are shown in Table IX. From a reference PDN layer combination with the top PDN layer of M6, we derive a total of 99 PDN layer combination variants. We obtain the shortest-path according to the coefficients of routability and WIR in the sensitivity graph and plot the tradeoff as a boundary in Fig. 22. Unlike the 28-nm FDSOI experiment, the best layer combination in 14-nm technology is the same as the reference PDN (M2 rail, M5, and M6).

B. PDN Layer Configuration Pathfinding for Foundry 14 nm

We first train routability and WIR models by generating PDNs with a total of 256 different configurations from the best PDN combination. Fig. 23 shows the tradeoff between WIR and routability, based on the ground-truth data, for the best PDN combination. Then, we solve the trained regression model of WIR and routability as a linear program to obtain the best PDN configuration, as follows: {wM5, wM6, sM5, sM6, pM5, pM6} = {1.353, 1.005, 13.5, 11.99, 42, 42} μm . Note that the best PDN configuration is obtained from the linear program; hence, the configuration is from a larger solution space than that defined in Table IX.

Fig. 24 shows the routed layout and rail analysis results of the AES and JPEG designs. The best PDN achieves 25.4% (respectively, 10.6%) improved routability while satisfying 16.7% (respectively, 18.5%) improved WIR in the AES (respectively, JPEG) designs, as shown in Table X.

The positive slack of the most critical endpoint is reduced by 43 ps in the JPEG testcase, but there is no timing violation in all cases, while the routing resource usage difference is less than 1%.

X. CONCLUSION

In this work, we present a novel two-stage power delivery pathfinding methodology for emerging 3-D F2F integration technology. Our proposed methodology is capable of navigating the tradeoff between IR drop and routability of PDN designs in the 3-D IC context, where VIs and TSVs introduce additional challenges. We augment the previous per-layer PDN configuration pathfinding of [16] with PDN layer combination pathfinding capability. Using our methodology, we demonstrate the rank-ordering of PDN designs for a given BEOL stack, considering the tradeoff between WIR drop and routing capacity in the design space. We validate our proposed pathfinding methodology with mesh-like placements, as well as with real design blocks in 28-nm technologies. The extra degree of solution space exploration afforded by PDN layer combinations leads to improvements of more than 10% in WIR and routability metrics. Exploring both PDN layer combination and layer configuration, we can achieve better results and estimate suboptimality compared with an industrial reference PDN in 14-nm foundry technology. Our future works include: 1) estimation of K_{th} and WIR values for placement in real SoC designs based on modeled K_{th} and WIR values; 2) extension of our approach to heterogeneous integration technologies beyond F2F integration of two dies; 3) extension of this work to comprehend dynamic IR drop (e.g., optimize the spacing between VDD and VSS power rails, as well as necessary accommodation for decoupling capacitor insertion); 4) extension of our approach to designs with macroblocks;

and 5) extension of this work to foundry sub-7-nm technology nodes where BEOL resistance (along with scaling boosters, such as supervias or buried power rails) will significantly expand the PDN-IR pathfinding solution space.

ACKNOWLEDGMENT

The authors would like to thank Dr. Kambiz Samadi for his contribution in the work in [16].

REFERENCES

- [1] K. Arabi, K. Samadi, and Y. Du, "3D VLSI: A scalable integration beyond 2D," in *Proc. Symp. Int. Symp. Phys. Des.*, 2015, pp. 1–7.
- [2] R. Bhooshan, "Novel and efficient IR-drop models for designing power distribution network for sub-100nm integrated circuits," in *Proc. 8th Int. Symp. Qual. Electron. Des. (ISQED)*, Mar. 2007, pp. 287–292.
- [3] U. Brenner and A. Rohe, "An effective congestion-driven placement framework," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 22, no. 4, pp. 387–394, Apr. 2003.
- [4] A. E. Caldwell, A. B. Kahng, S. Mantik, I. L. Markov, and A. Zelikovsky, "On wirelength estimations for row-based placement," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 18, no. 9, pp. 1265–1278, Aug. 1999.
- [5] W.-T.-J. Chan, P.-H. Ho, A. B. Kahng, and P. Saxena, "Routability optimization for industrial designs at sub-14nm process nodes using machine learning," in *Proc. ACM Int. Symp. Phys. Des.*, Mar. 2017, pp. 15–21.
- [6] W.-T. J. Chan, Y. Du, A. B. Kahng, S. Nath, and K. Samadi, "3D-IC benefit estimation and implementation guidance from 2DIC implementation," in *Proc. DAC*, 2015, pp. 1–6.
- [7] W.-H. Chang *et al.*, "Generating routing-driven power distribution networks with machine-learning technique," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 8, pp. 1237–1250, Aug. 2017.
- [8] K. Chang *et al.*, "Cascade2D: A design-aware partitioning approach to monolithic 3D IC with 2D commercial tools," in *Proc. ICCAD*, 2016, pp. 1–8.
- [9] K. Chang, S. Das, S. Sinha, B. Cline, G. Yeric, and S. K. Lim, "Frequency and time domain analysis of power delivery network for monolithic 3D ICs," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Des. (ISLPED)*, Jul. 2017, pp. 1–6.
- [10] K. Chang, A. Koneru, K. Chakrabarty, and S. K. Lim, "Design automation and testing of monolithic 3D ICs: Opportunities, challenges, and solutions: (Invited paper)," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des. (ICCAD)*, Nov. 2017, pp. 805–810.
- [11] C.-K. Cheng, A. B. Kahng, I. Kang, and L. Wang, "RePlAce: Advancing solution quality and routability validation in global placement," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 38, no. 9, pp. 1717–1730, Sep. 2019.
- [12] V. A. Chhabria, A. B. Kahng, M. Kim, U. Mallappa, S. S. Sapatnekar, and B. Xu, "Template-based PDN synthesis in floorplan and placement using classifier and CNN techniques," in *Proc. ASP-DAC*, 2020, pp. 44–49.
- [13] Y. Du, K. Samadi, and K. Arabi, "Emerging 3DVLSI: Opportunities and challenges," in *Proc. S3S*, 2015, pp. 1–5.
- [14] J. H. Friedman, "Multivariate adaptive regression splines," *Ann. Statist.*, vol. 19, no. 1, pp. 1–67, 1991.
- [15] A. Kahng, A. B. Kahng, H. Lee, and J. Li, "PROBE: A placement, routing, back-end-of-line measurement utility," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 7, pp. 1459–1472, Jul. 2018.
- [16] A. B. Kahng, S. Kang, S. Kim, K. Samadi, and B. Xu, "Power delivery pathfinding for emerging die-to-wafer integration technology," in *Proc. DATE*, 2019, pp. 836–841.
- [17] A. B. Kahng, B. Lin, and S. Nath, "Enhanced metamodeling techniques for high-dimensional IC design estimation problems," in *Proc. DATE*, 2013, pp. 1861–1866.
- [18] M.-C. Kim, J. Hu, D.-J. Lee, and I. L. Markov, "A SimPLR method for routability-driven placement," in *Proc. IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)*, Nov. 2011, pp. 67–73.
- [19] B. W. Ku *et al.*, "Physical design solutions to tackle FEOL/BEOL degradation in gate-level monolithic 3D ICs," in *Proc. Int. Symp. Low Power Electron. Des.*, 2016, pp. 76–81.
- [20] B. W. Ku, K. Chang, and S. K. Lim, "Compact-2D: A physical design methodology to build commercial-quality face-to-face-bonded 3D ICs," in *Proc. Int. Symp. Phys. Des.*, Mar. 2018, pp. 76–81.
- [21] W.-H. Liu, W.-C. Kao, Y.-L. Li, and K.-Y. Chao, "NCTU-GR 2.0: Multithreaded collision-aware global routing with bounded-length maze routing," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 5, pp. 709–722, May 2013.
- [22] S. A. Panth, K. Samadi, Y. Du, and S. K. Lim, "Design and CAD methodologies for low power gate-level monolithic 3D ICs," in *Proc. Int. Symp. Low power Electron. Des.*, 2014, pp. 171–176.
- [23] S. Panth, K. Samadi, Y. Du, and S. K. Lim, "Tier-partitioning for power delivery vs cooling tradeoff in 3D vlsi for mobile applications," in *Proc. DAC*, 2015, p. 92.
- [24] S. Panth, K. Samadi, Y. Du, and S. K. Lim, "Shrunk-2D: A physical design methodology to build commercial-quality monolithic 3D ICs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 10, pp. 1716–1724, Oct. 2017.
- [25] K. Pearson, "Note on regression and inheritance in the case of two parents," in *Proc. Roy. Soc. London*, 1895, pp. 240–242.
- [26] Y. Peng, D. Petranovic, K. Samadi, P. Kamal, Y. Du, and S. K. Lim, "Inter-die coupling extraction and physical design optimization for face-to-face 3D ICs," *IEEE Trans. NANO*, vol. 17, no. 4, pp. 634–644, Jul. 2017.
- [27] Z. Qi, Y. Cai, and Q. Zhou, "Accurate prediction of detailed routing congestion using supervised data learning," in *Proc. ICCD*, 2014, pp. 97–103.
- [28] J. A. Roy and I. L. Markov, "Seeing the forest and the trees: Steiner wirelength optimization in placement," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 26, no. 4, pp. 632–644, Apr. 2007.
- [29] S. K. Samal, K. Samadi, P. Kamal, Y. Du, and S. K. Lim, "Full chip impact study of power delivery network designs in gate-level monolithic 3-D ICs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 36, no. 6, pp. 992–1003, Jun. 2017.
- [30] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904.
- [31] Y. Xu, Y. Zhang, and C. Chu, "FastRoute 4.0: Global router with efficient via minimization," in *Proc. Asia South Pacific Des. Autom. Conf.*, Jan. 2009, pp. 576–581.
- [32] X. Yang, R. Kastner, and M. Sarrafzadeh, "Congestion estimation during top-down placement," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 21, no. 1, pp. 72–80, 2002.
- [33] Q. Zhou, X. Wang, Z. Qi, Z. Chen, Q. Zhou, and Y. Cai, "An accurate detailed routing routability prediction model in placement," in *Proc. ASQED*, 2015, pp. 119–122.
- [34] H. Reiter, *TSMC Details Family of Chip Stacks*. Accessed: Nov. 4, 2019. [Online]. Available: https://www.eetimes.com/author.asp?section_id=36&doc_id=1322075
- [35] J. Rudy, *Py-Earth*. Accessed: Nov. 4, 2019. [Online]. Available: <https://github.com/scikit-learn-contrib/py-earth>
- [36] G. Yeric, *Three Dimensions in 3DIC—Part I*. Accessed: Nov. 4, 2019. [Online]. Available: <https://community.arm.com/arm-research/b/articles/posts/three-dimensions-in-3dic-part-1>
- [37] G. Yeric, *Three Dimensions in 3DIC—Part II*. Accessed: Nov. 4, 2019. [Online]. Available: <https://community.arm.com/arm-research/b/articles/posts/three-dimensions-in-3dic-part-ii>
- [38] *OpenCores: Open Source IP-Cores*. Accessed: Nov. 4, 2019. [Online]. Available: <http://www.opencores.org>
- [39] Cadence Innovus User Guide. Accessed: Nov. 4, 2019. [Online]. Available: <https://www.cadence.com>
- [40] Cadence *Quantus QRC Extraction Users Manual*. Accessed: Nov. 4, 2019. [Online]. Available: <https://www.cadence.com>
- [41] Cadence *Voltus IC Power Integrity Solution User Guide*. Accessed: Nov. 4, 2019. [Online]. Available: <https://www.cadence.com>
- [42] ANSYS *RedHawk User Guide*. Accessed: Nov. 4, 2019. [Online]. Available: <https://www.ansys.com>
- [43] Synopsys *Design Compiler User Guide*. Accessed: Nov. 4, 2019. [Online]. Available: <http://www.synopsys.com>