Design Quality Trade-Off Studies for 3-D ICs Built With Sub-Micron TSVs and Future Devices

Dae Hyun Kim, *Student Member, IEEE*, and Sung Kyu Lim, *Senior Member, IEEE*

Abstract—Through-silicon vias (TSVs) have two negative effects in the design of three-dimensional integrated circuits (3-D ICs). First, TSV insertion leads to silicon area overhead. In addition, nonnegligible TSV capacitance causes delay overhead in 3-D signal paths. Therefore, obtaining all benefits such as wirelength reduction and performance improvement from 3-D ICs is highly dependent on TSV size and capacitance. Meanwhile, TSVs are downscaled to minimize their negative effects, and sub-micron TSVs are expected to be fabricated in the near future. At the same time, the devices are also downscaled beyond 32 nm and 22 nm, so future 3-D ICs will very likely be built with sub-micron TSVs and advanced device technologies. In this paper, we investigate the impact of sub-micron TSVs on the quality of today and future 3-D ICs. For future process technologies, we develop 22 nm and 16 nm libraries. Using these future process libraries and an existing 45 nm library, we generate 3-D IC layouts with different TSV sizes and capacitances and study the impact of sub-micron TSVs thoroughly.

Index Terms—Device, interconnect, three-dimensional integrated circuit (3-D IC), through-silicon via (TSV).

I. INTRODUCTION

THREE-DIMENSIONAL integrated circuits (3-D ICs) are expected to offer various benefits such as higher bandwidth, smaller form factor, shorter wirelength, lower power, and better performance than traditional 2-D ICs. These benefits are obtained by die stacking and use of through-silicon vias (TSVs) for inter-die connections. However, TSVs have two negative effects, occupation of silicon area and nonnegligible capacitance, in the design of 3-D ICs. The fact that TSVs occupy silicon area has great effects not only on silicon area, but also on wirelength, critical path delay, and power. The reason is as follows. If larger TSVs are inserted in a 3-D IC layout, footprint area of the design becomes larger, so the average wirelength increases [1]. This wirelength overhead leads to longer critical path delay and higher dynamic power consumption because of increased wire capacitance. In addition, nonnegligible TSV capacitance also has a negative effect on critical path delay and dynamic power consumption. One thing to notice is that small TSVs do not necessarily have smaller capacitance than large TSVs. The reason is because TSV capacitance is dependent not only on the TSV diameter and height, but also on other factors such as the liner thickness and doping concentration of the substrate [2].

Similarly as devices are scaled, TSVs are also downscaled [3]–[5]. Therefore, negative effects of TSVs will be reduced if smaller TSVs are used.[1] However, since process technology is also advancing, future 3-D ICs will very likely be fabricated with smaller TSVs and state-of-the-art process technology. In this case, negative effects of TSVs might remain the same or even increase.

In this paper, we investigate the impact of sub-micron TSVs on the area, wirelength, critical path delay, and power of today and future 3-D ICs based on GDSII-level layouts. For future process technologies, we develop 22 nm and 16 nm process and standard cell libraries. Using these future process technologies as well as an existing 45 nm library, we generate 3-D IC layouts with different TSV sizes and capacitances and study the impact of TSVs thoroughly. The contributions of this paper are as follows.

- To investigate the impact of sub-micron TSVs on future 3-D ICs, we develop a 22 nm and a 16 nm process and standard cell libraries. These libraries enable us to obtain very trustable simulation results.
- We generate layouts with various device and TSV technology combinations and obtain area, wirelength, critical path delay, and power. Therefore, we not only cross-compare 3-D ICs built with different process technologies, but also compare 3-D ICs built with the same process technology and different TSV sizes and capacitances.
- We also cross-compare 2-D designs built with more advanced process technology and 3-D designs built with older process technology. Our simulation results show that 3-D ICs built with an nth generation process technology could be beaten by 2-D ICs built with an $(n + 2)$th generation process technology.[2]

The rest of this paper is organized as follows. In Section II, we review negative effects of TSVs and show the motivation of this paper. Section III demonstrates the development flow of our 22 nm and 16 nm process and standard cell libraries, and compares a 45 nm and our 22 nm and 16 nm libraries. In Section IV, we explain our full-chip 3-D IC design and analysis methodology. In Section V, we present our simulation results, then we conclude in Section VI.

[1]If we assume that only the TSV size and height are downscaled while other parameters such as the liner thickness and doping concentration are fixed, TSV capacitance decreases as TSVs are downscaled.

[2]This observation is strongly dependent on TSV capacitance used at each process node.

II. PRELIMINARIES

A. Negative Effects of TSVs

The use of TSVs in 3-D ICs have two negative effects on the quality of 3-D ICs: area and delay overhead. According to recent research on TSV area overhead [6], silicon area occupied by TSVs is quite significant, which in turn reduces the wirelength benefit of 3-D ICs. In addition, TSV capacitance could make a delay degradation problem on 3-D signal paths [7]. Although buffer insertion can reduce delay overhead caused by TSV capacitance, buffer insertion itself also causes another problem: additional silicon area for buffer insertion and power overhead.

The degree of negative effects of TSVs on 3-D ICs is dependent on various technology and design parameters. For example, if we use 5 μm TSVs[3] with state-of-the-art process technology such as 32 nm technology in 3-D IC designs, these TSVs may cause a huge area overhead. However, if we use the same TSVs with relatively old technology such as 0.18 μm technology, these TSVs may not cause any area overhead because the latter (0.18 μm) has a smaller ratio between the TSV area and the gate area than the former (32 nm). On the other hand, small TSVs (e.g., 1 μm TSVs) could have huge capacitance depending on the liner thickness and doping concentration of the substrate. In this case, small TSVs may not cause area overhead, but they will cause serious delay overhead.

B. Motivation

Downscaling of devices reached 22 nm node [8], [9] in 2012, and 16 nm and 11 nm technologies are currently under development. As devices are downscaled, TSVs are also downscaled as TSV manufacturing technology advances. Recently, it was demonstrated that 0.7-μm-diameter TSVs could also be fabricated reliably [5]. In addition, according to the ITRS prediction, TSV diameter will continue to decrease while TSV aspect ratio will increase. Therefore, we expect that sub-micron TSVs will be developed and be ready for use within the next few years.

However, all the existing work on the impact of TSVs on the quality of 3-D IC designs focuses on micron-size TSVs and current (e.g., 45 nm) or even old (e.g., 130 nm) process technologies. For example, a 45 nm technology and 1.67 μm TSVs are used in [10] and a 45 nm technology and TSVs whose width is approximately 4 μm are used in [11]. However, none of them discuss what will occur if sub-micron TSVs are used with a 45 nm technology or if the same-size TSVs are used with different process technologies (e.g., a 90, 32, or 22 nm). However, it is crucial to accurately predict the impact of new TSV technology on the design quality of 3-D ICs in order to refine the technology or justify the investment and cost. Therefore, the goal of this paper is to investigate the impact of sub-micron TSVs on the area, wirelength, critical path delay, and power of today and future 3-D IC designs.

[3]A "X μm TSV" in this paper denotes a TSV whose width(= for square − shaped TSVs) or diameter(= for cylindrical − type TSVs) is X μm.
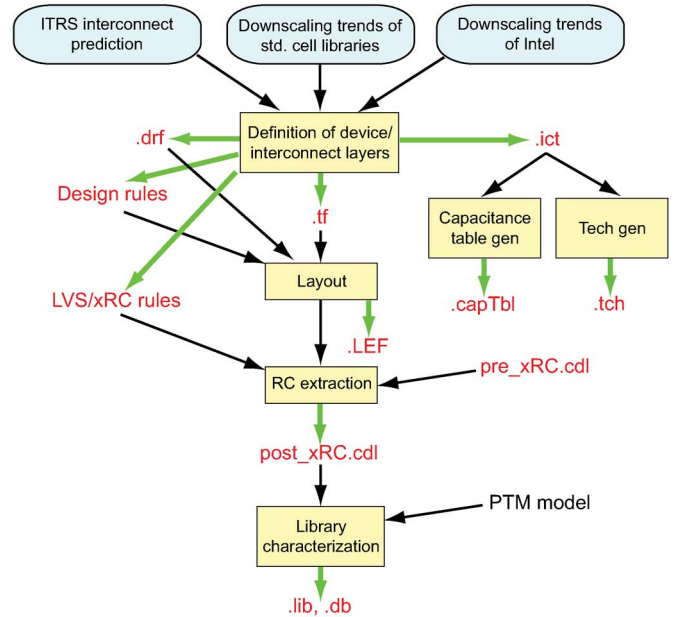

Fig. 1. Development flow of our 22 nm and 16 nm process and standard cell libraries.

III. LIBRARY DEVELOPMENT FLOW

In this section, we demonstrate the development flow of our 22 nm and 16 nm process and standard cell libraries. For 22 nm and 16 nm transistor models, we use the the predictive technology model (22 nm and 16 nm PTM HP model V2.1) [12].

A. Overall Development Flow

For the development of 22 nm and 16 nm process and standard cell libraries, we follow a typical library development flow illustrated in Fig. 1. We first define device and interconnect layers from which we create a tech file (.tf), a display resource file (.drf), an interconnect technology file (.ict), a design rule file, a layout-versus-schematic (LVS) rule file, and an RC parasitic extraction rule file. With the tech file and the display resource file, we draw standard cell layouts. After the layout generation, we perform abstraction to create a library exchange format file (.LEF), and run RC extraction and create SPICE netlists (post_xRC.cdl). With these SPICE netlists and the PTM transistor models, we perform library characterization to create timing and power libraries (.lib and.db). We also generate a capacitance table and a.tch file for sign-off RC extraction and timing analysis.

B. Interconnect Layers

We define interconnect layers based on ITRS interconnect prediction [17], downscaling trends of other standard cell libraries, and the downscaling trends of Intel process technology [13]–[15]. According to ITRS prediction on interconnect layers, for example, the pitch of the metal 1 wire at 22 nm is about 72 nm and that at 16 nm is about 48 nm, and the pitch of a semi-global wire at 22 nm is about 160 nm and that at 16 nm is about 130 nm. From these values as well as extrapolation of interconnect layers of Intel process technology and other standard cell libraries, we predict interconnect layers at 22 nm and

TABLE I
INTERCONNECT LAYERS OF 65 nm [13], 45 nm [14], 32 nm [15], 22 nm, AND 16 nm PROCESS TECHNOLOGY. THE 22 nm AND THE 16-nm LAYERS ARE FROM OUR PREDICTION

Layer	Pitch (nm)				
	$65nm$	$45nm$	$32nm$	$22nm$	$16nm$
Contacted Gate	220	160	112.5	86	62
Metal 1	210	160	112.5	76	46
Metal 2	210	160	112.5	76	46
Metal 3	220	160	112.5	76	46
Metal 4	280	240	168.8	130	72
Metal 5	330	280	225.0	206	98
Metal 6	480	360	337.6	206	146
Metal 7	720	560	450.1	390	240
Metal 8	1080	810	566.5	390	240

TABLE II
WIDTH (w) AND THICKNESS (t) OF METAL LAYERS USED IN OUR 22 nm AND 16 nm PROCESS LIBRARIES. THE ASPECT RATIO FOR THE 22 nm LIBRARY IS 1.8 AND THAT FOR THE 16 nm LIBRARY IS 1.9

Layer	$22nm$		$16nm$	
	w (nm)	t (nm)	w (nm)	t (nm)
Metal 1, 2, 3	36	64.8	22	41.8
Metal 4	60	108	32	60.8
Metal 5	96	172.8	44	83.6
Metal 6	96	172.8	66	125.4
Metal 7, 8	180	324	110	209

TABLE III
STANDARD CELLS IN OUR 22 nm AND 16 nm STANDARD CELL LIBRARIES

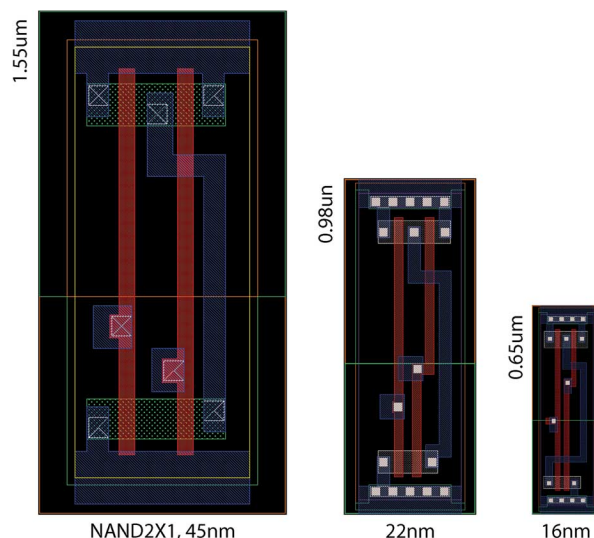| Type	Available sizes
AND2/3/4, AOI21/211/221	$1\times$, $2\times$, $4\times$
BUF, INV	$1\times$, $2\times$, $4\times$, $8\times$, $16\times$, $32\times$
LOGIC 0, LOGIC 1	$1\times$
MUX2	$1\times$, $2\times$
NAND2/3/4/, NOR2/3/4	$1\times$, $2\times$, $4\times$
OAI21/22/211/221/222	$1\times$, $2\times$, $4\times$
OAI33	$1\times$
OR2/3/4	$1\times$, $2\times$, $4\times$
XNOR2, XOR2	$1\times$, $2\times$
DFF	$1\times$, $2\times$
FA, HA	$1\times$

Fig. 2. The smallest ($1\times$) two-input NAND gates of the 45 nm [16], and our 22 nm and 16 nm libraries (drawn to scale).
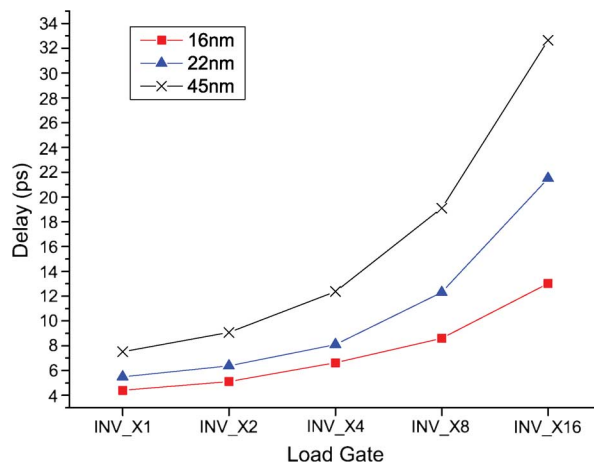

Fig. 3. Delay of a minimum-size inverter driving an $N\times$ inverter ($N = 1, 2, 4, 8, 16$), where both inverters are in the same process. RC parasitics are included.

16 nm as listed in Table I. Table II lists widths and thicknesses of all metal layers of our 22 nm and 16 nm process libraries. The aspect ratio of the 22 nm library is set to 1.8 and that of the 16 nm library is set to 1.9. Since we assume that low-k inter-layer insulator material is used, we use 1.9 for the dielectric constant of the inter-layer dielectric material and 3.8 for the dielectric constant of the barrier material for both the 22 nm and the 16 nm libraries.

C. Standard Cell Library

We first create a tech file defining device and interconnect layers and a set of design rules such as minimum poly-to-contact spacing and minimum metal-to-metal spacing. Then, we draw standard cell layouts with this tech file and the design rules.[4] We created about 90 cells and Table III lists the standard cells except antenna and filler cells. The placement site width and height of our 22 nm standard cell library are 0.1 μm and 0.9 μm, respectively, and those of our 16 nm library are 0.06 μm and 0.6 μm, respectively. Fig. 2 shows the smallest ($1\times$) two-input

[4]We referred to standard cell layouts of the Nangate 45 nm standard cell library [16].

NAND gates of the 45 nm, our 22 nm, and 16 nm standard cell libraries. After creating the standard cell layouts, we perform DRC and LVS for each layout and extract parasitic RC of each standard cell. We also characterize all standard cells to create timing and power libraries.

D. Comparison of 45 nm, 22 nm, and 16 nm Libraries

Before we proceed to the comparison of 2-D and 3-D ICs in Section V, we should verify the validity of our libraries. Therefore, we compare transistor characteristics and the Nangate 45 nm, our 22 nm, and our 16 nm standard cell libraries using commercial tools only.

1) Gate Delay and Input Capacitance: Gate delay and drive strength are determined by transistor characteristics and the gate size. Therefore, our first simulation is to compare the transistor characteristics. The simulation setting is as follows. A minimum-size inverter in each process library drives another minimum-size inverter, which drives an $N\times$ inverter of the same library. We obtain the delay of the second minimum-size inverter (driving the $N\times$ inverter) by SPICE simulation. Fig. 3 shows

TABLE IV
FO4 DELAY, STANDARD CELL HEIGHTS, WIRE SHEET
RESISTANCE, AND UNIT WIRE CAPACITANCE (fF/μm)

	$45nm$	$22nm$	$16nm$
FO4 delay	$15.15ps$	$13.63ps$	$12.28ps$
Std. cell. height	$1.4\mu m$	$0.9\mu m$	$0.6\mu m$
Wire sheet resistance (Metal 1)	0.38	0.26	0.40
(Metal 4)	0.21	0.16	0.28
(Metal 7)	0.08	0.05	0.08
Unit wire capacitance (Metal 1)	0.20	0.15	0.16
(Metal 4)	0.20	0.15	0.13
(Metal 7)	0.20	0.14	0.14

TABLE V
INPUT CAPACITANCE OF SELECTED STANDARD CELLS IN
THE 45 nm, THE 22 nm, AND THE 16 nm LIBRARIES

Cell	Cap (fF)		
	$45nm$	$22nm$	$16nm$
AND2 1×	0.54 (1.00)	0.25 (0.46)	0.22 (0.41)
AOI211 1×	0.64 (1.00)	0.30 (0.47)	0.25 (0.39)
AOI21 1×	0.55 (1.00)	0.23 (0.42)	0.20 (0.36)
BUF 4×	0.47 (1.00)	0.28 (0.60)	0.29 (0.62)
DFF 1×	0.90 (1.00)	0.41 (0.46)	0.26 (0.29)
FA 1×	2.46 (1.00)	1.31 (0.53)	1.36 (0.55)
INV 4×	1.45 (1.00)	0.69 (0.48)	0.56 (0.39)
MUX2 1×	0.95 (1.00)	0.42 (0.44)	0.34 (0.36)
NAND2 1×	0.50 (1.00)	0.24 (0.48)	0.22 (0.44)
OAI21 1×	0.53 (1.00)	0.25 (0.47)	0.20 (0.38)
OR2 1×	0.60 (1.00)	0.26 (0.43)	0.20 (0.33)
XOR2 1×	1.08 (1.00)	0.55 (0.51)	0.45 (0.42)
Average	(1.00)	(0.48)	(0.40)

TABLE VI
BENCHMARK CIRCUITS

Circuit	# Gates	# Nets	Total cell area		
			$45nm$	$22nm$	$16nm$
BM1	352K	372K	0.632	0.218	0.098
BM2	518K	680K	1.288	0.437	0.198

TABLE VII
COMPARISON OF 2-D LAYOUTS

	BM1			BM2		
	$45nm$	$22nm$	$16nm$	$45nm$	$22nm$	$16nm$
Area (mm^2)	1.00	0.36	0.17	2.56	0.81	0.42
Wirelength (m)	10.65	4.22	2.75	15.17	8.90	6.19
Delay (ns)	3.19	2.61	2.38	6.51	4.10	3.93
Power (W)	0.352	0.0684	0.068	0.521	0.154	0.133

the delay. We observe that the 16 nm inverter has the shortest delay and the 45 nm inverter has the longest delay. Quantitatively, we observe approximately 30% improvement when the process moves from 45 nm to 22 nm and about 20% improvement when the process moves from 22 nm to 16 nm. Notice that this SPICE simulation does not consider interconnect parasitic resistance and capacitance. Table IV also shows the FO4 delay at each process technology.

Since gate input capacitance is also an important factor determining delay and power, we show input capacitances of 45 nm, 22 nm, and 16 nm standard cells in Table V. As shown in the table, the average input capacitance of the 22 nm standard cells is approximately 48% of the average input capacitance of the 45 nm standard cells. On the other hand, the average input capacitance of the 16 nm standard cells is approximately 83% of the average input capacitance of the 22 nm standard cells. Since two generation gap exists between 45 nm and 22 nm, the input capacitance difference between 45 nm and 22 nm is greater than that between 22 nm and 16 nm.

2) Interconnect Layers: Characteristics of interconnect layers also have a big effect on the performance of a library, so we show wire sheet resistance and unit wire capacitance of short, semi-global, and global metal layers in Table IV. The resistivity of the 45 nm technology is about 5.0×10^{-8}, so the sheet resistance of the library is relatively high compared to the 22 nm library. On the other hand, the resistivity of the 22 nm and 16 nm technology is 1.7×10^{-8}, which is the resistivity of copper. This is why the sheet resistances of the 22 nm metal layers are lower than those of the 45 nm metal layers although the thickness of the 45 nm metal layers is larger than that of the 22 nm metal layers. On the other hand, as the technology moves

from 22 nm to 16 nm, the sheet resistance goes up because both of them use the same resistivity, but the metal layer thickness of the 16 nm library is smaller than that of the 22 nm library.

The unit wire capacitance of the 45 nm library is also slightly higher than that of the 22 nm library. This is because the dielectric constant used for the 45 nm library is 2.5 while the 22 nm library uses 1.9 for its dielectric constant. If the same dielectric material ($\epsilon_r = 1.9$) is used for the 45 nm library, the unit wire capacitance becomes 0.15, which is close to the unit wire capacitance of the 22 nm library.

3) Full-Chip 2-D Design: In this simulation, we design 2-D circuits using the three standard cell libraries and compare the area, wirelength, critical path delay, and power. The simulation flow is as follows. We prepare two benchmark circuits shown in Table VI, synthesize, design, and optimize them using each standard cell library and commercial tools. We use the same area utilization (60%) for all libraries for fair comparison and find the fastest operation frequency for each library.

Table VII shows the comparison results for the 2-D designs. The chip area of the 45 nm designs is about three times larger than that of the 22 nm designs on average, and the chip area of the 22 nm designs is approximately two times larger than that of the 16 nm designs on average. In addition, the total wirelength of the 16 nm designs is approximately 1.48× shorter than that of the 22 nm designs, and 3.08× shorter than that of the 45 nm designs. Regarding the critical path delay, the 16 nm designs are 1.49× faster than the 45 nm designs on average and 1.07× faster than the 22 nm designs on average. Power consumption of the 16 nm designs is approximately 4.5× smaller than that of the 45 nm designs and 1.1× smaller than that of the 22 nm designs. Overall, the delay and power enhancement coming from 22 nm to 16 nm transition is not as significant as the enhancement coming from 45 nm to 22 nm transition because 45 nm and 22 nm technologies are two generations apart while 22 nm and 16 nm technologies are only one generation apart, and the quality (sheet resistance and unit wire capacitance) of the interconnect layers of the 45 nm library is worse than that of the 22 nm library.

IV. FULL-CHIP 3-D IC DESIGN AND ANALYSIS METHODOLOGY

To generate 3-D IC layouts, we use the 3-D RTL-to-GDSII tool obtained from [11]. This tool works as follows. For a given 2-D gate-level (flattened) netlist, this tool partitions gates in the

x-, y-, and z- directions iteratively to globally place gates in grids in 3-D. After the global placement, it constructs a 3-D Steiner tree for each 3-D net and inserts TSVs into each placement grid based on the locations of vertical edges of the 3-D Steiner tree. Then, it runs detailed placement in each placement grid using Cadence Encounter [18]. Routing for each die is also performed by Encounter. The output of the tool consists of a verilog netlist, a design exchange format (DEF) file containing TSV locations, and a standard parasitic exchange format (SPEF) file for each die, and a top-level verilog netlist containing die-to-die connections and a top-level SPEF file. One thing to notice is that the minimum number of TSVs to be inserted in the 3-D design is dependent on the cut sequence, which is the order of the x-, y-, and z- direction partitioning we apply for global placement. For example, if we apply the z-direction partitioning early, we are likely to obtain fewer inter-die connections. On the other hand, if we apply the z-direction partitioning later, we are likely to obtain more inter-die connections [11]. This variation of the number of TSVs enables us to produce different global placement solutions with different TSV counts.

After generating 3-D IC layouts, we perform 3-D timing optimization. We first perform initial timing optimization for each die. Then, we feed all the layouts, timing analysis results, and the target clock frequency into the 3-D timing optimization tool obtained from [19]. This 3-D timing optimization tool iterates the following steps: (a) it performs RC extraction and obtains an SPEF file for each die; (b) it performs 3-D timing analysis using the SPEF files and the top-level SPEF file using Synopsys PrimeTime [20]; (c) based on the timing analysis result and the target clock frequency, the tool determines the target delay of each 3-D path and creates a timing constraint file for each die; (d) since each die has its own netlist and timing constraint file, we perform timing optimization for each die separately using Cadence Encounter. We iterate this timing optimization process several times until the overall timing improvement saturates.

3-D power analysis needs (a) a netlist for each die and a top-level netlist, (b) an SPEF file for each die and a top-level SPEF file, and (c) switching activities of cells and nets. To obtain switching activities of cells and nets, we load verilog netlists generated by the 3-D RTL-to-GDSII tool obtained from [11] into Encounter and run power analysis. This power analysis internally generates and stores switching activities of the cells and nets, and we dump this information into an output file after the power analysis. Then, we load all netlists, SPEF files, and the switching activity files into PrimeTime and run power analysis. This power analysis method produces true full-chip 3-D power analysis results.

V. SIMULATION RESULTS

A. Simulation Settings

We use two benchmark circuits, BM1 and BM2, as shown in Table VI. For the 45 nm process node, we use the Nangate 45 nm standard cells library [16]. We also use four sets of TSV-related dimensions shown in Table VIII. In our simulations, we use 5 μm and 0.5 μm TSVs with the 45 nm technology, 1 μm and 0.1 μm TSVs with the 22 nm technology, and 0.5 μm and 0.1 μm TSVs with the 16 nm technology. Since the standard cell

TABLE VIII
TSV-RELATED DIMENSIONS, DESIGN RULES, AND TSV CAPACITANCE

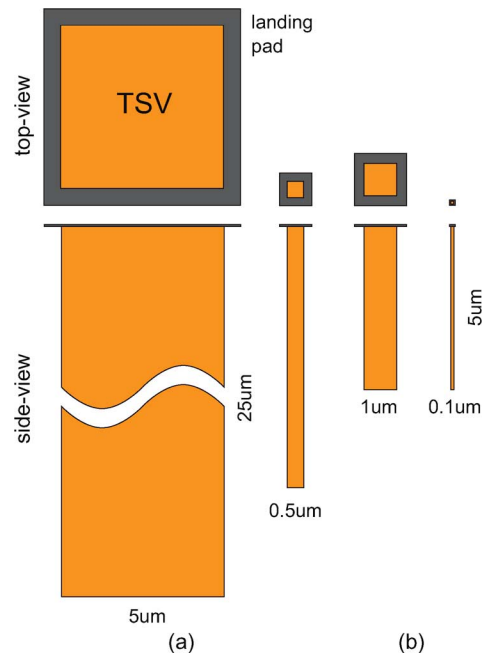| Dimensions	TSV-5	TSV-1	TSV-0.5	TSV-0.1
Width (μm)	5	1	0.5	0.1
Height (μm)	25	5	8	5
Aspect ratio	5	5	16	50
Liner thickness (nm)	100	30	20	10
Barrier thickness (nm)	50	15	10	5
Landing pad width (μm)	6	1.6	1	0.18
TSV-to-TSV spacing (μm)	2	0.8	0.6	0.1
TSV-to-device spacing (μm)	1	0.4	0.3	0.1
TSV capacitance (fF)	20	2.67	3.2	0.8
Used with	$45nm$	$22nm$	$45nm$ $16nm$	$22nm$ $16nm$

Fig. 4. Size comparison of the 4 TSVs used in our study: (a) 5 μm and 0.5 μm width used for 45 nm technology, (b) 1 μm and 0.1 μm width used for 22 nm technology.

height of the 45 nm library is 1.4 μm, a 5 μm TSV including its keep-out zone occupies five standard cell rows while a 0.5 μm TSV including its keep-out zone occupies one standard cell row. Similarly, a 1 μm TSV and a 0.1 μm TSV occupy three standard cell rows and 0.26 standard row, respectively, when they are used with the 22 nm standard cell library. If 0.5 μm and 0.1 μm TSVs are used for the 16 nm standard cell library, a 0.5 μm TSV occupies 1.33 standard cell rows and a 0.1 μm TSV occupies 0.5 standard cell row. Fig. 4 shows the four different TSVs in a top-down view and a side view and Fig. 5 shows GDSII images of TSVs and standard cells at 45, 22, and 16 nm technology. We also assume that face-to-back bonding and via-first TSVs are used for 3-D integration. Die thickness that we use for each TSV dimension set is same as the TSV height, which ranges from 5 to 25 μm. Although 5 μm thickness is extremely thin, it is practical [21], [22].

B. Impact on Silicon Area

Fig. 6 shows footprint area of 2-D designs and two-die 3-D BM1 designs at each technology node. If the TSV size is zero,
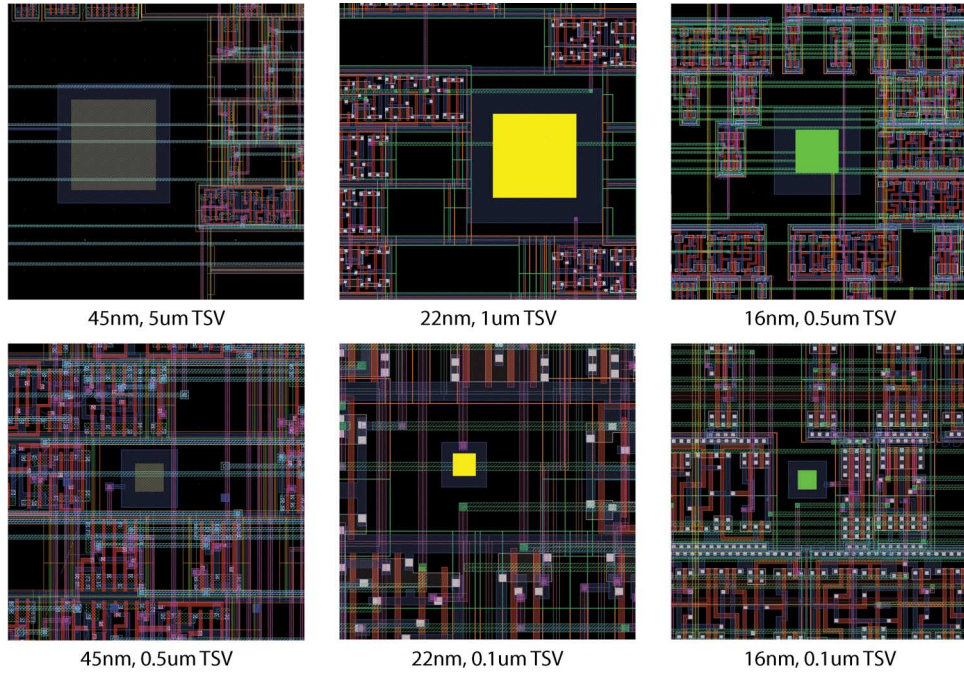
Fig. 5. Zoom-in GDSII layouts of the six types of designs studied in this paper. Each TSV is surrounded by its keep-out-zone.
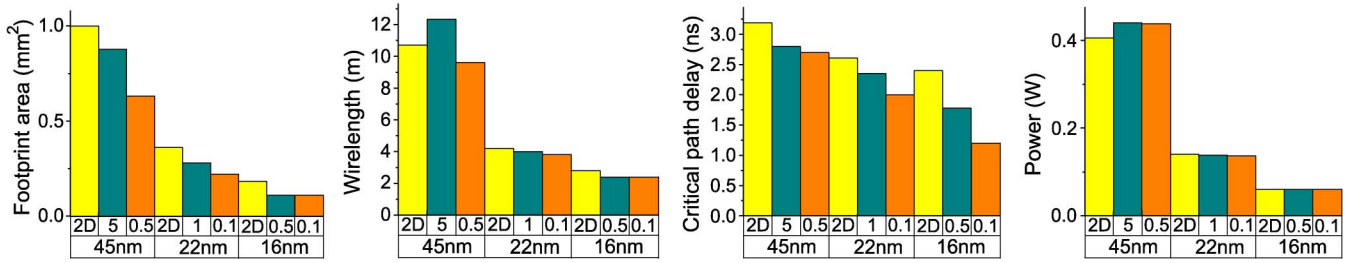

Fig. 6. Comparison of the optimized 2-D designs and two-die 3-D designs (BM1) in 45 nm, 22 nm, and 16 nm technology. The x-axis shows the technology combination (the first row shows TSV diameter in μm).
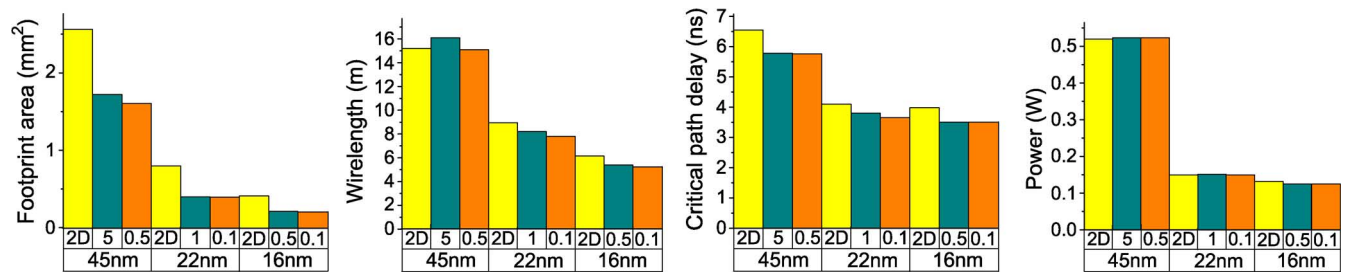

Fig. 7. Comparison of the optimized 2-D designs and two-die 3-D designs (BM2) in 45 nm, 22 nm, and 16 nm technology. The x-axis shows the technology combination (the first row shows TSV diameter in μm).

the footprint area of a two-die 3-D design should be approximately half of its 2-D counterpart. Since the TSV size is not zero, however, the footprint area of a two-die 3-D design is usually greater than half of its 2-D counterpart. For example, the area of the 45 nm 2-D design is 1.0 mm^2, but the area of the 45 nm 3-D design using 5 μm TSVs is about 0.85 mm^2, which is 85% of the 2-D design. Similarly, the area of the 45 nm 3-D design using 0.5 μm TSVs is about 0.63 mm^2, which is 63% of the 2-D design. The same trend is found in the 22 nm and 16 nm designs. However, if the TSV size is 0.1 μm, the footprint area of two-die 3-D designs become almost half of its 2-D counter-

part. We find similar trends in BM2 designs, as shown in Fig. 7.

All these trends depend on the TSV size and the number of TSVs used in the designs. Using smaller TSVs helps achieve smaller footprint area, which can reduce the chip cost. However, smaller TSVs could be more expensive due to manufacturing difficulties, so the use of smaller TSVs does not necessarily lead to lower chip cost. Using fewer TSVs also helps achieve smaller footprint area. However, several studies show that using more TSVs than the minimum number of TSVs helps reduce wirelength and improve performance [10], [11], [23].

Thus, trade-offs exist among the TSV size, the number of TSVs used in the design, and the chip cost.

C. Impact on Wirelength

Fig. 6 shows wirelength of BM1 designs. When 5 μm TSVs are used with the 45 nm technology, 3-D designs have longer wirelength than 2-D designs. However, when 0.5 μm TSVs are used with the 45 nm technology, the wirelength of the 3-D design is about 10% shorter than that of the 2-D design. When 1 μm and 0.1 μm TSVs are used with the 22 nm technology, however, we do not observe large wirelength reduction. On the other hand, when 0.5 μm and 0.1 μm TSVs are used with the 16 nm technology, we observe 15% wirelength reduction.

We find similar trends in BM2 designs as shown in Fig. 7. 45 nm 3-D designs have longer wirelength than 2-D designs. However, when 1 μm and 0.1 μm TSVs are used with the 22 nm technology, we observe 9% and 13% wirelength reduction, respectively. Similarly, when 0.5 μm and 0.1 μm TSVs are used with the 16 nm technology, we observe 12% and 15% wirelength reduction, respectively.

One thing to note is that 3-D designs at the nth generation process node could have longer wirelength than 2-D designs at the $(n+1)$th generation process node. For instance, the 22 nm 3-D layouts designed with 0.1 μm TSVs have longer wirelength than the 16 nm 2-D layouts in Figs. 6 and 7. Therefore, shrinking the TSV size is important to reduce the wirelength, but switching to advanced process nodes is also important for wirelength reduction. This observation also coincides with the prediction result presented in [24].

D. Impact on Performance

Fig. 6 shows the critical path delay of 2-D and 3-D designs for the BM1 benchmark circuit. As seen in the figure, the critical path delay of a 3-D design having longer wirelength than (or similar wirelength to) its 2-D counterpart can be smaller than that of the 2-D design. For example, the wirelength of the 3-D design built with 5 μm TSVs and the 45 nm technology is 15% longer than that of the 2-D design, but the critical path delay of the 3-D design is 12% smaller than that of the 2-D design. Similar trends are also found in the BM2 benchmark circuit as shown in Fig. 7.

One important observation is that the critical path delay of 3-D designs built with the nth generation process node could be smaller than the critical path delay of 2-D designs built with the $(n+1)$th generation process node. For example, the BM1 3-D design built with 0.1 μm TSVs with the 22 nm technology has approximately 20% smaller delay than the 2-D design built with the 16 nm technology. Similarly, the BM2 3-D design built with 0.1 μm TSVs with the 22 nm technology has about 9% smaller delay than the 2-D design built with the 16 nm technology.

For more in-depth analysis, we show the number of TSVs used in the critical paths in Table IX. If the TSV count is zero, the critical path is a 2-D path existing in a single die. If the TSV count is three, the critical path alternates three times (e.g., die0-die1-die0-die1) between two dies since all the layouts are two-die designs. Especially, if the TSV count is zero and the critical path delay is shorter than the critical path delay of its 2-D counterpart design, the shorter critical path delay of the 3-D

TABLE IX
ADDITIONAL TSV-RELATED STATISTICS. "C.P." DENOTES CRITICAL PATH

	BM1					
	45nm		22nm		16nm	
TSV diameter	5μm	0.5μm	1μm	0.1μm	0.5μm	0.1μm
# TSVs in c.p.	1	0	3	4	2	4

	BM2					
	45nm		22nm		16nm	
TSV diameter	5μm	0.5μm	1μm	0.1μm	0.5μm	0.1μm
# TSVs in c.p.	0	0	1	2	1	2

design is primarily due to the shorter wirelength achieved by the smaller footprint area. On the other hand, if the TSV count is nonzero, the critical path delay comes from both the smaller footprint area and the shorter wirelength.

E. Impact on Power

Figs. 6 and 7 show power consumption for BM1 and BM2 benchmark circuits, respectively. As seen in the figures, moving from 2-D ICs to 3-D ICs does not necessarily lead to power reduction even if 3-D designs have shorter wirelength than 2-D designs. The reason is as follows. Reduction in power consumption by building 3-D ICs comes from smaller dynamic power consumption due to shorter wirelength.[5] However, TSV capacitance can essentially be thought of as wire capacitance. Therefore, the total capacitance is the sum of the total TSV capacitance and the total wire capacitance. This means that the total TSV capacitance should be less than the reduced wire capacitance to achieve power reduction.[6] In other words, achievement of power reduction needs smaller TSV capacitance, use of fewer TSVs, and wirelength reduction. However, there again exist trade-offs among the number of TSVs, the amount of wirelength reduction, and power consumption. Inserting fewer TSVs may not reduce the total wirelength as much as expected. Similarly, the use of fewer TSVs may not reduce the dynamic power consumption. Inserting more TSVs, however, may reduce the total wirelength more than 10%–20% [10], but then the total TSV capacitance also increases, so the total capacitance could be larger than the total capacitance of 2-D designs.

Another reason that the total power does not decrease in 3-D designs is related to the wirelength distribution. If wirelength reduction is achieved by shortening short wires in a net, the load capacitance (capacitance of the input pins connected to the net) dominates the capacitance of the net, so the power consumption does not decrease. However, if long wires are shortened, the wire capacitance dominates the capacitance of the net, so we can achieve power reduction. In our simulation, however, we observe that the wirelength reduction comes from shortening short wires.

F. Area, Wirelength, Performance, and Power Versus # Dies

In the above sections, we build 3-D designs in two dies. However, the number of dies also has an impact on the area, wirelength, critical path delay, and power [10]. In this section, there-

[5]There exist many kinds of 3-D integration and some of them (e.g., core-DRAM stacking) provide a huge amount of power saving by removing long chip-to-chip connections.

[6]Note that this is a simplified analysis. In reality, the total power should be computed in a more sophisticated fashion taking switching activities of nets and gates into account.
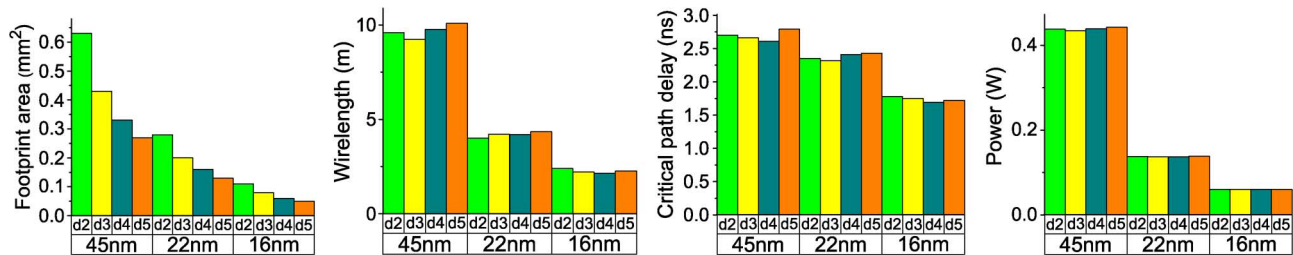
Fig. 8. Comparison of optimized 3-D designs (BM1) implemented in multiple dies. "dn" denotes n-die implementation. We use 0.5 μm, 1 μm, and 0.5 μm TSVs for the 45 nm, 22 nm, and 16 nm technologies, respectively.

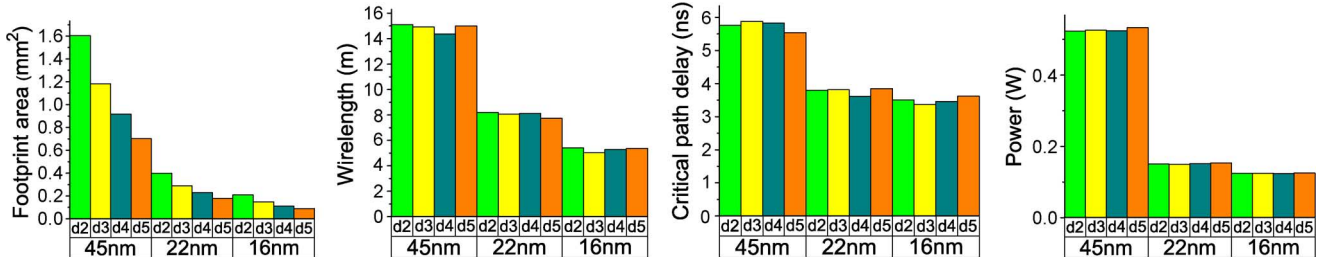Fig. 9. Comparison of optimized 3-D designs (BM2) implemented in multiple dies. "dn" denotes n-die implementation. We use 0.5 μm, 1 μm, and 0.5 μm TSVs for the 45 nm, 22 nm, and 16 nm technologies, respectively.

fore, we vary the number of dies and study the impact of TSVs on the four metrics. Figs. 8 and 9 show footprint area, wirelength, critical path delay, and power for BM1 and BM2 benchmarks when the number of dies varies from two (d2 cases in the figure) to five (d5 cases). To limit the simulation space size, we use 0.5 μm TSVs for the 45 nm technology, 1 μm TSVs for the 22 nm technology, and 0.5 μm TSVs for the 16 nm technology.

As the number of dies increases, the footprint area decreases as expected. Assuming that the TSV size is zero and the same utilization is used for all layouts, the footprint area of an n-die design of a circuit is approximately A_{2D}/n where A_{2D} is the area of the 2-D design of the circuit. However, the TSV size is not negligible and more TSVs are inserted as more dies are stacked, so the footprint area of the circuit designed in n dies is slightly larger than A_{2D}/n.

On the other hand, stacking more dies does not necessarily result in shorter wirelength although stacking more than two dies helps reduce the wirelength in most cases of our simulation. The largest wirelength reduction ratio between more-than-two-die designs and two-die designs is about 11% in our simulation (the 16 nm two-die implementation versus the 16 nm four-die implementation of BM1). In addition, in many cases of our simulation, stacking five dies does not produce shorter wirelength than two- to four-die designs. The main reason is because stacking more dies needs more TSVs in general, and inserting more TSVs causes wirelength overhead because of area overhead.

Regarding the critical path delay, stacking three or four dies reduces the critical path delay more effectively than stacking two dies. The largest critical path delay ratio between more-than-two-die designs and two-die designs is about 5% in our simulation (the 16 nm four-die implementation of BM1). However, stacking more than four dies does not reduce the critical path delay effectively. On the other hand, power consumption varies in a very small range. The reason is because gate internal

power is dominant, so the combination of reducing wirelength (a positive effect) and inserting more TSVs (a negative effect due to TSV capacitance) leads to the very small change in total power consumption.

VI. CONCLUSION

In this paper, we investigated the impact of sub-micron TSVs on the quality of today and future 3-D ICs using on GDSII-level layouts. To generate 3-D IC layouts of future 3-D IC layouts, we developed 22 nm and 16 nm process and standard cell libraries based on the ITRS prediction and downscaling trends of other standard cell libraries and Intel process technology. With these realistic libraries, we generated today and future 3-D IC layouts and compared footprint area, wirelength, critical path delay, and power consumption. The simulation results show that 1) footprint area is strongly dependent on the TSV size, so the use of sub-micron TSVs is the most important factor for area reduction; 2) wirelength is also dependent on the TSV size, but if the TSV size is sufficiently small (0.5 μm TSVs for 16 nm technology), shrinking the TSV size further does not help wirelength reduction; 3) critical path delay is strongly dependent on the TSV capacitance, but footprint area also has a nonnegligible effect on critical path delay; 4) transition from 2-D ICs to 3-D ICs does not necessarily lead to less power consumption even when the TSV capacitance is small.

REFERENCES

[1] D. H. Kim, S. Kim, and S. K. Lim, "Impact of sub-micron through-silicon vias on the quality of today and future 3-D IC designs," in *Proc. ACM/IEEE Int. Workshop Syst. Level Interconnect Predict.*, Jun. 2011, pp. 1–8.

[2] G. Katti, M. Stucchi, K. D. Meyer, and W. Dehaene, "Electrical modeling and characterization of through silicon via for three-dimensional ICs," *IEEE Trans. Electron Devices*, vol. 57, no. 1, pp. 256–262, Jan. 2010.

[3] R. E. Farhane, M. Assous, P. Leduc, A. Thuaire, and D. Bouchu *et al.*, "A successful implementation of dual damascene architecture to copper TSV for 3-D high density," in *Proc. IEEE Int. 3-D Syst. Integrat. Conf.*, 2010, pp. 1–4.

[4] M. Motoyoshi, "Through-Silicon Via (TSV)," *Proc. IEEE*, vol. 97, no. 1, pp. 43–48, Jan. 2009.

[5] M. Koyanagi, T. Fukushima, and T. Tanaka, "High-density through silicon vias for 3-D LSIs," *Proc. IEEE*, vol. 97, no. 1, pp. 49–59, Jan. 2009.

[6] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, "Through-silicon-via aware interconnect prediction and optimization for 3-D stacked ICs," in *Proc. ACM/IEEE Int. Workshop Syst. Level Interconnect Predict.*, Jul. 2009, pp. 85–92.

[7] D. H. Kim and S. K. Lim, "Through-silicon-via-aware delay and power prediction model for buffered interconnects in 3-D ICs," in *Proc. ACM/IEEE Int. Workshop Syst. Level Interconnect Predict.*, Jun. 2010, pp. 25–31.

[8] M. Radosavljevic *et al.*, "Electrostatics improvement in 3-D tri-gate over ultra-thin body planar InGaAs quantum well field effect transistors with high-K gate dielectric and scaled gate-to-drain/gate-to-source separation," in *Proc. IEEE Int. Electron Devices Meeting*, Dec. 2011, p. 33.1.1.

[9] S. Hsu *et al.*, "A 280 mV-to-1.1 V 256b reconfigurable SIMD vector permutation engine with 2-dimensional shuffle in 22 nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2012, pp. 178–180.

[10] D. H. Kim, K. Athikulwongse, and S. K. Lim, "A study of through-silicon-via impact on the 3-D stacked IC layout," in *Proc. IEEE Int. Conf. Computer-Aided Design*, Nov. 2009, pp. 674–680.

[11] M. Pathak, Y.-J. Lee, T. Moon, and S. K. Lim, "Through-silicon-via management during 3-D physical design: When to add and how many?," in *Proc. IEEE Int. Conf. Computer-Aided Design*, Nov. 2010, pp. 387–394.

[12] PTM, Predictive technology model [Online]. Available: http://ptm.asu.edu

[13] P. Bai *et al.*, "A 65 nm logic technology featuring 35 nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 μm^2 SRAM cell," in *Proc. IEEE Int. Electron Devices Meet.*, Dec. 2004, pp. 657–660.

[14] K. Mistry *et al.*, "A 45 nm logic technology with high-k+metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging," in *Proc. IEEE Int. Electron Devices Meet.*, Dec. 2007, pp. 247–250.

[15] P. Packan *et al.*, "High performance 32 nm logic technology featuring 2nd generation high-k + metal gate transistors," in *Proc. IEEE Int. Electron Devices Meet.*, Dec. 2009, pp. 1–4.

[16] Nangate, Nangate FreePDK45 Open Cell Library [Online]. Available: http://www.nangate.com

[17] ITRS, International Technology Roadmap for Semiconductors 2007 Edition Interconnect [Online]. Available: http://www.itrs.net

[18] Cadence Design Systems, Encounter Digital Implementation System [Online]. Available: http://www.cadence.com

[19] Y.-J. Lee and S. K. Lim, "Timing analysis and optimization for 3-D stacked multi-core microprocessors," in *Proc. Int. 3-D Syst. Integrat. Conf.*, Nov. 2010, pp. 1–7.

[20] Synopsys, PrimeTime [Online]. Available: http://www.synopsys.com

[21] Y. S. Kim, A. Tsukune, N. Maeda, H. Kitada, and A. Kawai *et al.*, "Ultra thinning 300-mm wafer down to 7-um for 3-D wafer integration on 45-nm node CMOS using strained silicon and Cu/low-k interconnects," in *Proc. IEEE Int. Electron Devices Meet.*, Dec. 2009, pp. 14.6.1–14.6.4.

[22] D. H. Kim, K. Athikulwongse, M. B. Healy, M. M. Hossain, and M. Jung *et al.*, "3-D-MAPS: 3-D massively parallel processor with stacked memory," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2012, pp. 188–190.

[23] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, "TSV-aware interconnect length and power prediction for 3-D stacked ICs," in *Proc. IEEE Int. Int. Technol. Conf.*, Jun. 2009, pp. 26–28.

[24] D. H. Kim and S. K. Lim, "Impact of through-silicon-via scaling on the wirelength distribution of current and future 3-D ICs," in *Proc. IEEE Int. Interconnect Technol. Conf.*, May 2011, pp. 1–3.

Dae Hyun Kim (S'08) received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 2002, and received the M.S. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, GA in 2007. Currently, he is working toward the Ph.D. degree in the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA. His research interests are physical design algorithms for 3-D ICs including design methodology, placement, routing, and design for manufacturability.

Sung Kyu Lim (S'94–M'00–SM'05) received the B.S., M.S., and Ph.D. degrees from the Computer Science Department, University of California, Los Angeles (UCLA), in 1994, 1997, and 2000, respectively. From 2000 to 2001, he was a Post-Doctoral Scholar at UCLA, and a Senior Engineer at Aplus Design Technologies, Inc. He joined the School of Electrical and Computer Engineering, Georgia Institute of Technology in 2001, where he is currently an Associate Professor. His research focus is on the physical design automation for 3-D ICs, 3-D System-in-Packages, microarchitectural physical planning, and field-programmable analog arrays. He is the author of *Practical Problems in VLSI Physical Design Automation* (Springer, 2008).

Dr. Lim received the Design Automation Conference (DAC) Graduate Scholarship in 2003 and the National Science Foundation Faculty Early Career Development (CAREER) Award in 2006. He was on the Advisory Board of the ACM Special Interest Group on Design Automation (SIGDA) during 2003–2008 and received the ACM SIGDA Distinguished Service Award in 2008. He is currently an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS (TVLSI) and served as a Guest Editor for the ACM Transactions on Design Automation of Electronic Systems (TODAES). He has served the Technical Program Committee of several ACM and IEEE conferences on electronic design automation.