Routing Congestion Aware Cell Library Development for Monolithic 3D ICs

Chen Yan and Emre Salman
Department of Electrical and Computer Engineering
Stony Brook University (SUNY), Stony Brook, NY 11794
Email: {chen.yan,emre.salman@stonybrook.edu}

***Abstract*—According to International Roadmap for Devices and Systems (IRDS), after 2024, there is no headroom for 2D geometry scaling. By 2024, IRDS predicts that monolithic 3D integration technology will be one of the most critical performance boosters. This prediction follows the highly promising and relatively recent improvements on the fabrication of second-tier devices on a single substrate. In this study, a fully functional and open source cell library is developed to design and characterize large-scale monolithic 3D ICs. The second version of the 3D cell library is equipped with the required files for integration into existing design automation tools, thereby enabling chip-level benchmarking. Furthermore, multiple versions of the library are proposed to investigate the tradeoffs among routability, timing, power, and area characteristics. The 3D library can also be used to analyze some of the important issues in monolithic 3D ICs, such as chip-level thermal characteristics, efficacy of various thermal management methodologies, and design-for-test methods for monolithic 3D integration.**

I. Introduction

Three-dimensional (3D) integrated circuits (ICs) have emerged as an effective solution to some of the critical issues encountered in planar technologies due to promising characteristics in reducing global interconnect delay, increasing transistor density, and enabling heterogeneous integration [1], [2]. Monolithic 3D integration enables significantly higher interconnect density as compared to other vertical integration methods. Monolithic inter-tier vias (MIVs) have comparable size to conventional on-chip metal vias, provided by the high alignment precision and thin top layer [3].

Transistor-, gate-, and block-level design methods have been proposed for monolithic 3D integration [4]. In transistor-level monolithic 3D integration, as focused in this paper, nMOS and pMOS transistors within a circuit are separated into two different tiers, as depicted in Fig. 1. This approach not only achieves fine-grained 3D integration with intra-cell MIVs, but also enables the individual optimization of the bottom and top tier devices.

In this study, multiple versions of a cell library for transistor-level monolithic 3D integration are developed. The power and timing characteristics of each cell within each library are fully characterized with both SPICE-level simulations and a commercial library characterization tool to ensure accuracy. The effects of the number of routing tracks on area, power, and delay characteristics are investigated by developing three versions of the cell library with different cell heights. This analysis is important since routing congestion is one of the primary physical design issues in monolithic 3D ICs.

This research is supported by the National Science Foundation CAREER grant under Grant CCF-1253715.

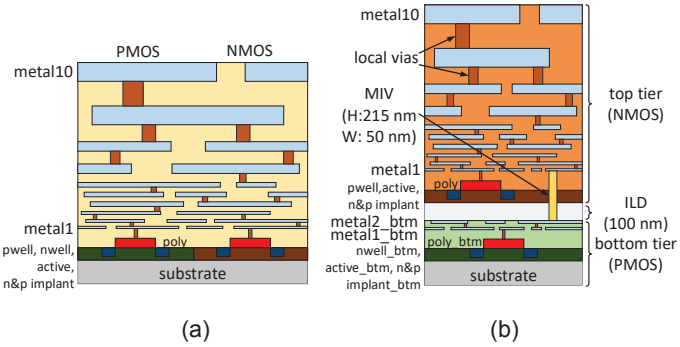

Fig. 1. **Cross-sections of the (a) conventional 2D and (b) transistor-level monolithic (TL-Mono) 3D technology where the top tier hosts nMOS transistors whereas pMOS transistors are placed within the bottom tier.**

The rest of the paper is organized as follows. Contributions of this paper are summarized in Section II. The details of the proposed open source cell libraries, characterization, and comparison with 2D cells are provided in Section III for each version. The effects of the number of routing tracks on congestion, power, and timing characteristics are investigated in Section IV by developing a large-scale 3D FFT core. Finally, the paper is concluded in Section V.

II. Contributions of this Paper

An open source cell library for monolithic 3D ICs has recently been developed [5] and is publicly available [6], facilitating future research on monolithic 3D technology [7], [8]. In this paper:

- Additional versions of the 3D monolithic cell library with different cell heights are developed and fully characterized.
- These additional libraries are utilized to investigate the effects of number of tracks on chip-level routing congestion (e.g., number of interconnect related design rule check violations), power, and timing (worst negative and total negative slack) characteristics.
- It is demonstrated that an optimum number of routing tracks exists to minimize chip-level power and delay characteristics in monolithic 3D ICs.

978-1-5386-1553-9/17/$31.00 ©2017 IEEE

III. Cell Library for Monolithic 3D ICs

The characteristics of the proposed cell library and the design flow to integrate the proposed library into the design process are summarized in Section III-A. Cell-level simulation results and comparison of 3D cells (in each version of the library) with 2D cells are provided in Section III-B.

A. Routing Congestion Aware Cell Library

In this work, the previously developed standard cell library for transistor-level monolithic 3D technology (*Mono3D*) is extended to consider different number of routing tracks. *Mono3D* consists of two tiers where each tier is based on the 2D 45 nm process design kit *FreePDK45* from North Carolina State University (NCSU) [9]. Thus, the process and physical characteristics (transistor models and characteristics of the on-chip metal layers) are obtained from the *FreePDK45*. Similar to [10], [11], the pull-down network of a CMOS gate (nMOS transistors) is built within the top tier whereas the pull-up network (pMOS transistors) is fabricated within the bottom tier. Note that the processing temperature of the top tier is constrained to be less than 500-600°C to not damage the transistors within the bottom tier [12]. This relatively low processing temperature, however, degrades the quality of the top tier devices. Thus, pMOS devices (that already have lower mobility) are placed within the bottom tier. The nMOS and nMOS device characteristics are the same as in 2D *FreePDK45*. However, the impact of novel devices and manufacturing steps for 3D monolithic integration can be captured by replacing/modifying the device models within the provided design kit. System-level effects of varying device characteristics can therefore be investigated.

In the proposed *Mono3D*, two metal layers are allocated to the bottom tier (metal1_btm and metal2_btm), as illustrated in Fig. 1. These metal layers are primarily for routing the intra-cell signals. The top tier is separated from the bottom tier with an inter-layer dielectric (ILD) with a thickness of 100 nm. Inter-tier coupling is minimized at this thickness, as experimentally validated [3]. All of the other parasitic components (including MIV) can be extracted by the modified extraction rule files. The 10 metal layers that exist in 2D *FreePDK45* are maintained the same for the top tier in *Mono3D*. The intra-cell connections that span the two tiers are achieved by MIVs. Each MIV has a width of 50 nm and height of 270 nm [4].

Currently, 20 standard cells exist in *Mono3D*, as listed in Table I. In addition to the fundamental cells, multiple clock buffers and a latch are included. Each cell is developed with a full-custom design methodology using a cell stacking technique. A specific track is allocated for intra-cell MIVs, which are distributed within the cell to minimize the interconnect length and reduce the cell height. Each cell within the 2D *NanGate* library has 14 routing tracks. Alternatively, in this study, three monolithic 3D cell libraries are developed with different number of tracks: 8-track (*Mono3D_v1*), 9-track (*Mono3D_v2*), and 10-track (*Mono3D_v3*). Number of tracks plays an important role on chip-level routing congestion, a primary issue in monolithic 3D ICs (see Section IV for more details). The cell heights in *Mono3D_v1*, *Mono3D_v2*, and *Mono3D_v3* are, respectively, 1.33 μm, 1.52 μm, and 1.71 μm. These cell heights are, respectively, 46%, 38%, and 31% shorter than the standard cell height (2.47 μm) in *NanGate* cell library [13].

TABLE I
List of standard cells in the monolithic 3D library.

AND2X1	INVX2
AOI21X1	INVX4
BUFX2	LATCHNEG
BUFX4	MUX2X1
CLKBUF1	NAND2X1
CLKBUF2	NOR2X1
CLKBUF3	OAI21X1
DFFPOSX1	OR2X1
FILL	XNOR2X1
INVX1	XOR2X1

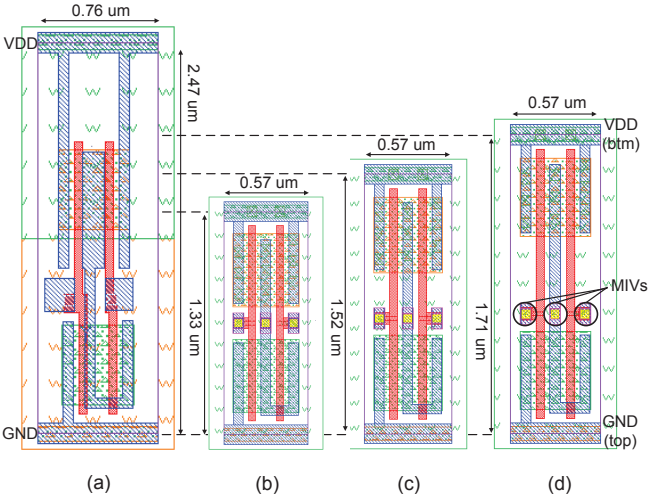

Fig. 2. **Comparison of the layout views of a NAND gate in (a) traditional 2D technology with 14 routing tracks in each cell, (b) monolithic 3D technology with 8 routing tracks, (c) monolithic 3D technology with 9 routing tracks, and (d) monolithic 3D technology with 10 routing tracks, illustrating the three MIVs used to connect the top and bottom tiers.**

The layout of a NAND cell is illustrated in Fig. 2 in both 2D and 3D monolithic technologies with three different cell heights. Cell dimensions and the three MIVs are highlighted. The design flow and the modifications required for 3D monolithic technology are described in [5].

B. Cell-Level Evaluation

1) Area: Cell-level area reduction varies from 6.5% to 64.1% in *Mono3D_v1*, -6.9% to 59.0% in *Mono3D_v2*, and -13.5% to 53.8% in *Mono3D_v3*, depending upon the specific cell. An average improvement of 32%, 22%, and 14% is achieved for, respectively, *Mono3D_v1*, *Mono3D_v2*, and *Mono3D_v3*. Note that a negative percent implies that the cell area increases as compared to the 2D cell. This behavior occurs for cells where the reduction in cell height causes a considerable increase in cell width. Similarly, the average area reduction is not as large as the reduction in cell height since, on average, the cell width slightly increases due to MIVs and intra-cell routing within the reduced cell footprint.

TABLE II
AVERAGE DELAY AND POWER CHARACTERISTICS OF 2D AND MONOLITHIC 3D CELLS WITH 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), AND 10 (*Mono3D_v3*) ROUTING TRACKS. THE PERCENT CHANGES WITH RESPECT TO 2D CELLS ARE LISTED.

Cells	**Delay (ps)**				**Power (μW)**			
	2D	3D_v1	3D_v2	3D_v3	2D	3D_v1	3D_v2	3D_v3
AND2X1	17.63	19.27 (9.3%)	19.5 (10.6%)	19.52 (10.7%)	2.82	2.98 (5.7%)	3.01 (6.7%)	3.03 (7.4%)
AOI21X1	13.68	13.58 (-0.7%)	13.69 (0.1%)	13.80 (0.9%)	3.32	3.33 (0.3%)	3.34 (0.6%)	3.35 (0.9%)
BUFX2	17.89	17.56 (-1.8%)	17.86 (-0.2%)	17.94 (0.3%)	14.04	13.71 (-2.4%)	13.92 (-0.9%)	13.97 (-0.5%)
BUFX4	15.82	14.97 (-5.4%)	15.29 (-3.4%)	15.76 (-0.4%)	29.00	28.97 (-0.1%)	28.98 (-0.1%)	29.14 (0.5%)
CLKBUF1	27.01	27.28 (1.0%)	27.32 (1.2%)	27.34 (1.2%)	64.07	62.17 (-3.0%)	62.52 (-2.4%)	62.84 (-1.9%)
CLKBUF2	39.57	40.05 (1.2%)	40.17 (1.5%)	40.24 (1.7%)	93.25	90.43 (-3.0%)	90.88 (-2.5%)	91.05 (-2.4%)
CLKBUF3	51.73	52.74 (2.0%)	53.04 (2.5%)	53.27 (3.0%)	121.4	118.6 (-2.3%)	119.0 (-2.0%)	119.1 (-1.9%)
DFFPOSX1	41.69	34.54 (-17.2%)	34.72 (-16.7%)	34.93 (-16.2%)	26.75	27.13 (1.4%)	27.18 (1.6%)	27.39 (2.4%)
INVX1	6.73	6.10 (-9.4%)	6.42 (-4.6%)	6.50 (-3.4%)	4.69	4.64 (-1.1%)	4.68 (-0.2%)	4.72 (0.6%)
INVX2	6.54	6.08 (-7.0%)	6.32 (-3.4%)	6.40 (-2.1%)	9.31	9.15 (-1.7%)	9.24 (-0.8%)	9.35 (0.4%)
INVX4	6.44	6.08 (-5.6%)	6.29 (-2.3%)	6.38 (-0.9%)	18.01	17.99 (-0.1%)	18.16 (0.8%)	18.36 (1.9%)
MUX2X1	16.25	17.21 (5.9%)	17.23 (6.0%)	17.27 (6.2%)	5.81	6.14 (5.7%)	6.15 (5.9%)	6.17 (6.2%)
NAND2X1	10.22	9.76 (-4.5%)	9.78 (-4.3%)	9.89 (-3.2%)	1.63	1.57 (-3.7%)	1.58 (-3.1%)	1.61 (-1.2%)
NOR2X1	11.41	11.78 (3.2%)	12.16 (6.6%)	12.18 (6.8%)	1.63	1.66 (1.8%)	1.69 (3.7%)	1.73 (6.1%)
OAI21X1	12.89	12.72 (-1.3%)	12.88 (-0.1%)	12.93 (0.3%)	3.27	3.24 (-0.9%)	3.25 (-0.6%)	3.26 (-0.3%)
OR2X1	18.33	20.89 (14.0%)	21.12 (15.2%)	21.14 (15.3%)	2.54	2.84 (11.8%)	2.85 (12.2%)	2.87 (12.9%)
XNOR2X1	36.05	41.32 (14.6%)	41.76 (15.8%)	41.92 (16.3%)	12.66	14.12 (11.5%)	14.13 (11.6%)	14.18 (12.0%)
XOR2X1	35.59	41.95 (17.9%)	42.41 (19.2%)	42.69 (19.9%)	12.53	14.16 (13.0%)	14.24 (13.6%)	14.28 (13.9%)
Average	21.42	21.88 (2.15%)	22.11 (3.22%)	22.23 (3.78%)	23.71	23.49 (-0.93%)	23.60 (-0.46%)	23.69 (-0.08%)

TABLE III
COMPARISON OF AREA, WIRELENGTH, AND NUMBER OF DRC VIOLATIONS IN 2D FFT AND MONOLITHIC 3D FFT WITH 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), AND 10 (*Mono3D_v3*) ROUTING TRACKS IN EACH CELL. *DRC vios* REFERS TO THE NUMBER OF DESIGN RULE CHECK (DRC) VIOLATIONS.

Circuit	Freq (GHz)	Design Style	Area (mm^2)	Wirelength (m)	DRC Vios
FFT128	0.5	2D	2.54	12.2	-
		3D_v1	1.55 **(-39.1)**	9.4 **(-23)**	-
		3D_v2	1.59 **(-37.2)**	9.2 **(-24)**	-
		3D_v3	1.77 **(-30.3)**	10.2 **(-16)**	-
	1.5	2D	2.94	15.2	0
		3D_v1	1.76 **(-40.2)**	11.6 **(-24)**	568
		3D_v2	1.84 **(-37.5)**	11.4 **(-25)**	7
		3D_v3	2.04 **(-30.6)**	13.0 **(-15)**	0

2) Delay and Power Consumption: HSPICE simulations are performed on the extracted 3D netlists to compare monolithic 3D technology with the conventional 2D technology at the cell level. At 1.1 V power supply, 50 ps transition time, and 27°C temperature, average delay and power consumption are analyzed, as listed in Table II for 2D and each version of the 3D technology. According to this table, *Mono3D_v1* cells have, on average, 2.15% (3.22% in *Mono3D_v2* and 3.78% in *Mono3D_v3*) higher propagation delay and 0.93% (0.46% in *Mono3D_v2* and 0.08% in *Mono3D_v3*) lower power consumption as compared to the 2D standard cells. This slight increase in delay is due to denser cell layout, producing additional coupling capacitances and MIV impedances. Note that in a DFF cell, both delay (clock-to-Q delay) and power are improved as compared to 2D cells since the DFF cell has relatively longer average interconnect length where the monolithic 3D technology is helpful. Also note that the cell-level change in delay and power highly depends upon the individual cell layout, interconnects, and MIVs.

IV. EXPERIMENTAL RESULTS

The proposed open cell libraries are used to develop a parallel 128-point FFT core with approximately 330K cells, operating at 500 MHz and 1.5 GHz. The layout views of the 2D and 3D versions of the FFT core are depicted in Fig. 3, illustrating the effect of number of tracks on footprint.

Specifically, as compared to the 2D version, the footprint and overall wirelength are reduced, respectively, by 40% and 24% in *Mono3D_v1*, 38% and 25% in *Mono3D_v2*, and 31% and 15% in *Mono3D_v3*, when the operating frequency is 1.5 GHz. These results are listed in Table III. At 500 MHz, no DRC violations are reported. At 1.5 GHz, however, *Mono3D_v1* exhibits approximately 600 violations due to routing congestion. These violations are reduced to 9 for *Mono3D_v2*, and are completely eliminated for *Mono3D_v3*.

At 1.5 GHz, the FFT core implemented with *Mono3D_v2* library consumes approximately 20% less interconnect power as compared to 2D technology. The internal power is also reduced by approximately 13%, partly due to the type of cells used in the design and partly due to reduction in short-circuit power. Overall, the monolithic 3D technology achieves approximately 15% reduction in power, as listed in Table IV.

The timing characteristics of the 2D and monolithic 3D circuits are compared in Table V where the worst negative slack (WNS), total negative slack (TNS), and number of timing violations are listed. The target clock frequencies are 500 MHz and 1.5 GHz. An important observation from Table V is that the timing characteristics are degraded when 8 routing tracks (*Mono3D_v1*) are considered. This degradation is due to 1) higher average cell delay for monolithic 3D technology and 2) routing congestion. However, if the number of routing tracks in each cell is increased to 9 (*Mono3D_v2*), the timing characteristics of all of the 3D benchmarks outperform 2D designs at both 500 MHz and 1.5 GHz. At 500 MHz, the positive slack increases. At 1.5 GHz, the WNS, TNS, and number of

TABLE IV

COMPARISON OF POWER CONSUMPTION IN 2D FFT AND MONOLITHIC 3D FFT WITH 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), AND 10 (*Mono3D_v3*) ROUTING TRACKS IN EACH CELL. *INT, SWI*, AND *LK* REFER, RESPECTIVELY, TO INTERNAL, SWITCHING (NET), AND LEAKAGE POWER.

Operating Frequency		500 MHz				1.5 GHz			
Circuit	Design Style	Power component (mW)				Power component (mW)			
		INT	SWI (Change)	LK	Total (Change)	INT	SWI (Change)	LK	Total (Change)
FFT128	2D	2,365	924.8 (-)	119.5	3,510 (-)	7,891	2,859 (-)	144.9	10,895 (-)
	3D_v1	2,340	750.3 **(-18.87%)**	119.4	3,210 **(-8.55%)**	6,936	2,351 **(-17.77%)**	146.5	9,437 **(-13.38%)**
	3D_v2	2,309	726.0 **(-21.50%)**	118.8	3,154 **(-10.14%)**	6,863	2,309 **(-19.24%)**	145.9	9,318 **(-14.47%)**
	3D_v3	2,302	749.7 **(-18.93%)**	119.9	3,172 **(-9.63%)**	6,884	2,333 **(-18.40%)**	145.3	9,363 **(-14.06%)**

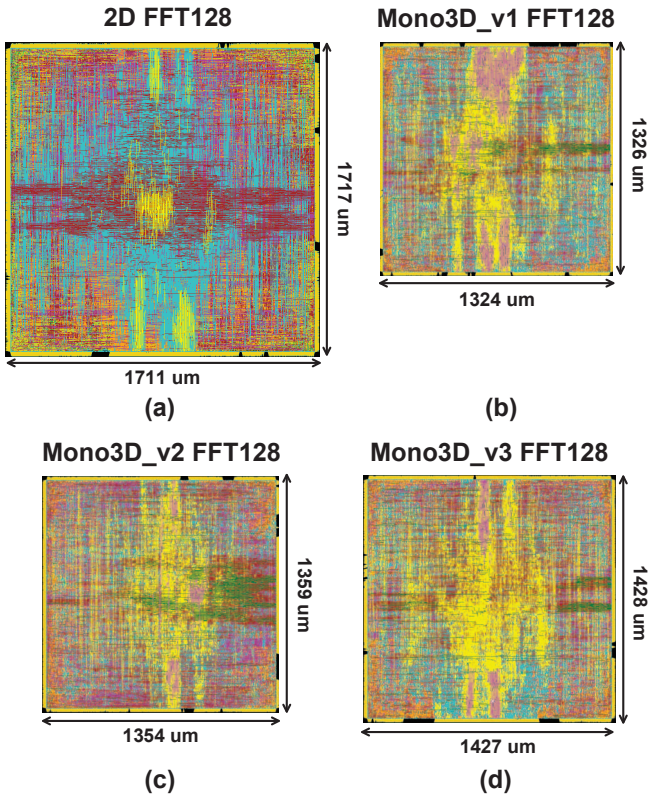

Fig. 3. **The layout views of a highly parallelized 128-point FFT core in (a) conventional 2D technology with 14 routing tracks in each cell, (b) transistor-level monolithic 3D technology with 8 routing tracks, (c) monolithic 3D technology with 9 routing tracks, and (d) monolithic 3D technology with 10 routing tracks.**

TABLE V

COMPARISON OF TIMING CHARACTERISTICS IN 2D FFT AND MONOLITHIC 3D FFT WITH 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), AND 10 (*Mono3D_v3*) ROUTING TRACKS IN EACH CELL. *WS, WNS* AND *TNS* REFER, RESPECTIVELY, TO WORST SLACK, WORST NEGATIVE SLACK AND TOTAL NEGATIVE SLACK.

Operating Frequency		500 MHz	1.5 GHz		
Circuit	Design Style	WS (ns)	WNS (ns)	TNS (ns)	# Violations
FFT128	2D	0.021	-0.104	-516.100	8,097
	3D_v1	0.016	-0.116	-725.474	9,118
	3D_v2	0.023	-0.091	-302.582	6,221
	3D_v3	0.022	-0.097	-319.003	6,574

violations are reduced. Thus, similar to power characteristics, *Mono3D_v2* achieves the best timing characteristics.

V. CONCLUSION

An open source transistor-level monolithic 3D cell library is developed and integrated into a digital design flow. The proposed library is used to investigate several important characteristics of monolithic 3D ICs such as 1) footprint, timing and power consumption at both relaxed and tight timing constraints, 2) routing congestion, and 3) the effect of number of routing tracks in each cell. The results of a large-scale FFT core operating at 1.5 GHz demonstrate that the monolithic 3D technology can reduce the footprint and overall power consumption by, respectively, 38% and 14%. The effect of routing congestion on timing characteristics is stronger in monolithic 3D technology, where the cell-level number of routing tracks plays a critical role. An optimum number of routing tracks exists that achieves the largest improvements in both power and timing characteristics.

REFERENCES

[1] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*. McGraw-Hill Professional, Aug. 2012.

[2] V. F. Pavlidis and E. G. Friedman, *Three-Dimensional Integrated Circuit Design*. Morgan Kaufmann, July 2010.

[3] P. Batude *et al.*, "GeOI and SOI 3D Monolithic Cell Integrations for High Density Applications," in *Proc. of the IEEE Int. Symp. on VLSI Technology*, June 2009, pp. 166–167.

[4] 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. of the ACM ISLPED*, Aug. 2014, pp. 171–176.

[5] C. Yan, S. Kontak, H. Wang, and E. Salman, "Open Source Cell Library Mono3D to Develop Large-Scale Monolithic 3D Integrated Circuit," in *Proc. of IEEE ISCAS*, May 2017, pp. 2581–2584.

[6] "Mono3D, Open Source Cell Library for Transistor-Level Monolithic 3D Integration." [Online]. Available: http://nanocas.ece.stonybrook.edu/mono3d/

[7] C. Yan, J. Dofe, S. Kontak, Q. Yu, and E. Salman, "Hardware-Efficient Logic Camouflaging for Monolithic 3D ICs," *IEEE Transactions on Circuits and Systems II: Express Briefs*, to appear.

[8] J. Dofe, C. Yan, S. Kontak, E. Salman, and Q. Yu, "Transistor-Level Camouflaged Logic Locking Method for Monolithic 3D IC Security," in *Proc. of the IEEE Asian HOST*, December 2016.

[9] "FreePDK45." [Online]. Available: http://www.eda.ncsu.edu/wiki/FreePDK45:Contents

[10] Y.-J. Lee, D. Limbrick, and S. K. Lim, "Power Benefit Study for Ultra-High Density Transistor-Level Monolithic 3D ICs," in *Proc. of the ACM DAC*, May 2013.

[11] J. Shi *et al.*, "On the Design of Ultra-High Density 14nm Finfet Based Transistor-Level Monolithic 3D ICs," in *Proc. of the IEEE ISVLSI*, July 2016, pp. 449–454.

[12] F. B. Claire *et al.*, "FDSOI Bottom MOSFETs Stability versus Top Transistor Thermal Budget Featuring 3D Monolithic Integration," *Solid-State Electronics*, vol. 113, pp. 2–8, Nov. 2015.

[13] "Nangate Cell Library." [Online]. Available: http://www.nangate.com