More Power Reduction With 3-Tier Logic-on-Logic 3-D ICs

Taigon Song, *Member, IEEE*, Shreepad Panth, *Member, IEEE*, Yoo-Jin Chae, *Student Member, IEEE*, and Sung Kyu Lim, *Senior Member, IEEE*

***Abstract*—Low-power is one of the key driving forces in modern very large scale integration systems. Recent studies show that 3-D integrated circuits (ICs) offer a significant power saving over 2-D ICs. However, these studies are mainly limited to two-tier (2-tier) designs. Thus, in this paper, we extend our target to three-tier (3-tier) 3-D ICs. This paper first shows that the one additional tier available in 3-tier 3-D ICs does offer more power saving compared with their 2-tier 3-D IC counterparts, but more careful floorplanning, through-silicon via management, and block folding considerations are required. Second, we find that the 3-tiers can be bonded in several different ways: 1) face-to-back only; 2) face-to-face and face-to-back combined; and 3) back-to-back and face-to-face combined. This paper shows that these choices pose additional challenges in design optimizations for more power saving. Lastly, we develop effective computer-aided-design solutions that are seamlessly integrated into commercial 2-D IC tools to handle 3-tier 3-D IC power optimization under various bonding style options. With our low-power design methods combined, our 3-tier 3-D ICs provide −14.8% more power reduction over 2-tier 3-D ICs, and −36.0% over 2-D ICs in microprocessor cores under the same performance. In full-chip microprocessors, our 3-tier 3-D ICs provide −27.2% more power reduction over 2-D ICs.**

***Index Terms*—3D IC, floorplanning, TSV, F2F (face-to-face), low power, power reduction.**

I. Introduction

As WE reach the mobile era, power reduction is the keyword that integrated circuit (IC) industry considers as top priority. Not only for mobile devices that require long battery life and energy efficiency, but also for data centers that wish to increase their GHz/Watt performance require to tackle this power reduction issue and have it set as their top priority goal. Power reduction directly links to packaging and cooling cost, and the power consumption of ICs has significant impact on manufacturing yield and reliability. In terms of device perspectives, the development of ultrathin body silicon-on-insulator (fully-depleted SOI) and fin field effect transistor devices also correlates with this power reduction trend [1].

Due to the increasing challenges in design, power, and cost issues that industries were facing beyond 32–22 nm nodes, many have started searching for alternative solutions. In this effort, 3-D ICs using through-silicon vias (TSVs) have gained a great deal of attention as a viable solution for low-power IC designs. Black *et al.* [2] showed that −15% power reduction and +15% performance gain can be achieved by an optimized 3-D floorplan in a two-tier (2-tier) microprocessor, and Jung *et al.* [3] achieved −21.2% power reduction when 3-D floorplan and design techniques were applied. Lee and Lim [4] reported that −21.5% power reduction can be achieved by reducing the bus power in graphics processing units. Kang *et al.* [5] demonstrated 50% leakage and 25% dynamic power reduction in 3-D dynamic random-access memory.

In this paper, we try to answer the following question: "If logic ICs are designed in many-tiers, how much more power reduction can 3-D ICs achieve?" Knowing that previous 3-D IC studies focused on reporting the power reduction in 2-tiers [2]–[4], [6], [7], we try to answer our question by designing three-tier (3-tier) 3-D ICs and study the impact. We visualize the unique design challenges and benefits of 3-tier 3-D ICs, which 2-tier 3-D ICs did not have. We develop computer-aided-design (CAD) tools for various 3-tier 3-D IC design styles, build GDSII-level 3-D IC layouts, and perform optimization and analysis using sign-off CAD tools. Our contributions include the following.

1) To the best of authors' knowledge, we are the first that reported the largest power reduction that 3-D ICs have. Our 3-tier core results show −36% power reduction to the 2-D counterpart [3] and −27.2% in full-chip [6], which is the biggest power reduction achieved among all other previous studies.
2) Three-tier 3-D IC design in mixed bonding styles [e.g., face-to-face and face-to-back combined (F2F+F2B)] help reduce more power. We develop CAD tools and implement various mixed bonding styles to reveal these benefits.
3) Block-folding technique reduce significant power in 3-tier. However, careful design management must be followed, and different bonding styles in mixed bonding impact the design quality in 3-tier block-folding.

Manuscript received October 1, 2015; revised January 15, 2016; accepted March 17, 2016. Date of publication April 5, 2016; date of current version November 18, 2016. This work was supported in part by the Semiconductor Research Corporation under Grant 2239.0001, and in part by Intel Corporation under Grant ICSS Task 2293. This paper was recommended by Associate Editor Y. Chen.

T. Song is with Synopsys Inc., Mountain View, CA 94043 USA (e-mail: taigon.song@synopsys.com).

S. Panth is with Altera Corporation, San Jose, CA 95134 USA (e-mail: spanth@altera.com).

Y.-J. Chae is with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: yoojin@kaist.ac.kr).

S. K. Lim is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: limsk@ece.gatech.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCAD.2016.2550583

0278-0070 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

TABLE I
PDN SPECIFICATIONS USED IN OUR 2-D AND 3-D DESIGNS. # TRACKS SHOW THE MAXIMUM NUMBER OF SIGNAL WIRES THAT CAN FIT IN BETWEEN TWO ADJACENT P/G WIRES

	Local	Intermediate	Global		
	M3	M4 - M6	M7	M8	M9
Metal width/pitch	56/152nm	112/228nm	224/456nm		
PDN density (%)	10.5	14.9	18.0	21.4	24.9
PDN width (nm)	208	340	2048		
PDN pitch (um)	1.976	2.28	11.4	9.576	8.208
# tracks	11	8	20	16	13

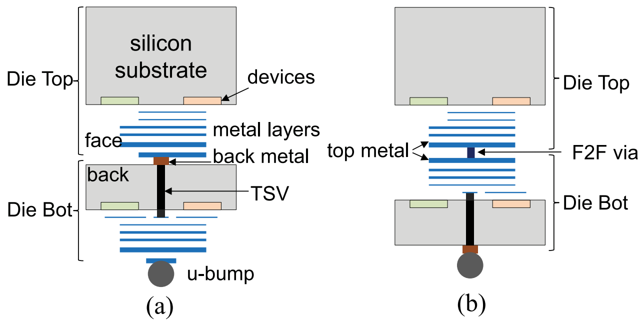


Fig. 1. Basic 2-tier die bonding styles. (a) F2B. (b) F2F.

TABLE II
3-D INTERCONNECT SETTINGS

	Diameter (μm)	Height (μm)	Pitch (μm)	R (Ω)	C (fF)
TSV	3	18	6	0.1	8.4
F2F via	0.5	0.38	1	0.043	0.2

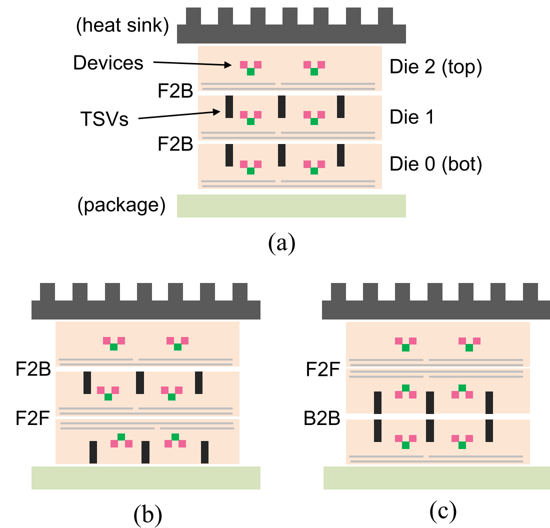


Fig. 2. 3-tier die bonding styles. (a) F2B-only. (b) F2F+F2B. (c) B2B+F2F. Thermal TSVs may be designed on die 2 for heat management (thermal TSVs not assumed in this paper).

II. SIMULATION SETTINGS

This section describes the simulation settings used in this paper. Regarding benchmark, Section II-A describes the benchmark used in Sections IV and V. Benchmark used in Section VI is detailed in Section VI-A.

A. Benchmark

For our 3-tier study, we use OpenSPARC T2 Core (T2 Core) [8] as our benchmark. T2 Core consists of 12 functional unit blocks including two integer execution units (EXUs), a floating point and graphics unit (FGU), an instruction fetch unit (IFU), a load/store unit (LSU), and a trap logic unit (TLU). We synthesize and design our benchmark using Synopsys 28 nm process design kit (PDK) [9]. The PDK allows to use up to nine metal layers, and we use dual-V_{th} [regular V_{th} and high V_{th} (HVT)] standard cells during our design. We include power distribution network (PDN) in our designs. We place fixed PDN at the initial design stage before placement and routing and is targeted to have a density of 25% (M9) to 10% (M3). Table I describes the details of our PDN design. We do not place a fixed PDN for M1 and M2. This is because for M1, standard cells already contain VDD/VSS lines, and a fixed PDN for M2 acts as placement blockages.

B. 3-D Bonding Technology

When stacking 3-D ICs in 2-tier, two bonding styles are possible: 1) face-to-back (F2B) and 2) face-to-face (F2F) (see Fig. 1). In F2B bonding, TSVs are used for vertical interconnects. However, since TSVs penetrate through the silicon substrate and occupy area, using excessive TSVs lead to area overhead, which designers should avoid. On the other hand, F2F is a bonding style where it uses F2F vias for vertical interconnects. Unlike TSVs, F2F vias do not occupy any silicon area due to its advantageous bonding style. Table II summarizes the bonding-style-related settings used in this paper. We assume that TSVs are much bigger than F2F vias since manufacturing reliable submicrometer TSVs are challenging. Resistances and capacitances of the TSVs are calculated based on [10].

In this paper, we study the impact of three different types of bonding styles in 3-tier 3-D ICs: 1) face-to-back only (F2B-only); 2) F2F+F2B; and 3) back-to-back and face-to-face combined (B2B+F2F). As in Fig. 2, each shows F2B-only, F2F+F2B, and B2B+F2F, respectively. In all bonding styles, die 0 is the bottom die where it connects to the package/PCB, and die 2 is the top die where heat sink attaches. For Fig. 2(a), F2B-only is a bonding style that only uses TSVs for 3-D interconnects. For Fig. 2(b), F2F+F2B uses F2F vias for 3-D interconnects between die 0 and die 1, and one TSV layer (in die 1) for die 1 and die 2. The advantage of F2F+F2B is that die 0 and die 1 suffer less from 3-D interconnect penalty (smaller *R* and *C* from F2F vias than TSVs). In addition, since F2F vias do not occupy any silicon area and are smaller than TSVs, more dense and optimal 3-D connection can be made. For Fig. 2(c), B2B+F2F uses F2F vias for die 1 and die 2, and two TSV layers for both die 0 and die 1. Since two TSV layers are used instead of one, B2B+F2F may provide less advantages than Fig. 2(b). However, for systems that have many external I/O connections to the package/PCB would consider B2B+F2F more beneficial than F2F+F2B. In this regard, it makes sense to use B2B+F2F bonding.

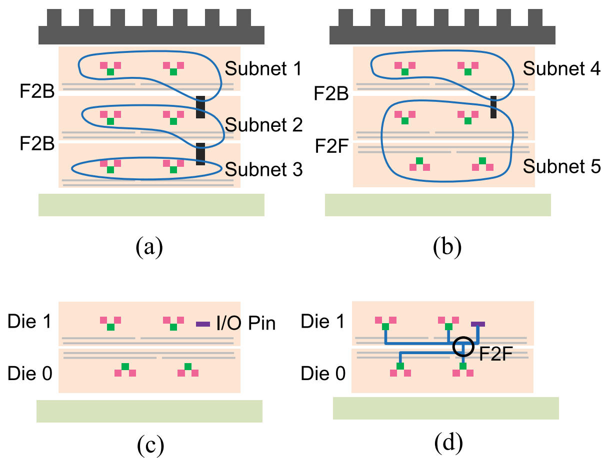


Fig. 3. Net handling and routing in 3-tier mixed bonding. (a) 6-pin net with two TSVs split into one subnet per tier in F2B-only case. (b) F2F bonding does not cause net splitting. (c) Subnet 5 from (b), where the TSV is defined as an I/O pin. (d) Sample routing topology for (c).

III. CAD Tool for 3-Tier 3-D ICs

This section first discusses existing CAD approaches for F2B and F2F 3-D ICs. It also discusses why these approaches are not directly applicable to mixed bonding. Next, it describes how a 3-tier F2B+F2F mixed bonding 3-D IC circuit can be constructed, and it finally shows the modifications required to support a B2B+F2F mixed bonding 3-D IC.

A. Need for New Tools

Kim *et al.* [11] provided a framework for handling TSVs arbitrarily in a many-tier F2B-only 3-D IC. However, the authors primarily compared wirelength, and when it comes to power studies, only 2-tier 3-D ICs have been considered in many previous papers [2]–[4], [6], [7].

In the placement framework proposed in [11], the gates are first partitioned into as many tiers as required. Next, TSVs are inserted into the netlist as large cells. The placement is an iterative force-directed process, with two main forces. The net force $\mathbf{F}_{\mathbf{net}}$ tries to bring all the cells of a given net together, and the move force $\mathbf{F}_{\mathbf{move}}$ tries to remove overlap between cells and TSVs of a given tier. The authors have also demonstrated that it is more beneficial to treat the 3-D net as one subnet per tier (including the TSV), instead of as a single 3-D net, as it leads to more accurate wirelength estimation. This is shown in Fig. 3(a).

When it comes to F2F integration, the placement engine remains more or less the same, with a few differences [6]. First, the F2F vias are not inserted into the netlist, and second, the nets are not split into subnets per tier. This is because the F2F vias are so small that they will be inserted by tricking a 2-D router. Once the placement is complete, the entire 3-D stack is fed into a commercial router to extract 3-D via locations. However, this is limited to two tiers, with at most seven metal layers per tier, as commercial 2-D tools cannot handle more than a total of 15 metal layers.

Clearly, these approaches cannot directly be applied for a circuit with mixed bonding. TSV-based engines require TSVs to be inserted during placement, while F2F engines do not. In addition, the TSV-based engine employs net splitting, while the F2F engine does not. Finally, the F2F planner can handle at most two tiers due to commercial tool limitations. Moreover, B2B requires special handling as the TSVs in both the tiers with the B2B interface needs to be aligned. We now present techniques to handle both F2F+F2B and B2B+F2F mixed bonded 3-D ICs in the following sections.

B. CAD Tool for F2B+F2F Bonding

The modifications made to the placement engine to handle this style of mixed bonding are shown in Fig. 3(b). We perform two major modifications. First, TSVs are inserted into the netlist only in those tiers that are F2B. Next, we perform net splitting, but do not split the nets at the F2F interface. Therefore, a 3-D net spanning three tiers will have only two subnets, instead of three as in the all F2B case. We then perform placement to give us the (x, y) locations of all the gates in the netlist, as well as the TSV locations for the F2B tier.

Now, we need to insert F2F vias using a commercial router in the F2F interface. However, as mentioned previously, commercial tools can only handle two tiers. So, we extract the netlist of those two tiers that are part of the F2F interface as shown in Fig. 3(c). In addition to extracting the connectivity and location of gates, we also need to create additional I/O pins in the same location as where the TSV would have existed. This ensures that the router will construct an accurate topology including the TSV, as shown in Fig. 3(d). Once the F2F locations are extracted, we create separate Verilog/DEF files for each tier, then place, route, and optimize them separately.

C. CAD Tool for B2B+F2F Bonding

Handling B2B+F2F bonding is similar to the F2B+F2F mixed bonding case. We perform net splitting at the B2B interface, and once the placement is complete, we extract the two F2F tiers only to feed into the commercial router. The major difference is that the placer now needs to determine the location of B2B TSVs instead of an F2B TSV.

In the B2B TSV interface, both the TSVs need to be aligned. This implies that the B2B TSV can only be placed in aligned whitespace in both tiers of the B2B interface. First, we enforce the alignment constraint by treating the B2B TSV in both tiers as a single object with a single (x,y) location rather than two separate objects in each tier that need to be aligned. Next, the move force that removes overlap needs to consider both tiers. We achieve this by considering two move forces for this single TSV object–$\mathbf{F}_{\mathbf{move},\mathbf{1}}$ and $\mathbf{F}_{\mathbf{move},\mathbf{2}}$. Each force is computed separately on a per-tier basis to try and remove overlap in that tier. The aggregate move force is then the vector average of these two. Finally, once the placement is done, this B2B TSV is snapped to aligned whitespace in both tiers.

D. 3-Tier 3-D IC Design Flow

To design an optimized 3-tier 3-D IC, we first synthesize the netlist using Synopsys Design Compiler. Then, we perform 3-tier floorplanning using the mixed-bonding tools from

TABLE III
AREA PERCENTAGE OF THE FUNCTIONAL UNIT BLOCKS IN T2 CORE

block	Area (%)	block	Area (%)
LSU	32.1	MMU	5.3
IFU	22.3	IFU_IBU	3.2
FGU	11.5	PKU	1.4
TLU	8.4	GKT	1.3
EXU0	6.3	PMU	1.3
EXU1	6.3	DEC	0.6

the previous sections. We design and layout each die separately using Cadence SoC Encounter. Once the 3-D CAD tools generate the TSV/F2F locations, cells and memory macros are placed using Cadence SoC Encounter. We then extract the parasitics of each die and perform static timing analysis using Synopsys PrimeTime to obtain new timing constraints for each die. With the new timing constraints, we perform timing and power optimizations using Cadence SoC Encounter. We perform several iterations of these optimization steps (from obtaining timing constraints by Synopsys PrimeTime to design optimization in each die using Cadence SoC Encounter). By these steps, we obtain a timing-closed and power optimized design for 3-tier 3-D ICs.

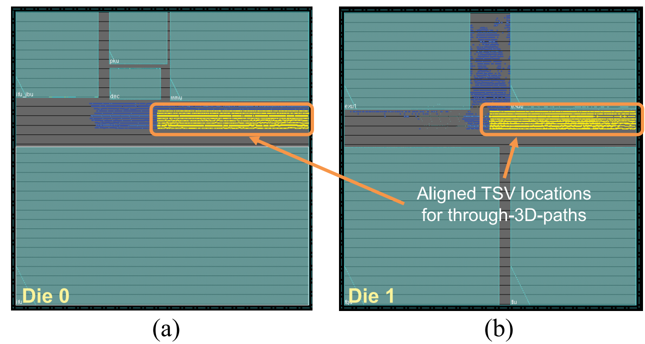
IV. BENEFITS OF 3-TIER 3-D IC

This section studies the challenges and benefits of 3-tier 3-D ICs. Due to the broad scope, this section is limited to F2B-only bonding style in block-level (nonfolded) T2 Core designs.

A. New Design Challenges

When floorplanning a 3-D IC, many design constraints must be considered such as the connection between blocks and top-level pins to external connections. In addition to these constraints, area balance limits many partitioning options in a 3-tier 3-D IC. For T2 Core, Table III shows the area ratio between the blocks inside. We see that the two biggest modules (LSU and IFU) occupy 32.1% and 22.3% of the total T2 Core area. This means that, e.g., when a designer decides to have LSU and IFU at the same die, this die will be significantly larger than the other two since these two blocks consume more than half (54.4%) of the total area. Considering area balance, LSU should not be partitioned to be at the same die with any large blocks (such as IFU, FGU, TLU, EXU, or memory management unit), and the die including IFU should also be carefully partitioned. Having this area balance issue, 3-tier partitioning becomes very challenging, and partitioning becomes even more challenging in many-tier designs.

In T2 Core, several blocks such as an LSU connect to other blocks on all three dies. If a die partition places a block (e.g., LSU) in die 0 and the other connecting block in die 2, die 1 must support the paths that connect blocks in die 0 and die 2. We will call these "through-3-D-paths". Knowing that every block interacts with other blocks in T2 Core, these through-3-D-paths become as many as half of the total TSV count. Many through-3-D-paths enter die 1 through a TSV



(a) (b)

Fig. 4. TSV layers aligned in T2 Core to provide through path for die 0–die 2 connecting nets (through-3-D-paths) in F2B-only (blue dots: regular TSVs, yellow dots: through-3-D-Path TSVs). (a) Die 0. (b) Die 1.

TABLE IV
2-D VERSUS 2-TIER 3-D VERSUS. 3-TIER 3-D (NONFOLDING, F2B-ONLY) IN T2 CORE. ALL PERCENTAGE VALUES ARE WITH RESPECT TO 2-D RESULTS

	2D [3]	2-tier 3D [3]	3-tier 3D (non-folding)
Clock period	1.5ns	1.5ns	1.5ns
Footprint (mm^2)	**3.08**	**1.44 (-53.2%)**	**1.00 (-67.5%)**
Si. Area (mm^2)	**3.08**	**2.88 (-6.5%)**	**3.00 (-2.6%)**
Wirelength (m)	**22.4**	**18.0 (-19.6%)**	**14.3 (-36.2%)**
# Cells	523.4K	420.8K (-19.6%)	403.9K (-22.8%)
# Buffers	**221.7K**	**130.8K (-41.0%)**	**130.7K (-41.0%)**
HVT cells	370.6K	408.3K	377.4K
# TSV	-	6,562	4,118
Total power (mW)	**348.3**	**271.7 (-22.0%)**	**248.1 (-28.8%)**
Cell power (mW)	71.6	62.9 (-12.2%)	62.6 (-12.6%)
Net power (mW)	175.7	137.9 (-21.5%)	117.3 (-33.2%)
Leak. power (mW)	101.1	70.9 (-29.9%)	68.2 (-32.5%)

from die 0 and leave die 1 by a TSV. In this regard, die 1 handles double the number of 3-D connections than the other two tiers. Therefore: 1) providing sufficient white space and 2) an actual "through-path" for through-3-D-paths is very important in 3-tier design. As in Fig. 4, we align white space of the top and bottom 3-D connections so that these through-3-D-paths do not need to detour.

B. 2-D Versus 2-Tier 3-D Versus 3-Tier 3-D

We now compare our 2-D and 3-D block level T2 Core designs in TSV only bonding style. First, all our designs run in a target clock period of 1.5 ns (=677 MHz). Note that the run speed of our designs are much slower than UltraSPARC T2, a commercial product of OpenSPARC T2, that runs at 1.4 GHz [12]. This is because some custom memory blocks in T2 Core such as content-addressable memory are synthesized with cells, because a general memory compiler cannot handle these kind of memories. Unfortunately, these synthesized memories run slower than the memory macros generated by a memory compiler. Second, our baseline 2-D and 2-tier 3-D follow the floorplan and designs done in [3]. However, since the designs in [3] did not have PDN, we include PDN in our 2-D and 2-tier 3-D designs and made minor modifications to meet the timing.

Table IV compares various metrics between 2-D, 2-tier 3-D, and 3-tier 3-D in T2 Core designs, and Fig. 6(a) and (b) shows

GDSII layouts of our 2-D and 3-tier nonfolded 3-D design in F2B-only bonding, respectively. 2-tier 3-D applies all design techniques proposed in [3]. First, by having 3-tier 3-D design, we reduce the total wirelength by −36.2% and cell count by −22.8%. Compared to 2-tier 3-D, we reduce −16.6% more wirelength and −3.2% more cell count. The significant wirelength reduction comes from the smaller footprint and better top-level floorplanning.

Second, and most importantly, 3-tier 3-D (nonfolding) reduces the total power by −28.8%, where 2-tier 3-D (block-folding) reduces −22.0% (Note that our 2-tier 3-D design reduces −0.8% more power than reported in [3]). In spite of not applying block-folding in our 3-tier 3-D yet, better 3-tier floorplan gives more net power reduction than in 2-tier 3-D (−20.6 mW more). Three-tier 3-D achieves power reduction by cell count reduction, and wirelength saving. However, significant wirelength saving largely contributes to this power reduction than reduction in cell count which is not as significant (small cell and leakage power reduction). Last, we reduce the footprint by −67.5%. This is −14.3% more reduction than the 2-tier 3-D design. In terms of silicon area, 3-tier 3-D still uses −2.6% less area than 2-D. Three-tier 3-D uses more silicon area than 2-tier 3-D since it requires to manage more TSVs on the top-level. However, the footprint/silicon area reduction stems from the significant wirelength and cell count reduction.

V. Bonding Style Impact Study

The previous section showed 3-tier designs in F2B-only (TSV) bonding. Thus, this section studies how various 3-tier bonding styles enhance design quality and reduce power in T2 Core.

A. Bonding Impact on Floorplan

1) F2B-Only Versus F2F+F2B Bonding: As described in Section II-B, F2F bonding provides many advantages over the F2B bonding. Even in 2-tier 3-D ICs, F2F reduces more power than F2B-only bonding style. Thus, it is advantageous to use F2F bonding in 3-tier designs too. However, if one layer is bonded in F2F style, the other 3-D layer must be designed in F2B as bonding style. Therefore, having nonfolded F2B-only T2 Core as our baseline, we compare how the top-level design quality changes when we apply F2F+F2B bonding in 3-tier.

Fig. 5 compares how the top-level design changes in die 0 of T2 Core in F2F+F2B bonding. Note that the floorplan is exactly the same in both designs. First, we see that F2F placement quality is much better than that of the TSV placement. Many top-level 3-D connections form between die 0 and die 1 (2176 TSV/F2F vias), and placing 2176 TSVs consumes a large space due to the relatively large TSV size. In addition, TSV landing pads in die 1 must not overlap with the top-metal PDN. In this regard, placing 2176 TSVs on the top-level requires more space than before. This forces the TSVs to be placed on suboptimal locations. As in Fig. 5(a), we see that TSVs are crowded and their locations become suboptimal. However, since F2F vias occupy smaller footprint than TSVs, F2F vias can be placed on its optimal location and become less affected by the PDN. Second, because of the better F2F

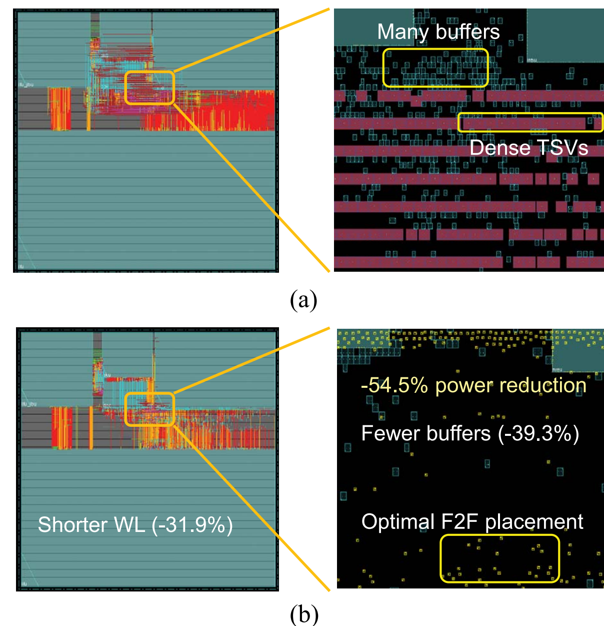


Fig. 5. F2F vias for better design in F2F+F2B bonding under the same floorplan in T2 Core. (a) F2B-only (TSVs for 3-D connection). (b) F2F+F2B (F2F vias for 3-D connection).

via locations and small *RC* parasitics, top-level design quality in F2F bonding improves significantly. In die 0, wirelength reduces by −31.9% and buffer count reduces by −39.3%. This translates to −54.5% top-level power reduction than F2B-only in die 0.

2) F2F+F2B Versus B2B+F2F Bonding: For various reasons, B2B+F2F bonding can be chosen over F2B-only or F2F+F2B bonding. The difference between F2F+F2B bonding and B2B+F2F bonding lies on the second 3-D interconnect layer [see Fig. 2(b) and (c)]. However, in B2B+F2F bonding style, TSVs must be placed at the same location in die 0 and die 1. Depending on designs, the initial floorplan may not align whitespace on both dies. In addition, TSV parasitics double in B2B+F2F because it uses two TSVs for 3-D connection instead of one.

Fig. 7 illustrates the design changes on die 1 of T2 Core in our B2B+F2F example compared with F2F+F2B. F2F+F2B and B2B+F2F has the same floorplan, but die 0 and die 2 are swapped to utilize the F2F bonding for layer with more 3-D connection. Fig. 7(b) shows that EXU changed its aspect ratio to provide white space for the top-level TSVs. LSU in die 0 occupies significant area, and this forces the TSVs in die 0 and die 1 to be placed on the top of the layout. However, due to this, die 1 in B2B+F2F bonding could not provide a through-3-D-path because the white space between die 0–die 2 cannot be aligned. Comparing the top-level design in die 1 (B2B+F2F versus F2F+F2B), the buffer count increases by +10.7% and wirelength increases by +14.3% in B2B+F2F design. In terms of the top-level power, this is +22.0% increase than the F2F+F2B in die 1.

B. Bonding Impact On Block-Folding

1) F2F+F2B Bonding on Folded Blocks: Block-folding in mixed bonding leaves the designer to choose the right 3-D

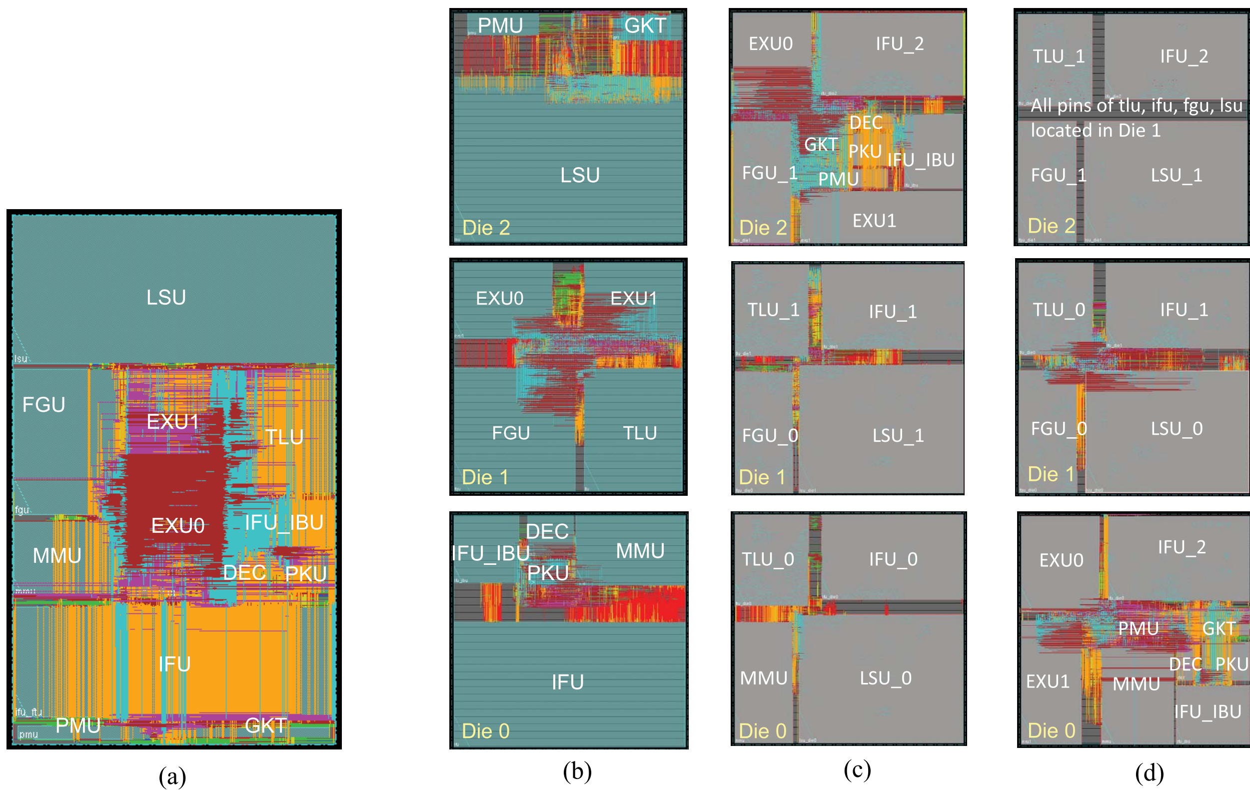


Fig. 6. GDSII layouts of various 3-tier T2 Core designs. (a) 2-D based on [3]. (b) 3-tier nonfolding in F2B-only. (c) 3-tier block-folding in F2B-only. (d) 3-tier block-folding in F2F+F2B.

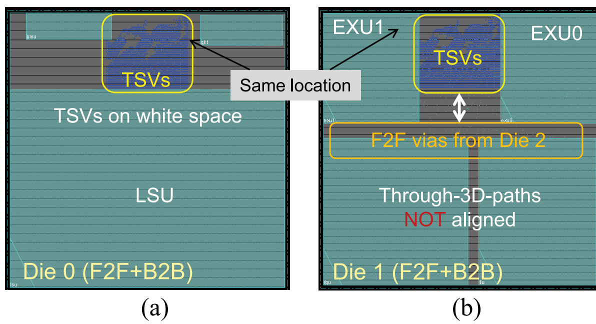


Fig. 7. Through-3-D-paths between die 1 TSV and die 2. F2F vias not aligned in B2B+F2F bonded T2 Core because TSVs must be placed both in die 1 and die 2 (see Fig. 4 for comparison).

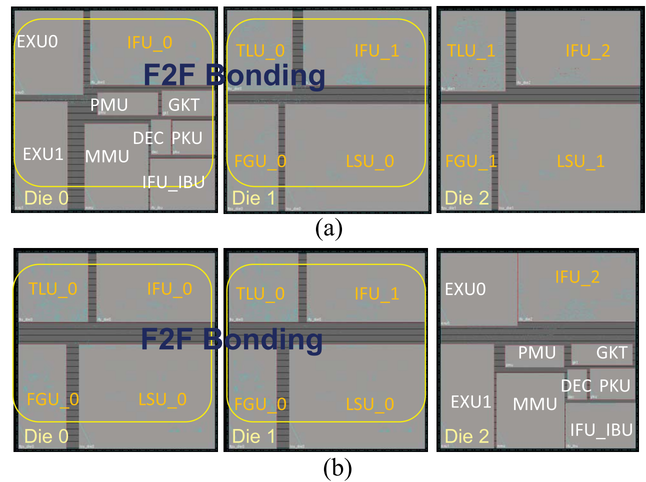


Fig. 8. F2F bonding choice for more power reduction in F2F+F2B bonded T2 Core. (a) F2F bonding for top-level. (b) F2F bonding for block-folding (folded blocks in orange font).

bonding for the right purpose. In a 2-tier design when the bonding style is decided to be F2F (or F2B), this means that both folded blocks and the top-level design utilize F2F layer. However, in 3-tier designs, we must decide how to utilize its F2F layer since it can have only one due to the bonding technology. The more the designer chooses to use F2F layer for block-folding, the less it can be used for top-level design, and vice versa. To study which is more beneficial in T2 Core, we studied two floorplans: 1) using F2F layer for top-level design (F2F+F2B V1) and 2) using F2F layer for block-folding (F2F+F2B V2) (see Fig. 8).

Our results show that F2F+F2B V1 reduces more power than F2F+F2B V2. F2F+F2B V1 showed −36.0% power reduction, but F2F+F2B V2 showed −34.7% power reduction than 2-D. We explain this through the following reasons.

First, extra power reduction from F2F bonding in folded blocks is not significant. Block-folding-based 3-tier designs must consider: 1) power reduction of the block itself from block-folding and 2) options for better connectivity in the top level. For power reduction of single blocks by block-folding in standalone designs, the total power reduction from F2F bonding is only −5.3 mW. This is −1.5% of the total T2 Core power. Note that we are not seeing significant power reduction from folded blocks in F2F bonding. This is because 3-tier floorplanning limits many partitioning options for block-folding in F2F.

TABLE V
COMPARISON AMONG 3-TIER T2 CORE DESIGNS BUILT WITH VARIOUS OPTIONS IN FOLDING AND BONDING STYLES. ALL FOLDED DESIGNS TARGET FOUR BLOCKS (LSU, IFU, TLU, AND FGU) TO BE FOLDED

	2D	Non-Folding			Block-Folding		
	[3]	F2B-only	F2F+F2B	B2B+F2F	F2B-only	F2F+F2B	B2B+F2F
Clock period	1.5ns	1.5ns	1.5ns	1.5ns	1.5ns	1.5ns	1.5ns
Footprint (mm^2)	3.08	1.44 (-53.2%)	1.44 (-53.2%)	1.44 (-53.2%)	1.44 (-53.2%)	1.44 (-53.2%)	1.44 (-53.2%)
Si. Area (mm^2)	3.08	3.00 (-2.6%)	3.00 (-2.6%)	3.00 (-2.6%)	3.00 (-2.6%)	3.00 (-2.6%)	3.00 (-2.6%)
Wirelength (m)	22.4	14.3 (-36.2%)	13.8 (-38.4%)	14.1 (-37.1%)	13.4 (-40.2%)	13.0 (-42.0%)	13.2 (-41.1%)
# Cells	523.4K	403.9K (-22.8%)	394.3K (-24.7%)	395.7K (-24.4%)	370.9K (-29.1%)	368.8K (-29.5%)	370.4K (-29.2%)
# Buffers	221.7K	130.7K (-41.0%)	124.9K (-43.7%)	126.4K (-43.0%)	117.8K (-46.9%)	114.2K (-48.5%)	115.6K (-47.9%)
HVT cells	370.6K [70.7%]	377.4K [93.4%]	372.0K [94.3%]	374.4K [94.6%]	348.6K [93.9%]	346.0K [93.8%]	347.9K [93.9%]
# TSV	-	4,118	4,118	6,060	8,688	9,231	12,986
Total power (mW)	**348.3**	**248.1 (-28.8%)**	**242.6 (-30.3%)**	**244.4 (-29.8%)**	**229.7 (-34.0%)**	**223.1 (-36.0%)**	**227.4 (-34.7%)**
Cell power (mW)	71.6	62.6 (-12.6%)	62.1 (-13.3%)	62.0 (-13.4%)	54.1 (-24.4%)	54.0 (-24.6%)	54.1 (-24.4%)
Net power (mW)	175.7	117.3 (-33.2%)	113.3 (-35.5%)	115.0 (-34.5%)	107.7 (-38.7%)	102.0 (-41.9%)	105.4 (-40.0%)
Leak. power (mW)	101.1	68.2 (-32.5%)	67.1 (-33.6%)	67.4 (-33.3%)	67.9 (-32.8%)	67.2 (-33.5%)	67.9 (-32.8%)

Second, top-level design quality in F2F+F2B V1 is better than F2F+F2B V2. F2F+F2B V1 and V2 uses 52% more top-level 3-D connections (TSV count: 2573) than F2B-only–block-folding design for top-level connection (TSV count: 1693). However, since the optimal white spaces for TSV location are limited, this leads to worse TSV locations and design quality in the top level. In fact, the top-level design quality in V2 is worse than F2B-only–block-folding design. However, note that F2F+F2B V1 uses F2F layer for top-level design. Despite the increased top-level F2F via count than F2B-only–block-folding design, F2F+F2B V1 provides better top-level design quality, and provides more power reduction than F2B-only–block-folding design (top-level design quality: F2F+F2B V1 > F2B-only–block-folding design > F2F+F2B V2). Comparing the top-level design quality, V1 achieves −17.3% cell count and −20.4% wirelength reduction and −29.4% total top-level power reduction than F2B-only–block-folding design. Better top-level design quality leads to more power reduction in blocks, because it requires the blocks to use less resources to optimize the boundaries. Therefore the design quality impact by better top-level design cannot be ignored.

2) B2B+F2F Bonding on Folded Blocks: B2B+F2F bonding leads to a 3-D layer using B2B bonding. Therefore, if top-level design uses F2F layer, blocks must use B2B layer for block-folding. Since Section V-A2 revealed the impact of B2B bonding on the top-level, it is important to study how the design quality of folded blocks change in B2B bonding. We designed 2-tier standalone blocks in T2 Core (LSU, FGU, TLU, and IFU), and results showed that F2B, B2B, and F2F bonding reduces block power compared to 2-D (in average) by −5.9%, −2.4%, and −8.3%, respectively. We report that B2B bonding shows the least power reduction among all other bonding styles. This is mainly due to the increased TSV RC parasitics (2× than F2B), occupying silicon area and TSV alignment issues in B2B bonding.

C. Overall Comparison

Table V compares all designs that we have done in this paper based on whether block-folding technique is applied and the bonding style. GDSII layouts of our designs are illustrated in Fig. 6, and designs that are not shown in the figure (such as nonfolding–B2B+F2F) are based on a similar design as what is shown in Fig. 6. First, we emphasize that we achieve a maximum of −36% power reduction in block-folded–F2F+F2B design. This is 14.8% more reduction than what was reported in [3], and the most power reduction reported in any previous studies. Second, block-folding provides more power reduction than nonfolding. In terms of bonding style, F2F+F2B reduces most power, followed by B2B+F2F and F2B-only style. However, to visualize more power reduction from these design techniques, we note that more careful floorplanning and design must be done.

VI. DESIGN CHALLENGES IN FULL-CHIP

This section describes the design challenges and results in full-chip 3-tier T2. Bigger design scale provides unique challenges in various metrics. For a thorough and comprehensive study, we provide six different full-chip designs based on block-folding and different bonding styles.

A. Full-Chip OpenSPARC T2 Design

The full-chip scale OpenSPARC T2 consists of 53 blocks including eight SPARC cores (T2 Core), eight L2-cache data banks (L2D), eight L2-cache tags (L2T), eight L2-cache miss buffers (L2B), and a cache crossbar (CCX). We synthesize each block with Synopsys 28 nm cell libraries [9] as in T2 Core. We remove seven blocks that do not directly affect the CPU performance from our implementation including five SerDes blocks, an electronic fuse, and a miscellaneous I/O unit. In addition, we replace the phase-locked loop (analog block) in a clock control unit by ideal clock sources. Thus, a total of 46 blocks are floorplanned. We use the same netlist as in the previous work [6], and our baseline 2-D follows the full-chip T2 floorplan and designs done in [6]. However, since these designs did not have PDN, we include PDN in our 2-D and made minor modifications to meet the timing.

B. Area Management Challenges

In IC designs, managing a small area is important for low cost. Therefore, 3-D ICs should also be designed in

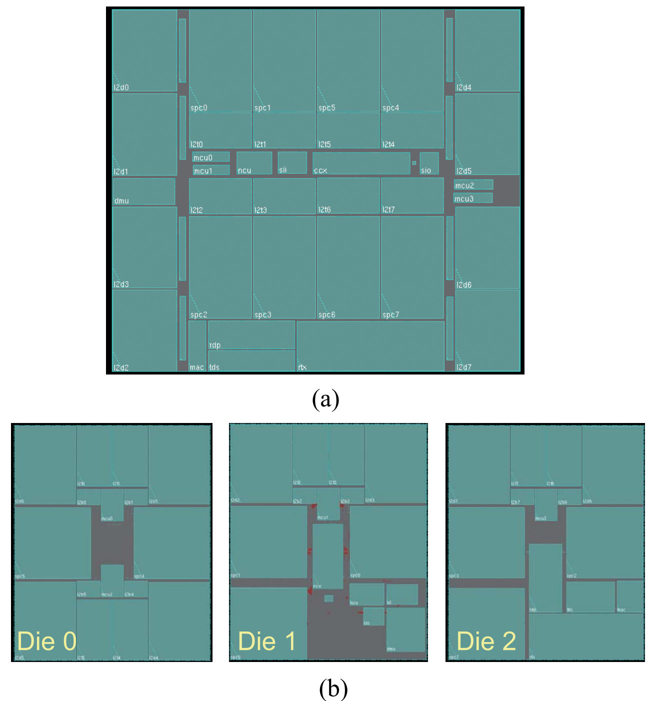
TABLE VI
AREA COMPARISON BETWEEN 2-D AND 3-D IN FULL-CHIP LEVEL STUDIES

	[7]	[6]	[4]	This study
2D silicon area (mm^2)	5.5225	71.1	8.2	71.1
3D footprint (mm^2)	3.1725	38.4	4.1	24.3
3D silicon area (mm^2)	6.345	76.8	8.2	72.9
Increase (%)	14.9	8.0	0	2.5

the smallest area possible. In Section IV-B and in previous studies [3], 3-D ICs are reported to have the benefit of designing modules in a smaller area due to the reduced wirelength and buffer count. However, this may not always be true when designing ICs in full-chip scale. Table VI shows how 3-D ICs are bigger to their counterpart 2-D in previous studies [4], [6], [7].

Notice that in full-chip scale studies, 3-D ICs do not consume less silicon area than the 2-D. For example, in [7], 2-D is 5.5225 mm^2 and their 2-tier 3-D is 6.345 mm^2 (+14.9% more area). In the previous full-chip scale study done in T2 [6], 3-D uses 8.0% more silicon area than 2-D. This is because of the following reason. We will explain this in example of T2: having 46 modules in full-chip requires significant effort on floorplanning to maintain a small footprint. In T2, the area difference between the biggest module (core) and the smallest module (SIO) is more than 16×. Therefore, managing a small-footprint floorplan is a challenging task in both 2-D and 3-D. However, floorplanning problem becomes more complicated in 3-D ICs. For example, 2-tier 3-D ICs require managing two seamless floorplans using only half of the number of total modules. Floorplanning becomes harder when there are less number of modules to place. In 3-tier 3-D ICs, it becomes even more challenging because designer must floorplan three surfaces using 1/3 of modules that the original 2-D has. Many design constraints must be met in full-chip design, and these design constraints conflict with area management. However, note that a more complicated floorplanning problem in 3-tier do not always lead to more area consumption. In comparison with [6], our 3-tier design consumes less silicon area (72.9 mm^2) than a 2-tier full-chip (76.8 mm^2). Fig. 9 shows a comparison between 2-D and 3-tier full-chip floorplan. Our 3-tier full-chip consumes more silicon area (+2.5%), but note that the white space inside the 3-tier floorplan is also larger than 2-D. In fact, all increased silicon area and the area saved from designing smaller modules in 3-D remains as empty space since floorplanning in 3-tiers is a challenging task.

Having different chip sizes in different dies may be a viable solution to area management. While wafer-to-wafer (W2W, [13]) bonding cannot have different sized ICs on each tier, chip-to-wafer (C2W, [14]) or chip-to-chip (C2C, [15]) bonding provides possibilities to use differently-sized dies in different tiers. However, C2W and C2C bonding comes with inferior accuracy and cost than using W2W bonding. Smaller dies are required to be handled with more advanced equipments, and in addition to this, handling smaller chip-scale dies result in reduced placement accuracy [16]. In some cases, smaller dies may not be able



(a)

(b)

Fig. 9. White space (= gray area) in T2 full-chip. (a) 2-D floorplan (9 mm × 7.9 mm). (b) 3-tier 3-D floorplan (4.5 mm × 5.4 mm). More silicon area used in 3-D remains as white space due to floorplanning challenges.

to be bonded in C2C or C2W style due to the equipments. Therefore, designers must choose the 3-D partition and floorplan wisely based on various design factors including these different chip bonding styles.

C. Block-Folding in Full-Chip

Block-folding in 3-tier is more challenging in full-chip due to the increased design complexity. We describe our block-folding strategies and show how this is different from 2-tier.

1) How Many Blocks Can We Fold?: In addition to regarding area balance in Section IV-A, the actual area that can be used for folding reduces due to the reduced footprint. Therefore, designers must properly choose what blocks to fold based on power reduction and floorplanning benefits. Fig. 10 shows how the area for folding reduces in 3-tier full-chip layout. As in Fig. 10(b), 2-tier 3-D allows to fold five different modules (core, RTX, L2D, L2T, and CCX) [6]. Because of the reduced footprint in the folding die in die 1, 3-tier only allows to fold four modules. However, notice that different number of tiers stem distinctive challenges. For example, a 4-tier 3-D will have different folding constraints of a 3-tier design. For example, 4-tier design can use die 0–die 1 and die 2–die 3 for folding since this would not overlap to each other.

2) Block-Folding Design Strategies in Full-Chip: We describe how our 3-tier full-chip floorplan with block-folding is done considering all challenges described in the previous sections. Though this is an example for OpenSPARC T2 architecture, the basic ideas can further extend to other microprocessor architectures as well. First, we perform 3-tier-folding only on cores and RTX. 3-tier folding may provide more power reduction than 2-tier folding. However, a 3-tier folded block

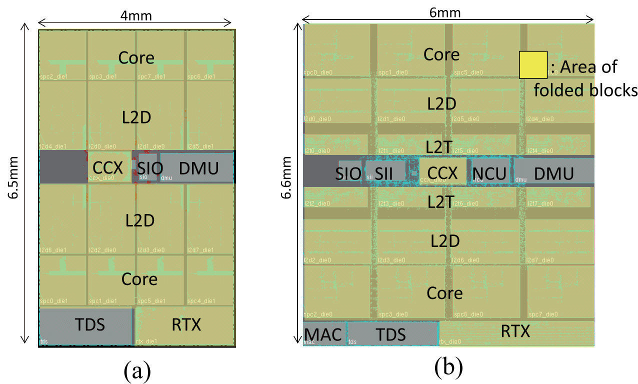


Fig. 10. How folding area reduces in 3-tier designs. Footprint reduction in 3-tier leads to less folded blocks. (a) Die 1 in 3-tier. (b) Die 1 in 2-tier [6].

becomes a floorplan/routing blockage in all 3-tiers. These folded blocks cause routing problems when they are placed in the middle of the die. Thus, these 3-tier blocks are placed on the top and bottom of the floorplan. Fig. 11 shows how our block-folding design strategy is applied in the layout.

Second, CCX and L2Ds are folded in 2-tiers. L2Ds do not reduce much power when it is folded, but we fold it for a better top-level floorplan. When deciding a floorplan, huge-sized modules are not preferable because of the reduced design freedom on the top-level. Especially for hard modules that the designer cannot change its size freely, it is more advantageous to have its size as small as possible. L2D is a module that consists of 32 memory macros so that the size changing is not easy. Therefore, we fold L2D into 2-tiers. L2Ds were the biggest module inside the top-level block-folding floorplan before folding, but the size of its 2-tier footprint is now comparable to other modules in the top-level floorplan.

Third, modules that are heavily connected to each other are gathered together. In fact, L2$s (L2D, L2T, L2B, and memory controller unit) are heavily connected to each other. To utilize the block-folding space efficiently, die 1 is used for folded L2Ds, and other L2$s are placed on die 0 and die 2. However, folding restriction from die 1 limits some L2Ds being placed on its suboptimal locations. Therefore, we choose die 0–die 1 L2Ds to be placed on the side which provides the best floorplan for L2$s, and die 1–die 2 L2Ds are placed on the middle of the chip. However, due to this, the L2$ floorplan in die 2 becomes inferior than die 0. For best L2$ connections, L2D4–L2D7 I/Os are assigned on die 0 and L2D0–L2D3 I/Os are assigned on die 2. In addition to L2$s, network interface unit (NIU) modules (TDS, RDP, MAC, and RTX) are heavily connected to each other and do not have many connections to other modules. Therefore, all NIU modules are gathered on the bottom of the chip. DMU, NCU, and SIU modules (SIO and SII) have many connections to each other, so they are gathered as well.

Finally, I/O pins of the folded modules are properly managed. In the OpenSPARC architecture, cores do not directly connect with L2$s. In fact, most of the core I/Os connect to CCX, and CCX connects to L2Ts. Having this architecture, and knowing that L2Ts are placed on die 0 and die 2,

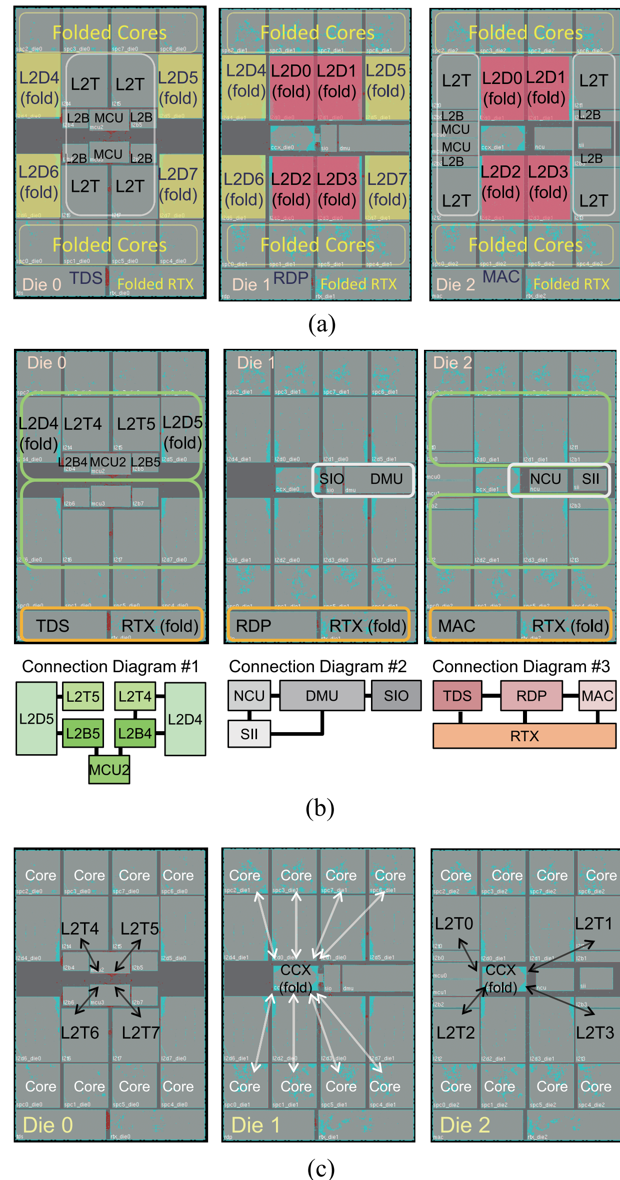


Fig. 11. Full-chip block-folding floorplan strategies. (a) 3-tier folded modules and L2$ floorplan. Die 1 is utilized to place folded L2Ds, and other L2$s are placed on die 0 and die 2. Corresponding L2D pins are placed on each die. (b) How highly-connective modules are placed closely to each other and its connection diagram. (c) L2T–CCX and CCX-core pin assignment—core-CCX I/Os are placed/routed in die 1, and L2T–CCX I/Os are placed/routed in die 0, and die 2 to reduce congestion.

core I/Os and CCX I/Os must be managed properly. Core-CCX I/Os are placed/routed on die 1 and CCX–L2T I/Os are placed/routed on die 0 and die 2. By this technique, we resolve significant congestion between CCX–L2T and CCX-core in top-level design.

D. Managing Bonding Styles in Full-Chip

Managing an adequate bonding style is also important for more power reduction in full-chip designs. Comparing Tables V and VII, we notice some differences that occur in nonfolded full-chip designs compared to single core designs: First, we do not obtain significant power reduction when we

TABLE VII
FULL-CHIP COMPARISON AMONG 3-TIER 3-D IC DESIGNS BUILT WITH VARIOUS OPTIONS IN FOLDING AND BONDING STYLES. ALL FOLDED DESIGNS TARGET FOUR BLOCKS (CORE, RTX, CCX, AND L2D) TO BE FOLDED

	2D	Non-Folding			Block-Folding		
	[3]	F2B-only	F2F+F2B	B2B+F2F	F2B-only	F2F+F2B	B2B+F2F
Clock period	2ns	2ns	2ns	2ns	2ns	2ns	2ns
Footprint (mm^2)	71.1	24.3 (-65.8%)	24.3 (-65.8%)	24.3 (-65.8%)	25.8 (-63.7%)	25.8 (-63.7%)	25.8 (-63.7%)
Si. Area (mm^2)	71.1	72.9 (+2.5%)	72.9 (+2.5%)	72.9 (+2.5%)	77.4 (+8.9%)	77.4 (+8.9%)	77.4 (+8.9%)
Wirelength (m)	343.0	248.1 (-27.7%)	247.0 (-28.0%)	246.6 (-28.1%)	234.4 (-31.7%)	227.1 (-33.8%)	228.7 (-33.3%)
# Cells	7.56M	6.48M (-14.3%)	6.43M (-14.9%)	6.44M (-14.8%)	5.99M (-20.8%)	5.92M (-21.6%)	5.95M (-21.3%)
# Buffers	3.05M	1.97M (-35.4%)	1.92M (-37.0%)	1.93M (-36.7%)	1.69M (-44.6%)	1.62M (-46.8%)	1.65M (-45.9%)
HVT cells	6.57M [86.9%]	6.06M [93.5%]	6.02M [92.9%]	6.03M [93.6%]	5.46M [91.1%]	5.44M [91.8%]	5.44M [91.4%]
# TSV	-	4,599	4,599	6,842	55,142	82,743	93,185
Total power (W)	**8.614**	**6.695 (-22.3%)**	**6.649 (-22.8%)**	**6.654 (-22.7%)**	**6.406 (-25.6%)**	**6.275 (-27.2%)**	**6.335 (-26.5%)**
Cell power (W)	1.757	1.525 (-13.2%)	1.521 (-13.4%)	1.523 (-13.1%)	1.431 (-18.6%)	1.421 (-19.1%)	1.425 (-18.9%)
Net power (W)	4.770	3.327 (-30.3%)	3.294 (-30.9%)	3.290 (-31.0%)	3.231 (-32.3%)	3.120 (-34.6%)	3.167 (-33.6%)
Leak. power (W)	2.087	1.843 (-11.7%)	1.835 (-12.1%)	1.841 (-11.8%)	1.744 (-16.4%)	1.734 (-16.9%)	1.743 (-16.5%)

move from F2B-only to F2F+F2B bonding. Second, the power penalty from F2F+F2B to B2B+F2F is not significant.

1) Advantages of F2F Bonding: In nonfolded T2 Core, we achieved -1.5% more power reduction when we chose from F2B-only to F2F+F2B bonding (Table V). However, in nonfolded T2 full-chip, we obtain only -0.6% more. We explain this through the following: in core, top-level routing required many I/Os to be connected between modules. Due to this, nonfolded core must have TSVs in particular spots. Therefore, TSVs were crowded on its suboptimal locations (see Fig. 5). However, in our full-chip, I/Os that are connecting to other blocks are relatively sparse compared to core due to careful I/O managing. Note that TSV count in die 0 is 2176 in core and 2356 in full-chip. Despite that the design size increased by more than $20\times$, TSV count is similar to each other.

To obtain more power reduction from F2F bonding, the initial F2B design requires to: 1) have many TSVs and 2) these TSVs should be congested so that it cannot find its optimal locations. F2F+F2B core could benefit more from F2F bonding since it met these two criteria. However, I/Os are managed to have less TSVs with less congestion in our full-chip. In addition, full-chip design has signficant white space for TSVs. TSVs already find its optimal spot during TSV placement. Therefore, we do not see significant benefit from F2F bonding. Comparing Fig. 5 from Fig. 12, notice that TSVs in full-chip are already placed in its optimal location. In summary, due to the good TSV locations full-chip F2B-only nonfolded design provide, it does not show significant power reduction when full-chip design moves to F2F+F2B bonding.

2) Managing B2B Bonding: In nonfolded T2 Core, B2B+F2F bonding consumes $+0.5\%$ more power than F2F+F2B bonding. However, in nonfolded full-chip, B2B+F2F bonding consumes only $+0.1\%$ more power than F2F+F2B bonding. This is because our B2B+F2F design did not have many issues with placing TSVs on both dies. B2B bonding becomes a significant design issue when TSVs cannot find white spaces to be placed on both dies. However, in full-chip level where TSVs have sufficient space to be placed, B2B bonding will not become a significant handicap compared to F2B bonding style. Notice that in

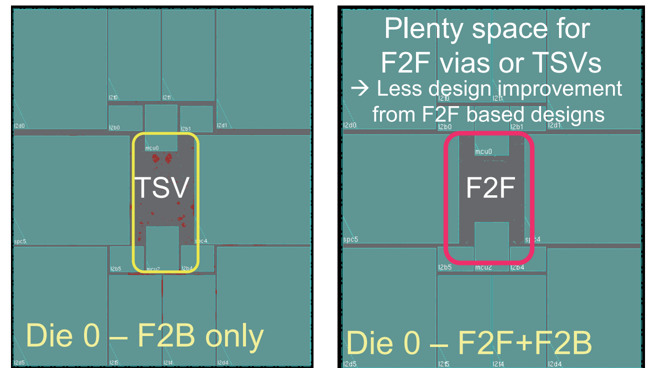


Fig. 12. TSV/F2F placement in full-chip. Because TSVs are placed in its optimal locations (left) due to less congestion and large whitespace, F2F bonding (right) does not provide significant benefits over TSVs.

our full-chip design, TSVs can easily be placed on both sides of chip, and this leads to almost negligible penalty when using B2B bonding. In summary, block-level full-chip designs did not show significant difference between different bonding styles. Maximum bonding style impacts came from block-folded full-chip designs, and this is because of the design benefits/issues that rise from more 3-D connections.

E. Overall Comparison in Full-Chip

Table VII compares all full-chip designs we have done based on whether block-folding technique is applied and the bonding style. GDSII layouts of our designs are illustrated in Fig. 13, and designs that are not shown in the figure (F2B-only and B2B+F2F bonding styles on both nonfolded and block-folded full-chip) are based on a similar floorplan of what is shown in Fig. 13. First, we emphasize that we achieve a maximum of -27.2% power reduction in block-folded-F2F+F2B design. This is -6.9% more reduction than what was reported in [6]. Note that our power reduction from 3-tier design is almost similar to one technology difference. We emphasize that this is the maximum power reduction reported in any kind of full-chip studies. Second, similar as T2 Core results in Section V-C, block-folding provides more power reduction than nonfolding. In terms of bonding style, F2F+F2B reduces most power, followed by B2B+F2F and F2B-only style. For maximum

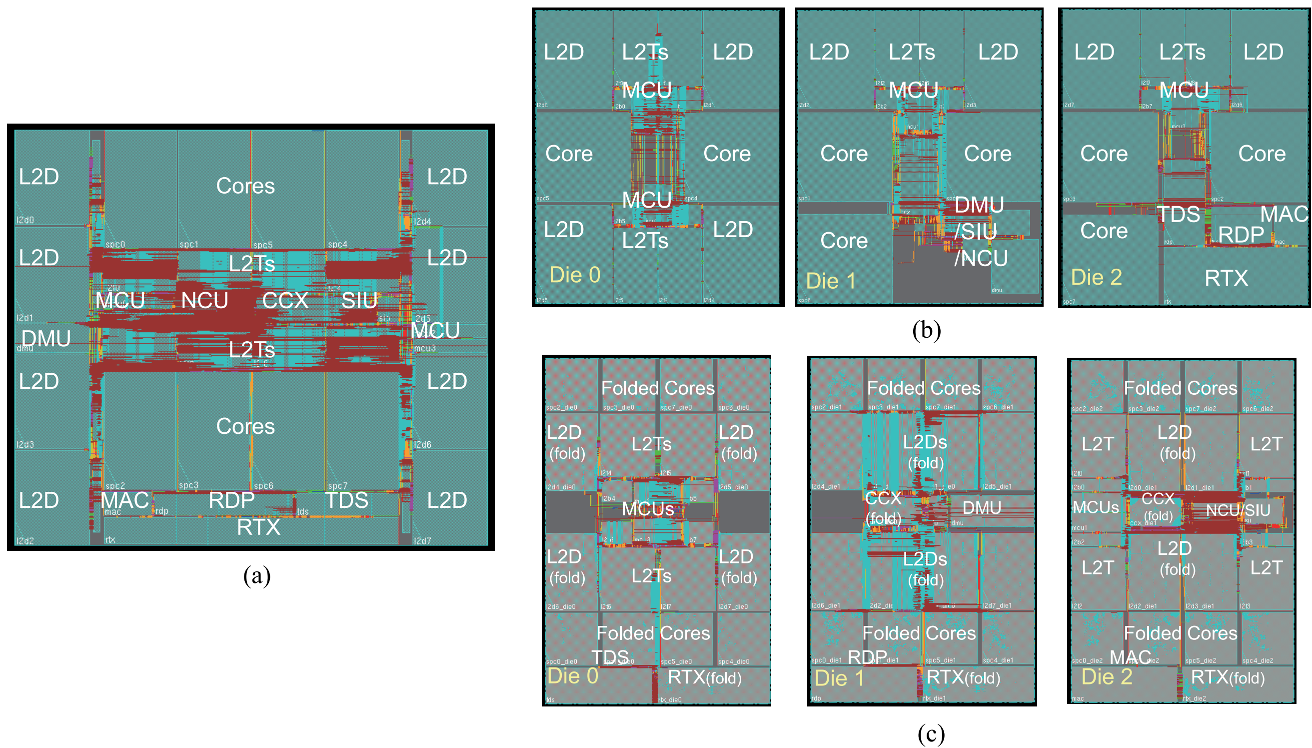


Fig. 13. GDSII layouts of various full-chip 3-tier 3-D IC designs in F2F+F2B bonding. (a) 2-D based on [6]. (b) 3-tier nonfolding. (c) 3-tier block-folding.

power reduction in 3-tier 3-D ICs, all 3-D design techniques we have mentioned in this paper such as floorplanning, pin assignment, block-folding, and TSV assignment should be carefully managed.

VII. Conclusion

In this paper, we demonstrated power reduction benefits that 3-tier 3-D IC design provides in OpenSPARC T2. First, we showed that one additional tier in 3-tier 3-D ICs offers more power savings than 2-tier 3-D ICs. Second, 3-tiers can be bonded in various mixed styles, and these various styles provide additional power reduction. Whenever possible, it is recommended to use F2F bonding over other bonding styles. However, more careful floorplanning, 3-D interconnect (TSVs and F2F) management, and block-folding considerations are required for the most power reduction when combined with advantageous bonding styles. Last, to demonstrate the maximum power reduction of 3-tier 3-D ICs, we developed CAD tools that seamlessly integrate into commercial 2-D tools for design and optimization. With aforementioned methods and design techniques combined, we achieved −36.0% total power saving against the 2-D counterpart in T2 Core, and −27.2% total power saving in full-chip T2 microprocessor.

References

[1] K. Ahmed and K. Schuegraf, "Transistor wars," *IEEE Spectr.*, vol. 48, no. 11, pp. 50–66, Nov. 2011.

[2] B. Black *et al.*, "Die stacking (3D) microarchitecture," in *Proc. 39th Annu. IEEE/ACM Int. Symp. Microarch.*, Orlando, FL, USA, 2006, pp. 469–479.

[3] M. Jung *et al.*, "How to reduce power in 3D IC designs: A case study with OpenSPARC T2 core," in *Proc. IEEE Cust. Integr. Circuits Conf. (CICC)*, San Jose, CA, USA, 2013, pp. 1–4.

[4] Y. J. Lee and S. K. Lim, "On GPU bus power reduction with 3D IC technologies," in *Proc. Design Autom. Test Eur. Conf. Exhibit. (DATE)*, Dresden, Germany, 2014, pp. 1–6.

[5] U. Kang *et al.*, "8GB 3D DDR3 DRAM using through-silicon-via technology," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, San Francisco, CA, USA, 2009, pp. 130–131, 131a.

[6] M. Jung, T. Song, Y. Wan, Y. Peng, and S. K. Lim, "On enhancing power benefits in 3D ICs: Block folding and bonding styles perspective," in *Proc. DAC*, San Francisco, CA, USA, 2014, pp. 1–6.

[7] S. H. Ok, K. R. Bae, S. K. Lim, and B. Moon, "Design and analysis of 3D IC-based low power stereo matching processors," in *Proc. IEEE Int. Symp. Low Power Electron. Design (ISLPED)*, Beijing, China, 2013, pp. 15–20.

[8] Oracle. *OpenSPARC T2*. [Online]. Available: http://www.oracle.com/technetwork/systems/opensparc/opensparc-t2-page-1446157.html

[9] Synopsys. *32/28nm Generic Library*. [Online]. Available: https://www.synopsys.com/COMMUNITY/UNIVERSITYPROGRAM/Pages/32-28nm-generic-library.aspx

[10] G. Katti, M. Stucchi, K. De 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.

[11] D. H. Kim, K. Athikulwongse, and S. K. Lim, "A study of through-silicon-via impact on the 3D stacked IC layout," in *IEEE/ACM Int. Conf. Comput.-Aided Design Dig. Tech. Papers*, San Jose, CA, USA, 2009, pp. 674–680.

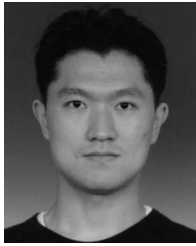
[12] U. G. Nawathe *et al.*, "An 8-core 64-thread 64b power-efficient SPARC SoC," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, San Francisco, CA, USA, 2007, pp. 108–590.

[13] C. Huyghebaert *et al.*, "Cu to Cu interconnect using 3D-TSV and wafer to wafer thermocompression bonding," in *Proc. Int. Interconnect Technol. Conf. (IITC)*, Burlingame, CA, USA, 2010, pp. 1–3.

[14] W.-C. Lo *et al.*, "An innovative chip-to-wafer and wafer-to-wafer stacking," in *Proc. 56th Electron. Compon. Technol. Conf.*, San Diego, CA, USA, 2006, p. 6.

[15] J. Shi *et al.*, "Direct chip powering and enhancement of proximity communication through anisotropic conductive adhesive chip-to-chip bonding," in *Proc. ECTC*, Las Vegas, NV, USA, Jun. 2010, pp. 363–368.

[16] J. Yannou *et al.*, *SET is Well Positioned and Prepared to Address the Challenges of the Fast Growing 3D System Integration Market*, Yole Développement, Villeurbanne, France, 2010.



Taigon Song (S'09–M'16) received the B.S. degree in electrical engineering from Yonsei University, Seoul, South Korea, in 2007, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2009, and the Ph.D. degree from the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2015.

He is currently a senior Research and Development engineer at Synopsys Inc. His research interests are in the modeling, design, extraction, and analysis in the advanced technologies including 3D integrated circuits. He has authored over 40 publications in international conferences and journals.

Dr. Song has served as a Reviewer for many journals such as the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, and the IEEE TRANSACTIONS ON COMPUTERS.

Shreepad Panth (S'11–M'15) received the B.S. degree from Anna University, Chennai, India, in 2009, and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, GA, USA, in 2011 and 2015, respectively.

He is currently a Design Engineer and a member of the technical staff with Altera Corporation, San Jose, CA, USA. He has authored over twenty publications. His current research interests include physical design for current and next generation 3-D integrated circuits.

Dr. Panth was a recipient of the Best Paper Award at ATS'12 and IITC'14 and the nominations for the Best Paper Award at ISPD'14 and DAC'14.

Yoo-Jin Chae (S'15) is currently pursuing the B.S. degree in electrical engineering with the Korea Advanced Institute of Science and Technology, Daejeon, South Korea.

Her current research interests include digital very large scale integration design, 3-D integrated circuits, and emerging technologies for smart automobiles.

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 at Los Angeles (UCLA), Los Angeles, CA, USA, in 1994, 1997, and 2000, respectively.

From 2000 to 2001, he was a Post-Doctoral Scholar with UCLA, and a Senior Engineer with Aplus Design Technologies, Inc., Los Angeles. In 2001, he joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, where he is currently an Associate Professor. He has authored the book entitled *Practical Problems in VLSI Physical Design Automation* (Springer, 2008). His current research interests include physical design automation for 3-D integrated circuits, 3-D system-in-packages, microarchitectural physical planning, and field-programmable analog arrays.

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