Heterogeneous 3D ICs: Current Status and Future Directions for Physical Design Technologies

Gauthaman Murali
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, USA
gauthaman@gatech.edu

Sung Kyu Lim
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, USA
limsk@ece.gatech.edu

***Abstract*—One of the advantages of 3D IC technology is its ability to integrate different devices such as CMOS, SRAM, and RRAM, or multiple technology nodes of single or different devices onto a single chip due to the presence of multiple tiers. This ability to create heterogeneous 3D ICs finds a wide range of applications, from improving processor performance by integrating better memory technologies to building compute-in-memory ICs to support advanced machine learning algorithms. This paper discusses the current trends and future directions for the physical design of heterogeneous 3D ICs. We summarize various physical design and optimization flows, integration techniques, and existing academic works on heterogeneous 3D ICs.**

***Index Terms*—heterogeneous 3D IC, 3D physical design, 3D integration, macro3D, pin3D**

I. Introduction

For several decades, Moore's law has flourished through CMOS technology's continued scaling reaching nanometer dimensions. After reaching the pinnacle of CMOS scaling, the focus shifted towards 3D IC integration to further facilitate giga-scale integration. However, not all components of a typical IC scale down at the same pace. Circuit components such as memories and analog blocks tend to scale at a slower pace than their logic counterpart. This becomes a bottleneck in 2D/3D integration, restricting the further growth of giga-scale integration.

With memory-intensive applications such as machine learning and computer vision proliferating, the need to improve the memory capacity and bandwidth of modern processors is more crucial than ever. Memories such as Spin Torque Transfer Magnetic RAM (STT-MRAM) and Resistive RAM (RRAM) have shown better performance than CMOS memories for such applications. Integrating them with current processor systems can provide significant performance improvements. However, the advancements across multiple device technologies are not similar, restricting us from integrating them. 2.5D integration technique provides an acceptable means to build such heterogeneous systems by combining chiplets of different blocks on various technology nodes on an interposer. However, the large footprint, longer wirelength, and huge wire parasitics may degrade the system's performance rather than improving it.

We need integration techniques that facilitate building heterogeneous systems in a single chip to overcome the issue of 2.5D designs. Thanks to the concept of heterogeneous 3D integration, we can now integrate different devices and devices of different technology nodes onto a single IC. This way, modern systems can benefit by using recent technology nodes for standard cells without worrying about the current technology node of memory blocks or other hard IPs. Several such works have been presented in the academia. Lingjun et. al designed a heterogeneous 3D design of a single tile of OpenPiton RISC-V architecture [1] with STT-MRAM in [2]. The heterogeneous 3D Openpiton is 34.7% faster, has 17.6% lower silicon area, 21.7% smaller wirelength, and 58.8% smaller footprint than the 2D counterpart. A heterogeneous monolithic 3D CIM architecture is presented in [3], with the RRAM cells and transistors designed at 40nm technology node on the top tier, and the digital periphery and ADCs designed at lower technology nodes of 28/16nm on the bottom tier. The heterogeneous integration improves overall energy efficiency of CIM tiles by 3.4x compared to its 2D counterpart designed at 40nm.

The field of designing heterogeneous 3D ICs has been growing in recent years due to their potential to bring drastic performance improvements in hardware systems needed to develop many futuristic applications. Significant design challenges are faced during the physical design stages of heterogeneous 3D ICs. We do not have commercial layout designing tools that support multiple technology nodes for a single IC design. Several works are being performed in academia to leverage existing tools and build new tools for designing such heterogeneous systems. This paper discusses various design flow methodologies, optimization flow, and integration techniques for heterogeneous 3D IC design developed so far, and open problems still being faced by physical design methodologies for heterogeneous 3D ICs.

The rest of the paper is organized as follows: The placement and routing flows for heterogeneous 3D integration are presented in section II, methods to improve timing and power performance, and signal and power integrity are discussed in Section III, Machine learning (ML) based techniques to

This research is partially funded by the DARPA ERI 3DSOC Program under Award HR001118C0096, the Semiconductor Research Corporation under Task 2929, and the National Research Foundation of Korea under NRF-2020M3F3A2A02082445.

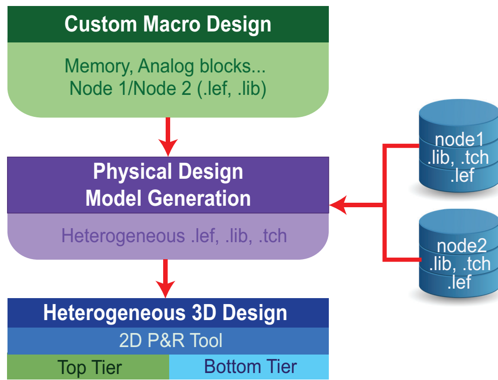

Fig. 1. Technology file generation for heterogeneous 3D ICs

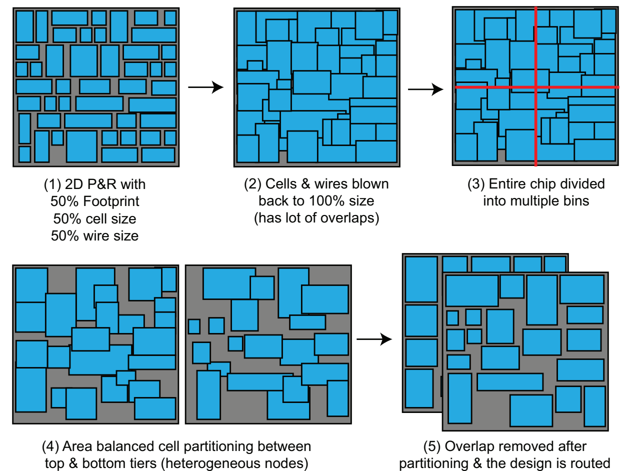

Fig. 2. Illustration of Shrunk2D design methodology for Heterogeneous 3D ICs

improve the performance and quality of 3D designs are presented in Section IV, potential heterogeneous 3D integration techniques are summarized in Section V, and finally, future directions for physical design methodology of heterogeneous 3D ICs are discussed in Section VI.

II. Design Flows for Heterogeneous 3D ICs

In this section, we explain how to generate technology files for performing the physical design of heterogeneous 3D ICs, how traditional homogeneous pseudo3D flows can be used for heterogeneous 3D IC designs containing just standard cells. We also summarize a macro-on-logic (MoL) 3D flow that helps design heterogeneous 3D IC designs with hard macros such as memories and analog blocks.

A. Technology File Generation

The physical design of heterogeneous 3D ICs requires heterogeneous 3D technology files during certain stages of the design flow, as shown in Fig. 2 and 3. These heterogeneous technology files are generated by combining technology files of different nodes involved in the design. The technology file generation flow for heterogeneous 3D ICs is shown in Fig. 1. The physical layout information of various cells, the front-end-of-line (FEOL), and the back-end-of-line (BEOL) of a particular node are contained in library exchange format (LEF) files. The cells' timing and power information are available in liberty (lib) files, and the parasitic (resistance and capacitance) details of a technology node are available in .tch files. The LEF files of various technology nodes are merged to create a heterogeneous 3D FEOL and BEOL stack. Similarly, .lib and .tch files are combined to generate 3D timing, power, and parasitic information. These files are input to P&R tools during physical design stages involving the entire 3D design. We explain different heterogeneous 3D design flows in the following subsections and how these 3D technology files are used in them.

B. Shrunk2D

Shrunk2D [4] is a pseudo3D flow in which standard cells and interconnects are shrunk to half of their area to first create a 2D design with a footprint equal to that of the target 3D design, using commercial 2D place and route (P&R) tool. Then, the 2D design is transformed into a 3D design by partitioning the cells using FM mincut algorithm [5] across two tiers, expanding cells to their original sizes, and then routing them tier-by-tier. The locations of inter-tier vias(monolithic inter-tier via (MIV) or face-to-face (F2F) via) are then fixed by using a via-planning stage, and the final routing is performed. Though this flow has been largely used for homogeneous 3D IC design, it is also equally capable of performing heterogeneous 3D IC designs.

To design heterogeneous 3D ICs using Shrunk2D flow, the initial 2D P&R is done using a single technology node. This is because 2D P&R tools cannot support multiple technology nodes in a single design. The FM mincut partitioning stage is slightly modified for heterogeneous designs to accommodate the area difference between the two different technology nodes. This way, the total area of both the tiers is balanced, and the difference is within a tolerable range. After partitioning, the individual dies are routed using their corresponding technology node files, followed by MIV/F2F-via planning and then rerouting the designs. The illustration of Shrunk2D design flow for heterogeneous 3D ICs is shown in Fig. 2.

C. Compact2D

An issue with Shrunk2D is its need to shrink both cells and routing geometries to half of their original area. For recent technology nodes, 2D P&R engines capable of handling future technology nodes are required. Compact2D [6] is pseudo3D flow that does not need any geometry shrinking. In compact2D, a 2D design is first implemented with RC parasitics of the interconnects scaled by a factor of $\sqrt{2}$. Then the 2D design is contracted to the desired footprint size for the 3D IC. Also, compact2D performs tier-by-tier post-layout optimization, thereby providing better timing closure and power optimization than shrunk2D designs. The design flow of compact2D post the initial 2D design phase is the same as that of shrunk2D flow. Fig. 3 shows the illustration of compact2D design flow for heterogeneous 3D ICs. Heechun et al. discuss

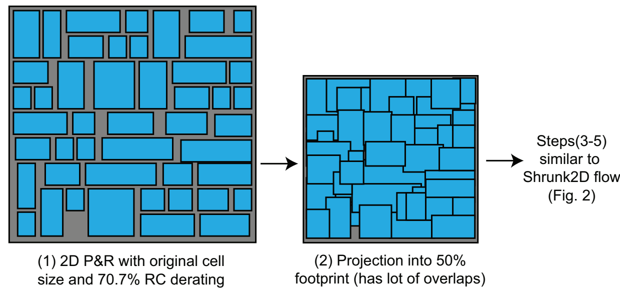

Fig. 3. Illustration of Compact2D design methodology for Heterogeneous 3D ICs

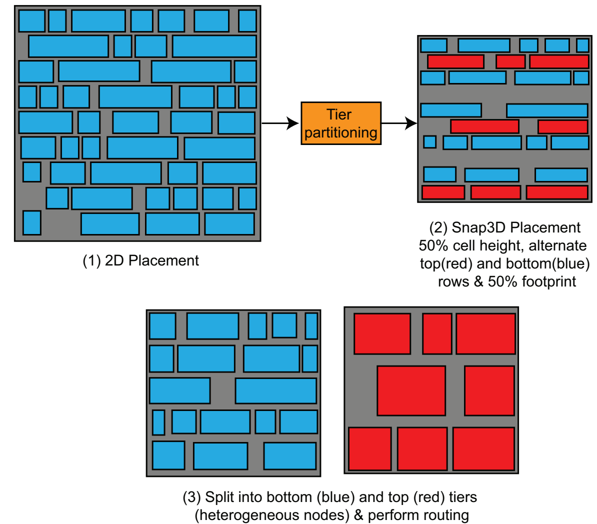

Fig. 4. Illustration of Snap3D design methodology for F2F Heterogeneous 3D ICs

the benefits of these pseudo3D flows (shrunk2D, compact2D, and cascade3D) for homogeneous monolithic 3D IC design in detail in [7]. On performing modifications to these flows as mentioned above, the benefits of pseudo3D flows can be observed in heterogeneous 3D IC designs as well.

D. Snap-3D

Pseudo3D tool flows estimate the parasitics of cell pins based on the assumption that all these pins are located on the same BEOL. However, post tier partitioning, the pins are located on different tiers. Failure to account for this difference in pin location leads to timing degradation and sub-optimal 3D designs. Iampikul et al. present a 3D design flow called Snap-3D in [8] to overcome this issue. In Snap-3D, a given 2D netlist is partitioned into two, the height of the cells is reduced by half, and a commercial 2D P&R tool is used to perform a pseudo3D placement on these half-sized cells. The placement rows are divided into top and bottom rows to enable simultaneous placement of cells in both tiers. In the case of designs with hardmacros additional placement blockages are added. Then pseudo-3D signal routing and timing closure are performed using a full-stack 3D BEOL as described in Section II A. Then the pseudo-3D design is split into two tiers, the cells are scaled back to the original sizes, final signal routing is performed, followed by timing and power optimization and design signoff. The illustration of snap3D design flow for heterogeneous 3D IC design is shown in Fig. 4

E. Macro3D - Macro on Logic Integration

Bamberg et al. in [9] propose a 3D integration flow based on commercial P&R tools for face-to-face stacked heterogeneous 3D ICs containing large hard macros called Macro3D. Macro3D is a macro/memory on logic (MoL) integration flow for two-tier designs, in which the top-tier is completely filled with hard macros and the bottom-tier has a mix of logic (standard cells) and hard macros. Even though Macro3D is intended for face-to-face (F2F) stacking, the concept can be extended to monolithic 3D ICs as well. Macro3D integration consists of four stages, as shown in Fig. 5. In the first stage, 2D floorplans of individual tiers are generated. The top-tier or macro-tier only consists of hard macros, and the bottom-tier or logic-tier consists of both hard macros and logic cells. The footprint of each 2D floorplan is equal to the footprint of the 3D design. After floorplanning, individual dies are stacked together in the second stage. This is achieved by shrinking the macros on the macro-tier to the minimum possible rectangular area of the technology node, called the site size, and then superimposing the shrunk macros on the logic-tier. However, the pin locations of the shrunk macros are retained at the original position. For this to be possible, Macro3D uses a full 3D-stack BEOL structure, as described in Section II A, containing both bottom and top tier metal layers along with inter-tier vias (Monolithic Inter-tier Vias in case of Monolithic 3D ICs and F2F vias in case of F2F stacked 3D ICs). This way, 2D commercial tools are tweaked to perform full 3D routing. In the third step, the super-imposed floorplan along with fully stacked 3D BEOL is fed to commercial 2D P&R tools to perform clock and signal routing and design optimizations. In the final step, the super-imposed designs are separated into individual dies, and the layouts of individual tiers are generated.

F. Pin-in-the-middle: Block-level 3D Integration

Many commercial processor designs involve block-level implementation for effective performance. In a 3D block-level integration, placing pins along the blocks' edges leads to longer inter/intra-block interconnects. As 3D integration allows vertical interconnections, the block-level pins need not necessarily be placed along the block boundaries. Placing these pins in the middle of the blocks helps significantly reduce the length of both inter and intra-block interconnects as shown in Fig. 6. Ku et al. propose a block-level 3D integration flow called Pin-in-the-middle in [10]. The flow takes a 2D netlist as an input, restructures into optimal blocks, and optimally assigns the blocks to different tiers. Then, a wirelength-driven 3D placement is performed, which optimally places the blocks and their pins in the middle of the corresponding blocks to improve the overall wirelength. Timing budget is then performed in which the wirelength savings provide additional block-level

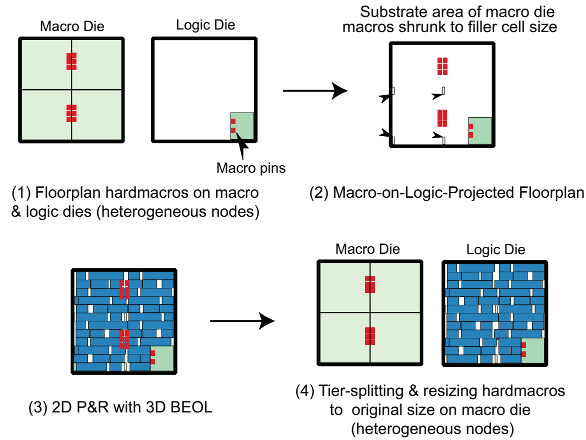

Fig. 5. Macro3D design methodology for F2F Heterogeneous 3D ICs

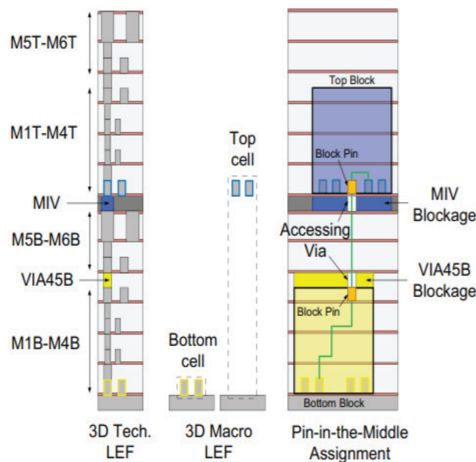

Fig. 6. Pin in the middle assignment (adopted from [10])

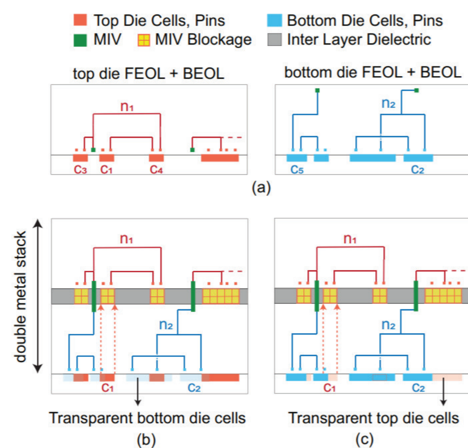

Fig. 7. Illustration of Pin3D flow (adopted from [11])

timing margin. Finally, the individual blocks are implemented, assembled on different tiers, and 3D timing closure and signoff is performed.

III. Optimizing Heterogeneous 3D ICs

In this section, we describe few optimization flows that can be used to improve the quality and PPA of heterogeneous 3D ICs.

A. Timing and Power Optimization

To sign off an IC design, both timing closure and power optimizations are critical. While post-layout optimization of homogeneous 3D ICs has been a target of research for a long time, for the first time, Pin3D [11] optimizer, based on 2D commercial P&R tools, has been shown to perform post-layout optimization of heterogeneous 3D IC designs designed out of pseudo3D flows. Pin3D places the cells of both top and bottom tiers together and uses a 3D routing stack similar to Macro3D. While optimizing the top tier, the placement area of bottom tier cells is made transparent. Only their pin locations are retained at the bottom tier and marked as fixed, allowing only the movement of top-tier logic cells. In contrast, the opposite occurs while optimizing the bottom-tier. This is illustrated in Fig. 7. The placement rows are updated according to the technology node of the tier being optimized. This way, 2D P&R tools are tricked that only one tier of cells (one FEOL) is to be optimized while having access to timing and power information of both cells and nets of both the tiers.

The clock routed placement result of any pseudo3D flow is fed as an input to the Pin3D flow. The cells are placed as described above, and the placement locations are legalized. The entire design is then routed using 3D BEOL with MIVs, and then post-routing optimization is performed to close design timing. As the tool has complete information on clock skews in both inter-tier and intra-tier paths, it is easier for pin3D flow to optimize the entire design for better performance. Also, as the entire design is loaded into the P&R tool, it is easier to use the tool's capability to perform engineering change of order (ECO) optimization for heterogeneous 3D ICs using this flow. Thus, pin3D provides a standard means to optimize heterogeneous 3D designs.

B. Signal and Power Integrity

One of the significant challenges in a heterogeneous 3D IC is to achieve better signal and power integrity. When multiple nodes are integrated together, the interconnect dimensions and power domains of the different nodes vary. Accurate parasitic extraction of combined heterogeneous design is the key to achieving signal and power integrity in heterogeneous 3D designs. It is shown in [12] that in F2F bonded 3D designs, ignoring inter-die parasitic coupling by performing tier-by-tier parasitic extraction can underestimate the total coupling capacitance by 35%. Hence, a holistic extraction considering all the tiers and inter-tier connections is necessary for better parasitic extraction and, therefore, better signal integrity. However, holistic extraction considering all the tiers together

is computationally challenging and time-consuming, especially when multiple technology nodes are involved. Peng et al. propose a technique called in-context extraction in [12] as an alternative. In-context extraction is similar to tier-by-tier parasitic extraction, but both the tiers are extracted with the knowledge of the interface layers. This drastically reduces the number of pre-calibrated structures involved in the computation. Especially with advanced nodes, which involve multiple design rules on small structures, in-context extraction ignores such rules. Still, it performs high-quality parasitic extraction with just 0.8% error in estimating total coupling capacitance and 0.9% error in estimating total ground capacitance. With such high-quality parasitic extraction, commercial P&R tools can be used to improve the signal and power integrity of heterogeneous 3D designs. To handle multiple power domains, appropriate level shifters are needed to improve the power integrity further.

IV. Machine Learning based heterogeneous 3D Design Flows

Physical design flows involve several parameters that can be tuned to optimize the design for better wirelength and PPA. However, manually tuning all the parameters can be a herculean task. Applying machine learning concepts to physical design helps reduce human efforts in tuning thousands of parameters and guarantees to provide highly optimized 3D designs. This section discusses two such ML-based frameworks enabling high-quality physical design flows for heterogeneous 3D ICs.

A. TP-GNN: ML-based Tier Partitioning

In pseudo3D flows presented in Section II, the bin-based tier-partitioning fails to consider global connections among bins leading to timing degradation. Min-cut based partitioning does not always give the best 3D integration. Also, as the partitioner completely ignores the RTL hierarchy information, such partitioning can lead to degraded 3D placement quality. Lu et al. propose a graph neural network (GNN) framework called TP-GNN for tier-partitioning in [13] to address these drawbacks in bin-based/min-cut based tier partitioning. As shown in Fig. 8, the GNN framework takes a projected 2D design as the input, creates a netlist hypergraph, and transforms it into an edge-contracted clique-based graph using a hierarchy-aware edge contraction algorithm. These instance-based graph representations are learned using GNNs. The features within a specific hop neighborhood of the target node are sampled and aggregated to learn accurate representations for the downstream clustering stage. The GNN based clustering then provides a high-quality tier-partitioned result, which can be subjected to the regular 3D P&R flow. Yu et al. demonstrated an improvement of 27.4% in the effective frequency of Openpiton designed using Shrunk2D flow and TP-GNN framework.

B. ML-based Wire RC Prediction

As mentioned earlier, the accuracy with which parasitics are predicted in a 3D IC plays a significant role in achieving signal and power integrity and timing closure. Pentapati et al. in [14] present a regression model based on boosted decision tree

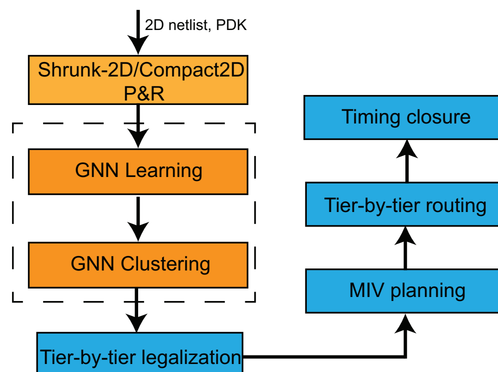

Fig. 8. Graph neural network (GNN) flow for tier-partitioning in pseudo3D design flows.

learning combined with pseudo3D flows to better predict the wire parasitics (RCs) in the final 3D design stage for achieving high-quality 3D designs. The net features are extracted from the projected 2D designs (pseudo3D) and passed to an ML model that predicts the RC values of 3D nets. The nets are annotated with these predicted parasitics, and then the usual optimization flow is performed. The optimization flow changes the netlist due to the addition of buffers. Hence, the RC prediction model is again run to re-extract net features and predict 3D RC values, and the nets are re-annotated. An incremental in-place optimization flow is run, which does not change the existing net structure (or rather a minimum change). Following this, the design is subjected to the usual tier-partitioning, and 3D routing flows with better design signoff. The ML-based RC prediction model provides an accuracy of up to 98.6%.

V. Integrating Heterogeneous 3D ICs

Though heterogeneous 3D ICs are still in evolution stages, we have reliable integration methodologies that can be used to implement them.

- Leti's 3D sequential integration technology called CoolCube™ allows stacking of transistors on top of each other with nanoscale resolution. It allows lower aspect ratios and small 3D contact fine-grain interconnects. However, the thermal budget of the top tier has to be limited to 500 °C to maintain the stability of bottom tier devices.
- Foveros [15] is a 3D IC packaging technology introduced by Intel that enables F2F stacking of heterogeneous dies. F2F chip-on-chip bonding is achieved through fine-pitched micro-bumps of diameter 36 microns each. FOVEROS consists of a base logic die that is also an active interposer with TSVs, to distribute signals and power, and a platform controller hub (PCH) to manage I/O signals. Additional active components such as a logic-die, a memory-die or RF/analog-die are placed on top of the base logic die with the help of F2F micro-bumps to leverage their advantages of easy interconnect density scaling and lower wire parasitics.

VI. Future Directions

A. 3D Placers

The physical design technology for heterogeneous 3D ICs is emerging to be efficient and successful. However, the closest we have come to true heterogeneous 3D placement is with MoL designs in macro3D flow, where the macro placements are fixed on macro-tier, and we perform logic cell placement on logic-tier. Nevertheless, we still do not have 3D synthesis and placement algorithms that support heterogeneous technology nodes. We still use 2D placement tools and pseudo3D techniques to implement 3D designs containing logic cells on both tiers. There are several 3D placement algorithms for homogeneous 3D designs such as Force-directed 3D [16], Non-linear 3D [17], ePlace-3D [18], and TSV-aware 3D [19]. These algorithms can be extended to handle multiple technology nodes by incorporating appropriate area balancing techniques while moving cells across different tiers to account for the difference in the standard cells areas of different technology nodes. Enabling true-3D placement for heterogeneous 3D ICs facilitates easy and efficient designing of monolithic 3D ICs with multiple nodes, as the MIV count in the design can be optimized for better wirelength and performance of the system.

B. Signal Integrity and Thermal Optimization

3D ICs, in general, are more prone to signal integrity and heating issues due to the stacking of multiple layers of devices on top of each other. Heterogeneous 3D ICs stacking memories and analog circuitry on top of logic cells further add to the heating issues. So, optimizing heterogeneous 3D ICs for signal integrity and thermal issues is critical for their stability and functionality. Chen et al. propose a heterogeneous 3D integration technique in [20], which helps achieve smaller form factor and shorter tier-to-tier connections. This helps in improving the signal and power integrity of 3D designs. To address thermal issues in heterogeneous 3D ICs, an air gap based thermal isolation technique has been proposed in [21]. Zhang et al. propose a solution combining interposer embedded heat sink, thermal bridge, and air gap isolation technique to address the thermal issues. These methods try to address signal integrity and thermal issues in heterogeneous 3D ICs during the integration stage. However, floorplanning, placement, number and nature of inter-tier vias, alignment of vias on different metal layers to facilitate thermal conduction from the bulk of the IC to the heat sink all play a significant role in affecting the signal integrity and thermal behavior of ICs. We need algorithms for these stages of the physical design flow, considering the signal integrity and thermal behavior across different technology nodes and thermal and parasitic coupling between different tiers to achieve a better performance in heterogeneous 3D ICs.

VII. Conclusion

In this paper, we explored the applications, summarized different placement, routing, and optimization flows for designs with and without hardmacros, various potential IC integration schemes, and finally discussed possible future directions for physical design methodologies of heterogeneous 3D ICs. Complete systems built using heterogeneous 3D integration show excellent performance benefits. Further improving existing physical design and optimization flows to build heterogeneous 3D ICs can lead to designing better performing processor architectures for a wide range of applications.

References

[1] J. Balkind *et al.*, “Openpiton: An open source manycore research framework,” in *International Conference on Architectural Support for Programming Languages and Operating Systems*, 2016.

[2] L. Zhu *et al.*, “Heterogeneous 3d integration for a risc-v system with stt-mram,” *IEEE Computer Architecture Letters*, 2020.

[3] G. Murali, X. Sun, S. Yu, and S. Lim, “Heterogeneous mixed-signal monolithic 3d in-memory computing using resistive ram,” *IEEE Transactions on VLSI Systems, in press*, 2020.

[4] S. Panth, K. Samadi, Y. Du, and S. K. Lim, “Shrunk-2-d: A physical design methodology to build commercial-quality monolithic 3-d ics,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2017.

[5] C. M. Fiduccia and R. M. Mattheyses, “A linear-time heuristic for improving network partitions,” in *Design Automation Conference*, 1982.

[6] 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 *International Symposium on Physical Design*, 2018.

[7] H. Park, B. W. Ku, K. Chang, D. E. Shim, and S. K. Lim, “Pseudo-3d approaches for commercial-grade rtl-to-gds tool flow targeting monolithic 3d ics,” in *ACM International Symposium on Physical Design*, 2020.

[8] P. V. -Iampikul, C. Shao, Y. C. Lu, S. S. Kiran Pentapati, and S. K. Lim, “Snap-3d: A constrained placement-driven physical design methodology for face-to-face-bonded 3d ics,” in *ACM International Symposium on Physical Design*, 2021.

[9] L. Bamberg, A. García-Ortiz, L. Zhu, S. Pentapati, D. E. Shim, and S. Kyu Lim, “Macro-3d: A physical design methodology for face-to-face-stacked heterogeneous 3d ics,” in *Design, Automation Test in Europe Conference Exhibition*, 2020.

[10] B. W. Ku and S. K. Lim, “Pin-in-the-middle: An efficient block pin assignment methodology for block-level monolithic 3d ics,” in *ACM/IEEE International Symposium on Low Power Electronics and Design*, 2020.

[11] S. S. K. Pentapati, K. Chang, V. Gerousis, R. Sengupta, and S. Lim, “Pin-3d: A physical synthesis and post-layout optimization flow for heterogeneous monolithic 3d ics,” 2020.

[12] Y. Peng, T. Song, D. Petranovic, and S. K. Lim, “Parasitic extraction for heterogeneous face-to-face bonded 3-d ics,” *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2017.

[13] Y. C. Lu, S. S. Kiran Pentapati, L. Zhu, K. Samadi, and S. K. Lim, “Tp-gnn: A graph neural network framework for tier partitioning in monolithic 3d ics,” in *ACM/IEEE Design Automation Conference*, 2020.

[14] S. S. Kiran Pentapati, B. W. Ku, and S. K. Lim, “ML-based wire rc prediction in monolithic 3d ics with an application to full-chip optimization,” in *ACM International Symposium on Physical Design*, 2021.

[15] D. B. Ingerly *et al.*, “Foveros: 3d integration and the use of face-to-face chip stacking for logic devices,” in *IEEE International Electron Devices Meeting (IEDM)*, 2019.

[16] D. H. Kim, K. Athikulwongse, and S. K. Lim, “Study of through-silicon-via impact on the 3-d stacked ic layout,” *IEEE Transactions on Very Large Scale Integration Systems*, 2013.

[17] G. Luo, Y. Shi, and J. Cong, “An analytical placement framework for 3-d ics and its extension on thermal awareness,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2013.

[18] J. Lu, H. Zhuang, I. Kang, P. Chen, and C.-K. Cheng, “Eplace-3d: Electrostatics based placement for 3d-ics,” in *International Symposium on Physical Design*, 2016.

[19] M. Hsu, Y. Chang, and V. Balabanov, “Tsv-aware analytical placement for 3d ic designs,” in *ACM/EDAC/IEEE Design Automation Conference*, 2011.

[20] M. Chen, F. Chen, W. Chiou, and D. C. H. Yu, “System on integrated chips (soic(tm)) for 3d heterogeneous integration,” in *IEEE Electronic Components and Technology Conference (ECTC)*, 2019.

[21] Y. Zhang, T. E. Sarvey, and M. S. Bakir, “Thermal challenges for heterogeneous 3d ics and opportunities for air gap thermal isolation,” in *International 3D Systems Integration Conference (3DIC)*, 2014.