Decoupling Capacitor Topologies for TSV-Based 3-D ICs With Power Gating

Hailang Wang and Emre Salman, *Member, IEEE*

***Abstract*— In traditional decoupling capacitor topologies, power gating can significantly degrade the system-wide power integrity of a 3-D integrated circuit since the decoupling capacitance associated with the power-gated block/plane becomes ineffective for the neighboring, active planes. Two topologies are investigated to alleviate this issue by exploiting: 1) relatively low-resistance through silicon vias (TSVs) and 2) ability of TSVs to bypass plane-level power networks when delivering the power supply voltage. In the proposed topologies, decoupling capacitors placed within a plane can provide charge to neighboring planes even when the plane is power gated, achieving up to 50% and 87% reduction in, respectively, rms power supply and power gating (in-rush current) noise at the expense of a moderate increase in physical area and peak power consumption.**

***Index Terms*— Power dissipation, three-dimensional integrated circuits, through-silicon vias, very large scale integration.**

I. Introduction

THROUGH silicon via (TSV)-based 3-D integrated circuits (ICs) alleviate the adverse effects of global interconnects, thereby enabling higher performance at lower power consumption, while potentially lowering cost [1]–[3]. Furthermore, the integration of disparate circuits is facilitated, making 3-D technology a key enabler not only for More-Moore, but also for heterogeneous More-than-Moore integration [4], [5].

A critical design challenge for TSV-based 3-D ICs is the reliable delivery of the power supply voltage while ensuring sufficient system-wide power integrity [6]–[8]. This requirement is particularly challenging for low-voltage 3-D ICs where performance degradation is partially compensated by high parallelism, producing relatively high current. This low voltage and high current challenge has received considerable attention with particular emphasis on the effect of TSVs, power grid architectures, and decoupling capacitors [9]–[11]. The effect of power gating on the power integrity of 3-D ICs, however, has not received much attention.

Power gating is a common method to significantly reduce the subthreshold leakage current in nanoscale technologies [12]. For 3-D systems, power gating is critical due to higher and heterogeneous integration where the amount of nonswitching circuits can be significantly high. Thus, the 3-D ICs are expected to be heavily power gated to sufficiently reduce the leakage power, as shown in Fig. 1.

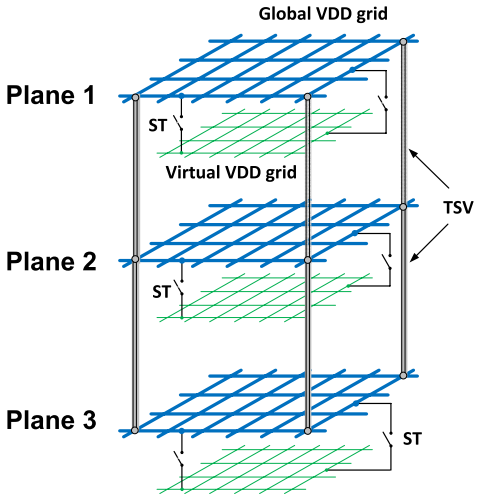


Fig. 1. Power distribution network for a three-plane 3-D IC with power gating, illustrating the global and virtual power grids, sleep transistors, and TSVs.

Todri *et al.* [13] investigated the effect of plane-level power gating on power and thermal integrity in 3-D ICs. However, the interdependence of decoupling capacitance with power gating has not been considered. Furthermore, the sleep transistors have been modeled with an equivalent resistance, preventing to accurately consider the transient, turn-ON characteristics of the sleep transistors while evaluating power integrity. According to [13], the decoupling capacitance placed within a plane is highly effective in reducing the power supply noise of the neighboring planes, as also observed in this paper. This behavior, however, holds unless the related blocks (or entire plane) are power gated, as described below.

In traditional topologies, the decoupling capacitance of a power-gated block (or plane) cannot provide charge to neighboring planes since these capacitors are typically connected to a virtual power network that is closer to the switching circuit. Placing the decoupling capacitors sufficiently close to the switching load is helpful in reducing the power supply noise [14], [15]. However, if the block or plane is power gated, those capacitors that provide charge to the block (or plane) are disconnected from the global power network, making these capacitors ineffective for the neighboring planes.

Manuscript received February 20, 2014; revised July 21, 2014 and September 29, 2014; accepted December 4, 2014. Date of publication May 13, 2015; date of current version November 20, 2015. This work was supported in part by the National Science Foundation CAREER under Grant CCF-1253715 and in part by the Office of the Vice President for Research through Stony Brook University, Stony Brook, NY, USA.

The authors are with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: hailang.wang@stonybrook.edu; emre.salman@stonybrook.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/TVLSI.2014.2386253

Since system-wide power integrity is a critical challenge in 3-D ICs, the effective use of intentional decoupling capacitance is crucial, even when power gating is adopted.

Xu *et al.* [16] proposed reroutable decoupling capacitors in 2-D ICs. The primary objective has been to relax the strong tradeoff between the power gating noise and the leakage power by connecting the decoupling capacitors to the global power grid when a block is power gated. Thus, these decoupling capacitors remain charged, reducing the power gating noise. This reduction in the power gating noise is utilized to decrease turn-on time, thereby saving additional leakage power. In [16], however, the efficacy of reroutable capacitors in reducing the power supply noise of other blocks (connected to a different virtual grid) has not been investigated. In this paper, reconfigurable capacitors are investigated in 3-D ICs to alleviate power supply noise of the neighboring, active planes. An always-on topology is also discussed to achieve the same objective. These topologies exploit low-resistance TSVs and ability of TSVs to bypass the plane-level power network when delivering the power supply voltage. Note that for via-first TSVs that connect the bottommost metal layer of an upper plane with the topmost metal layer of a lower plane, a sufficiently high number of metal vias is required to effectively bypass the plane-level power networks. These two characteristics increase the effective radius [14] of a decoupling capacitance in TSV-based 3-D ICs, making these topologies particularly applicable to vertical integration.

The rest of this paper is organized as follows. Related background on 3-D power distribution with power gating and decoupling capacitance is provided in Section II. Decoupling capacitor topologies for power-gated 3-D ICs are described in Section III. These topologies are evaluated in Section IV through a comprehensive case study. Finally, the conclusion is provided in Section V.

II. Power Gating in 3-D Power Distribution Networks

Background material related to power gating in 3-D ICs is described in Section II-A. The nonnegligible effect of decoupling capacitors in reducing power noise of the neighboring planes is discussed in Section II-B.

A. 3-D Power Distribution Network With Power Gating

Similar to 2-D planar technologies, sleep transistors with high-threshold voltage are utilized to achieve the power gating in 3-D ICs. Depending upon the type of TSVs (via-first/middle or via-last), distributed or lumped power gating topologies have been proposed to minimize the power supply noise while minimizing the leakage current [17], [18]. In the distributed topology (shown to be more appropriate for via-last TSVs), sleep transistors are distributed throughout the entire 3-D stack whereas in lumped topology (shown to be more appropriate for via-first/middle TSVs), all of the sleep transistors are located at the topmost plane [17]. Note that via-last TSVs pass through the metal layers and connect the topmost metal layer on each plane [7]. This characteristic facilitates the decoupling capacitor topologies discussed in this paper. However, via-first TSVs can also be used provided that a sufficiently high number of TSVs and metal vias is utilized to connect the two TSVs.

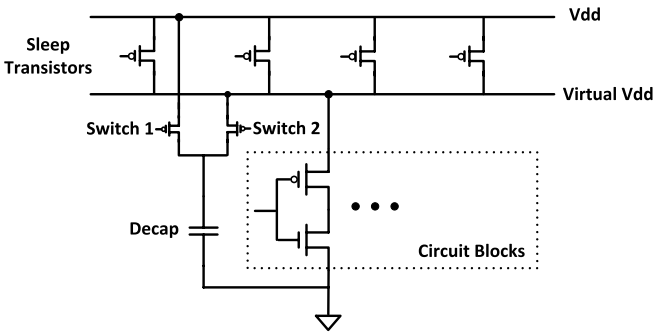


Fig. 2. Conceptual representation of the reconfigurable decoupling capacitor topology with power gating.

B. Effective Range of Decoupling Capacitance in 3-D ICs

In 3-D ICs, efficacy of a decoupling capacitor at one plane on the remaining planes is critical since it is highly challenging to satisfy the power supply noise as the number of planes increases. In existing work, it has been observed that the effective range of a decoupling capacitor exceeds single plane in 3-D ICs with low-resistance TSVs [13]. This phenomenon is important particularly when block- or plane-level power gating is performed due to significant decoupling capacitance that cannot provide charge to the remaining, active planes. Thus, when one or more number of planes is power gated, power supply noise in one of the remaining active planes may increase and violate the constraint despite a reduction in the overall switching current drawn from the power supply, as observed in this paper. It is, therefore, highly critical to ensure that the decoupling capacitors within the power-gated planes are not ineffective for the remaining planes.

Note that a similar issue exists in 2-D ICs with multiple power domains. However, due to longer global interconnects, it is relatively impractical to utilize the capacitance of a power-gated domain for the remaining, active domains. As demonstrated in [19], the required capacitance exponentially increases if the resistance between the capacitor and switching load exceeds a certain threshold.

III. Decoupling Capacitor Topologies for Power-Gated 3-D ICs

In this section, two decoupling capacitor placement topologies are discussed: 1) reconfigurable topology, as described in Section III-A and 2) always-on topology, as described in Section III-B. In both topologies, the decoupling capacitors on a power-gated plane can be utilized to suppress the power supply and the power gating noise within the neighboring active planes.

A. Reconfigurable Topology

In the reconfigurable topology, two switches are introduced (similar to [16]) to form a configurable decoupling capacitor, as conceptually shown in Fig. 2. If a certain plane is active, the

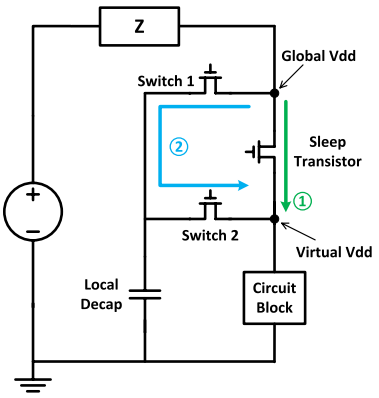


Fig. 3. Additional resistive path between the global and virtual power networks formed by the reconfigurable switches.

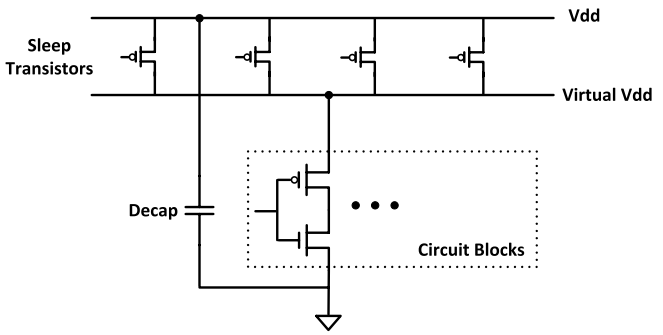


Fig. 4. Conceptual representation of the always-on decoupling capacitor topology with power gating.

decoupling capacitors on that plane are connected to the virtual V_{DD} grid through Switch 2, thereby reducing the power supply noise on that plane. Alternatively, if the plane is power gated (sleep transistors are turned OFF), the decoupling capacitors are connected to the global V_{DD} grid, bypassing the sleep transistors. Thus, even if the plane is power gated, the decoupling capacitors are effective for the remaining planes. The overhead of this topology includes the reconfigurable switches, metal resources required to route the related control signals, and a possible increase in overall power consumption depending upon how the capacitors are implemented, as further discussed and quantified in Section IV.

The design process (sizing and choosing an appropriate threshold voltage) of these switches exhibits similar and well-known tradeoffs as the design process of the sleep transistors [20], [21]. Similar to sleep transistors, high-V_{th} switches are used to minimize the voltage at the virtual V_{DD} grid when the plane is power gated (Switch 1 is ON and Switch 2 is OFF). Note that the two reconfigurable switches form an additional path from the global V_{DD} grid to the virtual V_{DD} grid, as shown in Fig. 3. Thus, the effective resistance of the sleep transistors and the effective resistance of the reconfigurable switches are in parallel, partially reducing the OFF-resistance between the global and the virtual V_{DD} grids. High-V_{th} switches are, therefore, required to maintain significant savings in the leakage current when the plane is power gated.

B. Always-On Topology

The reconfigurable placement methodology described above provides flexibility to exploit a decoupling capacitor located in a power-gated plane for the remaining planes. This flexibility is achieved at the expense of additional reconfigurable switches. To mitigate the overhead of reconfigurable topology, an always-on topology is considered for planes with low switching activity (such as a sensing plane of a heterogeneous 3-D IC that is periodically activated). In this topology, the decoupling capacitors are always connected to the global V_{DD} grid, thereby bypassing the sleep transistors, as conceptually shown in Fig. 4. Thus, the decoupling capacitance within a power-gated plane is available to suppress the power supply and the power gating noise of the neighboring active planes. The limitation of this topology is a possible increase in the power supply noise of the plane where the decoupling capacitors are located due to a greater impedance between the capacitor and the switching circuit [14]. This tradeoff and the additional decoupling capacitance required to mitigate this limitation are characterized in Section IV.

IV. Case Study

A comprehensive case study is developed to investigate the benefits and tradeoffs of the decoupling capacitor topologies discussed in this paper. The analysis setup is described in Section IV-A. Design issues, such as simultaneously sizing decoupling capacitors and switches, are investigated in Section IV-B. Simulation results are presented in Section IV-C, where the proposed topologies are compared with the traditional topology in terms of power supply noise, power gating noise, physical area, turn-on time, and overall power consumption. The effect of number of planes is also investigated.

A. Simulation Setup

1) 3-D Power Grid With Power Gating: A power distribution network for a three-plane 3-D IC with via-last TSVs is developed, as conceptually shown in Fig. 1. A 45-nm CMOS technology with 10 available metal layers in each plane is adopted [22]. A portion of the power network with an area of 1 mm $\times$ 1 mm is analyzed. Each plane consists of a global power network, virtual power network, distributed pMOS sleep transistors, distributed decoupling capacitance (implemented as MOS capacitors), and distributed switching load circuit consisting of inverter gates, as shown in Fig. 5. Note that the top plane also consists of C4 bumps to connect the on-chip grid with the flip-chip substrate.

The top two metal layers (9 and 10) on each plane are dedicated to global power distribution network with an interdigitated grid of 11×11 metal lines [23]. Metal layers 8 and 7 are used as the virtual power grid that is connected to the global grid through sleep transistors. Virtual V_{DD} network is also represented by an interdigitated grid of 21×21. Power gating is achieved using a distributed method where the sleep transistors that control a plane are placed within that plane [17].

Primary physical characteristics of the 3-D power grid are listed in Table I. The pitch and width of the metal lines are

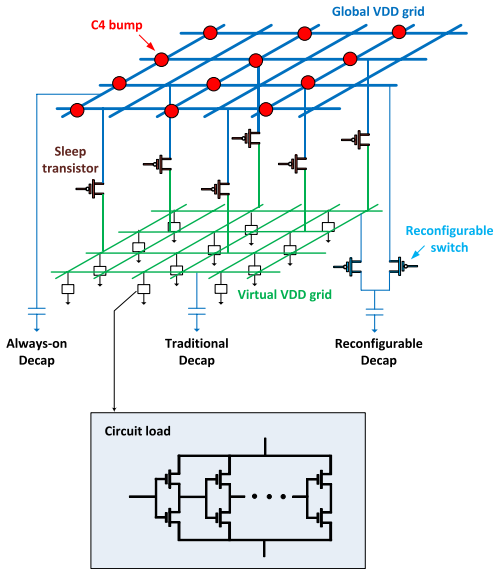


Fig. 5. Plane-level power network illustrating distributed sleep transistors, decoupling capacitors (traditional and proposed topologies), switching load circuits (gates with active devices), and the $C4$ bumps (for the top plane only).

TABLE I
PRIMARY PHYSICAL CHARACTERISTICS OF THE GLOBAL AND VIRTUAL POWER GRIDS [22]

Parameters		Values
Metal 10 & 9	Pitch	45 μm
	Width	40 μm
	Resistivity (ohm/sq)	0.03
Metal 8 & 7	Pitch	23.5 μm
	Width	20 μm
	Resistivity (ohm/sq)	0.075

TABLE II
PARASITIC IMPEDANCES OF THE UNIT INTERCONNECT SEGMENT WITHIN THE GLOBAL AND VIRTUAL POWER GRIDS

	R_{unit} (mΩ)	C_{unit} (fF)	L_{unit} (pH)
Global grid	67.50	27.00	37.70
Virtual grid	176.25	7.05	20.13

determined based upon the technology design rules [22] while also considering routing constraints. For each interconnect segment, an RLC π circuit is used to model the parasitic impedances of the power grid. The unit parasitic capacitance (extracted from FastCap [24]), inductance (extracted from FastHenry [25]), and resistance (based on sheet resistance [22]) values are listed in Table II.

The physical characteristics of the via-last TSVs (with copper as the filling material), $C4$ bumps, and package impedances are listed in Table III. A flip-chip package is assumed and modeled with a lumped resistance of 1 mΩ and inductance of 120 pH [26]. $C4$ bumps are regularly placed with a pitch of 200 μm over the 1-mm $\times$ 1-mm area [8]. Each $C4$ bump has a resistance of 5 mΩ and an inductance of 200 pH [8]. Clustered via-last TSVs are distributed throughout the area as a 5 $\times$ 5 array and connect the

TABLE III
PACKAGE, TSV, AND $C4$ BUMP PARASITIC IMPEDANCES AND PHYSICAL CHARACTERISTICS [8], [26]

Parameters	Values
Nominal power supply voltage V_{DD}	1.0 V
Lumped package resistance R_{package}	1 mΩ
Lumped package inductance L_{package}	120 pH
Single C4 bump resistance R_{C4}	5 mΩ
Single C4 bump inductance L_{C4}	200 pH
Via-last TSV diameter W	10 μm
Via-last TSV height H	60 μm
Via-last TSV dielectric thickness t_{ox}	0.2 μm
Resistivity of TSV filling material (copper) ρ_f	16.8 nΩ·m
Single via-last TSV resistance R_{tsv}	20mΩ
Single via-last TSV capacitance C_{tsv}	283 fF
Single via-last TSV inductance L_{tsv}	35 pH

global power grid on each plane. Each TSV cluster consists of four TSVs. Thus, 100 power TSVs are used to connect the two planes. The overall number of power TSVs in the three-plane stack is 200.

2) Switching Load Circuit: As opposed to using piecewise linear (PWL) current sources to model the switching load circuit (typical practice in existing work [12], [27]), gates with active devices are used since power gating is considered. Note that power gating noise cannot be accurately analyzed when PWL sources are used since the transient turn-ON characteristics of the active devices play an important role in power gating noise. Furthermore, savings in the subthreshold leakage current cannot be estimated with PWL sources. Finally, an active load enables to consider the negative feedback between the supply noise and the load current (larger supply noise reduces load current, which in turn reduces supply noise), enhancing the accuracy of the analysis.

Similar to [28], inverter pairs with varying number and size are used to model the switching load circuit. The overall area is divided into 30 segments and a switching circuit is connected to each segments to consider the spatial heterogeneity of the current loads. The spatial load current distribution and power densities are based on [29]. As an example, the current distribution of the top plane is shown in Fig. 6, where the peak current for each block is indicated. For the middle and bottom planes, the same switching circuits are used, but these circuits are placed at different locations throughout the power network. Note that according to the current profiles of the inverter pairs, the peak power density reaches 40 W/cm^2, which is comparable with the power density in modern processors [30]. Decoupling capacitors are also conceptually shown in Fig. 6. The specific location of the decoupling capacitors is determined based on the spatial power supply noise distribution, as described in Section IV-C.

B. Reconfigurable Switch and Decoupling Capacitor Sizing

The size of the reconfigurable switches should be sufficiently large to minimize the shield effect of these switches on the decoupling capacitors [3], [14]. Analyses demonstrate that the multiple pairs of decoupling capacitor and switch size satisfy the power supply noise constraint. Since both the decoupling capacitors and switches consume area, it is important to

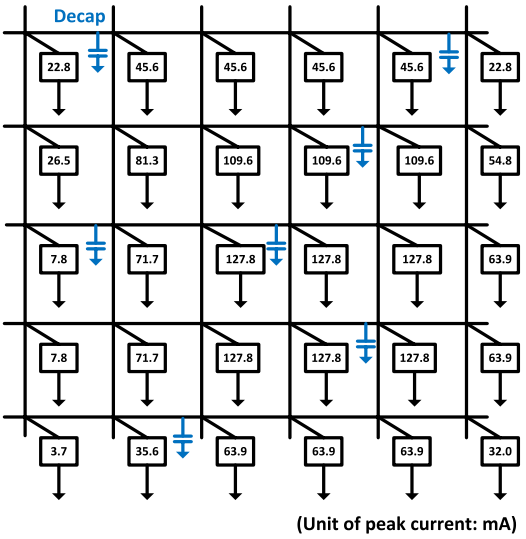


Fig. 6. Current distribution within the top plane based on [29]. Numbers: peak current drawn by the digital gates at each node. The peak power density reaches 40 W/cm^2, which is comparable with the power density in modern processors [30].

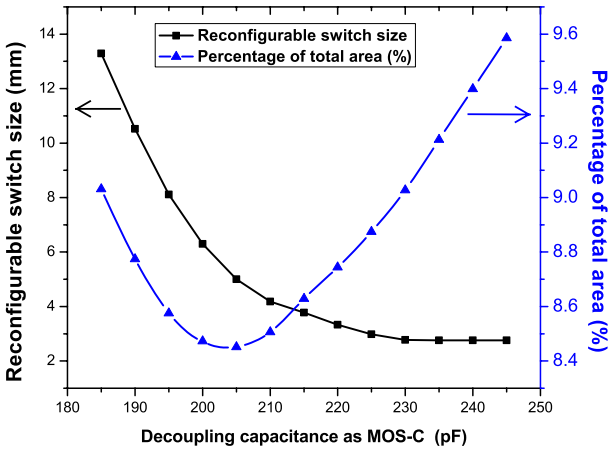


Fig. 7. Area consumption (as percent of the overall area) of different pairs of decoupling capacitance and reconfigurable switch size. Note that each pair satisfies the power supply noise constraint of 50 mV (5% of the supply voltage). Area overhead is minimized at a specific pair. Decoupling capacitors are implemented as MOS-C.

choose a pair that minimizes the physical area overhead. This characteristic is shown in Fig. 7. Each pair of switch size and decoupling capacitance (on the curve with square markers) satisfies the 5% power supply noise constraint (50 mV). The area overhead (determined as a percentage of the overall area) is shown by the curve with triangle markers. The decoupling capacitors are implemented as MOS capacitors in the 45-nm technology with an oxide thickness of 1 nm [22]. As shown in this figure, a small decoupling capacitor requires a very large switch size to satisfy the noise constraint. A large switch size not only increases the area overhead, but also increases the voltage at the virtual V_{DD} grid when the block/plane is power gated, thereby increasing the leakage current. Alternatively, if decoupling capacitance exceeds a certain threshold, the switch size cannot be reduced further, thereby increasing the overall

TABLE IV
SIZES OF THE DECOUPLING CAPACITORS AND RECONFIGURABLE SWITCHES

	ST	Decap	Switch
Traditional	36.3 mm	2.90 nF	N/A
Reconfigurable	36.3 mm	3.29 nF	160 mm
Always-on	36.3 mm	5.21 nF	N/A

area overhead. For this case study, the minimum area overhead (8.45%) occurs when the equivalent decoupling capacitor (from MOS-C) is approximately 205 pF and switch size is 5 mm. Note that 205 pF and 5 mm represent, respectively, a single decoupling capacitor and a single switch in the power network. The overall decoupling capacitance (per plane) is equal to approximately 3.3 nF, whereas the overall switch size (per plane) is equal to 160 mm since there are two switches per capacitor and the total number of capacitors per plane is 16. This amount of decoupling capacitance and switch size is used for the reconfigurable topology in the remaining portions of this paper.

C. Simulation Results

The efficacy of the decoupling capacitor topologies discussed in this paper is demonstrated by comparing these methods with the traditional topology. Design criteria, such as area overhead, power supply noise, power gating noise, and turn-on time, are analyzed. The effect of number of planes on the capacitor topologies is investigated. Power overhead of each topology is also quantified to demonstrate that the proposed topologies do not undermine the leakage savings achieved by power gating. All of the simulations have been performed using SPICE accurate SPECTRE simulator [31].

Several different scenarios are considered.

- *Scenario 1:* All of the three planes are active, representing the greatest workload.
- *Scenario 2:* The top and bottom planes are active, while the middle plane is power gated.
- *Scenario 3:* Only the bottom plane is active, while the middle and top planes are power gated.
- *Scenario 4:* The middle and bottom planes are active, while the top plane is power gated.
- *Scenario 5:* Only the middle plane is active, while the top and bottom planes are power gated.

1) Area Overhead: The size of the distributed decoupling capacitors (implemented as MOS-C) is determined based on Scenario 1 where all of the planes are active. For each topology, the decoupling capacitors are sized to ensure that the worst case power supply noise is within 5% of the V_{DD} (50 mV) throughout the entire power network. This constraint ensures that no additional performance penalty is introduced with the reconfigurable and always-on topologies. For the reconfigurable topology, the size of the switches and decoupling capacitors is determined to minimize the physical area overhead, while satisfying the power supply noise constraint, as described in Section IV-B. These sizes are listed in Table IV. Note that the size of the decoupling capacitors,

TABLE V
COMPARISON OF THE PHYSICAL AREA OVERHEAD OF THE TRADITIONAL, RECONFIGURABLE, AND ALWAYS-ON TOPOLOGIES

Area (μm^2)		ST	Decap	Switch	Area per plane
Traditional	MOS	1815	65413	N/A	67228 (6.72%)
	MIM		233870		235685 (23.56%)
Reconfigurable	MOS	1815	76733	8000	86548 (8.65%)
	MIM		265322		275137 (27.51%)
Always-on	MOS	1815	118158	N/A	119973 (11.99%)
	MIM		420161		421976 (42.19%)

sleep transistors, and switches (for the reconfigurable topology) listed in this table refers to per plane.

MOS and metal–insulator–metal (MIM) capacitors are considered to estimate the area overhead of the decoupling capacitors to demonstrate the tradeoff between area and leakage overhead. For MIM capacitors, the area overhead is analytically estimated by assuming a capacitance density of 12.4 fF/μm^2 based on [32]. The itemized area overhead of the traditional, reconfigurable, and always-on topologies are listed in Table V for both MOS and MIM capacitors. If MOS capacitor is used (as in the simulations), the area overhead (due to capacitors and sleep transistors) is 6.70% of the overall area for the traditional topology where the decoupling capacitors are connected to the virtual $V_{\rm DD}$ grid. The area overhead increases to 8.65% in the reconfigurable topology since the size of the decoupling capacitors should be moderately increased to compensate for the shield effect of the reconfigurable switches (which also contributes to the area overhead). Finally, in the always-on topology, the area overhead increases to 11.99% due to an increase in the decoupling capacitance, as listed in Table IV. For MIM capacitor, the area overhead (analytically determined) is significantly greater than MOS capacitor (23.56%, 27.51%, and 42.19%, respectively, for traditional, reconfigurable, and always-on topologies). The leakage current of MIM capacitor, however, is significantly less than MOS capacitor, as discussed in Section IV-C5. In addition, note that the MIM capacitors consume area within the metal layers rather than consuming transistor area.

2) Power Integrity: Power supply noise and power gating noise are analyzed for each scenario. The reconfigurable and always-on decoupling capacitor topologies achieve significant reduction in both peak and rms power supply noise, as listed in Table VI. Note that, in the simulations, all of the decoupling capacitors are implemented as MOS-C.

In Scenario 1, where all of the planes are switching, the peak power supply noise is equal to 50 mV for each topology since the decoupling capacitor, sleep transistor, and switch sizes are determined based on this scenario. Note that power supply noise is observed in the bottom plane except Scenario 5 where bottom plane is power gated. In this case, noise is observed in the middle plane. For Scenarios 3 and 5 (where two planes are power gated), both the reconfigurable and always-on topologies reduce the peak power supply noise by more than 20%. In these scenarios, the reconfigurable topology reduces the rms noise by 46%, whereas the always-on topology achieves 50% reduction in rms noise. For Scenarios 2 and 4 (where only one plane is power gated), the reduction in peak noise is approximately 9% and 10% for, respectively, reconfigurable and always-on topologies. For the same scenarios, proposed topologies achieve, respectively, at least 25% and 27% reduction in rms noise.

It is important to note that in the traditional topology, the peak noise in Scenarios 3 and 5 exceeds 50 mV despite a reduction in the overall switching current due to power gating. This characteristic is due to less decoupling in the power network since the decoupling capacitors in the power-gated planes cannot behave as charge reservoirs for the remaining, active planes.

Transient behavior of voltage noise at a specific node within the bottom plane is shown in Fig. 8 for each topology for Scenario 3, demonstrating the reduction in peak and rms noise. Similarly, the spatial distribution of peak power supply noise is shown in Fig. 9 for Scenario 3 where the first two planes are power gated and the bottom plane is active. Reduction in peak noise throughout the power network is illustrated for both reconfigurable and always-on topologies.

To investigate power gating noise, one of the power-gated plane transitions from sleep to active state in each scenario (except Scenario 1) and the voltage fluctuation due to in-rush current during the wake-up process is analyzed. Note that a gradual wake-up strategy is adopted where switching circuits on each plane are divided into five segments and each segment sequentially wakes up with a time interval of 100 ps based on [33]. Peak power gating noise is observed in the bottom plane except Scenario 5 where bottom plane has a transition. In this case, noise is observed in the middle plane. Results are listed in Table VII. Both reconfigurable and always-on topologies achieve approximately 80% reduction in peak and rms power gating noise. This considerable reduction in power gating noise is due to a significant amount of in-rush current in traditional topology that flows not only for the activated circuit, but also to charge the associated decoupling capacitors, as also observed in [16]. Thus, a greater in-rush current produces significantly high-power gating noise (particularly due to parasitic inductance). Alternatively, in both the reconfigurable and always-on topologies, the decoupling capacitors are connected to the global $V_{\rm DD}$ grid when the plane is power gated. Thus, even if the plane is power gated, these capacitors remain charged (significantly reducing in-rush current) and can behave as a charge reservoir once the plane transitions to active state. The transient behavior of the power gating noise is shown in Fig. 10 for Scenario 3. The middle plane transitions from sleep to active state at 1 ns and the in-rush current noise is observed on the bottom plane.

3) Turn-On Time: Turn-on time for each topology is investigated. A gradual wake-up strategy described in Section IV-C2 is adopted. Scenario 3 is considered where the middle plane is turned on while the top plane is power gated and bottom plane is active. Power supply voltage variation on the virtual grid of one of the circuit blocks is shown in Fig. 11 during the wake-up process. The wake-up time is determined when the voltage reaches 90% of the nominal $V_{\rm DD}$. For the traditional topology, the wake-up time of the circuit block is 0.80 ns. Alternatively, with the reconfigurable and always-on topologies, the wake-up time is reduced,

TABLE VI

PEAK AND RMS POWER SUPPLY NOISE OBTAINED FROM EACH SCENARIO AND NOISE REDUCTION ACHIEVED BY THE PROPOSED TOPOLOGIES (ALL OF THE DECOUPLING CAPACITORS ARE IMPLEMENTED AS MOS CAPACITORS)

	Power gating status			**Power supply noise (mV)**									
	Top	Middle	Bottom	Traditional		Reconfigurable				Always-on			
				Peak	RMS	Peak	Redtn.	RMS	Redtn.	Peak	Redtn.	RMS	Redtn.
Scenario 1	on	on	on	50	30.34	50	N/A	26.26	13.5%	50	N/A	26.45	12.8%
Scenario 2	on	off	on	48.16	23.40	43.48	9.7%	17.2	26.5%	43.40	9.9%	15.83	32.4%
Scenario 3	off	off	on	52.22	16.93	39.64	24.1%	9.19	45.7%	38.07	27.1%	8.57	49.4%
Scenario 4	off	on	on	48.55	23.03	44.50	8.3%	17.15	25.5%	42.89	11.7%	16.65	27.7%
Scenario 5	off	on	off	52.51	17.03	38.78	26.1%	9.21	45.9%	37.55	28.5%	8.46	50.3%

TABLE VII

PEAK AND RMS POWER GATING NOISE OBTAINED FROM EACH SCENARIO AND NOISE REDUCTION ACHIEVED BY THE PROPOSED TOPOLOGIES (ALL OF THE DECOUPLING CAPACITORS ARE IMPLEMENTED AS MOS CAPACITORS)

	Power gating status			**Power gating noise (mV)**									
	Top	Middle	Bottom	Traditional		Reconfigurable				Always-on			
				Peak	RMS	Peak	Redtn.	RMS	Redtn.	Peak	Redtn.	RMS	Redtn.
Scenario 1	on	on	on	N/A		N/A				N/A			
Scenario 2	on	off→on	on	93.46	74.55	19.45	79.2%	12.94	82.6%	17.98	80.8%	11.47	84.6%
Scenario 3	off	off→on	on	112.28	83.60	19.61	82.5%	12.80	84.7%	17.31	84.6%	11.49	86.3%
Scenario 4	off→on	on	on	93.65	74.34	19.40	79.3%	12.93	82.6%	17.24	81.6%	11.41	84.7%
Scenario 5	off	on	off→on	112.0	83.65	19.60	82.5%	12.82	84.7%	17.39	84.5%	11.49	86.3%

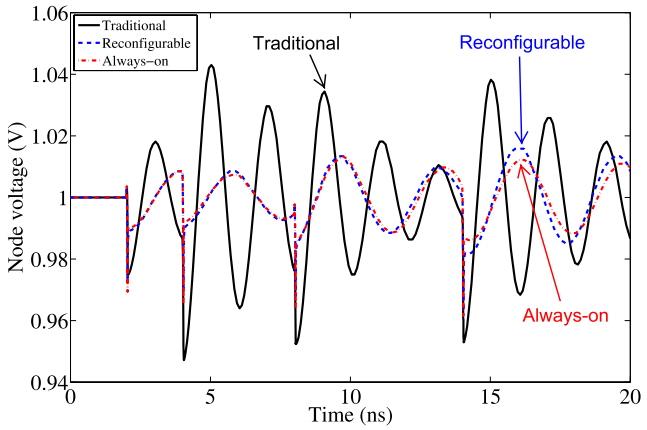


Fig. 8. Transient behavior of the ON-voltage at a specific node within virtual grid of the bottom plane for each topology for Scenario 3 (first two planes are power gated and the bottom plane is active).

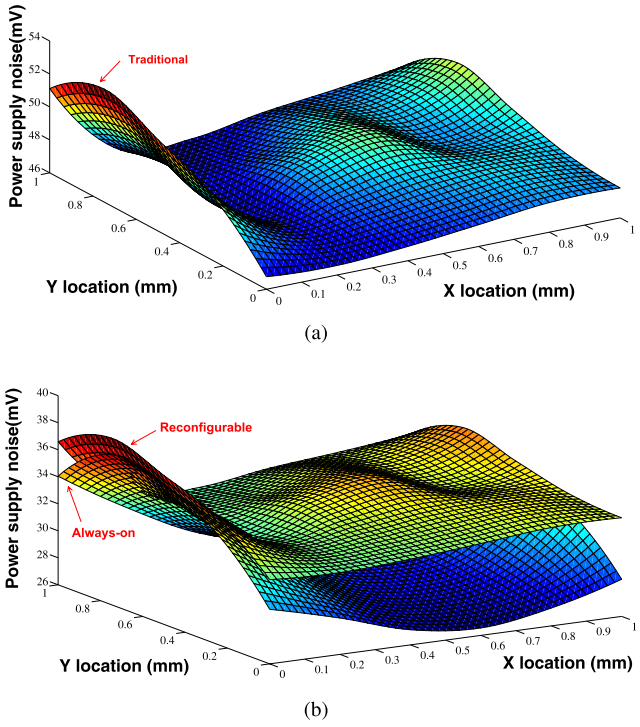


Fig. 9. Spatial distribution of the peak power supply noise on the bottom plane for Scenario 3 (first two planes are power gated and the bottom plane is active). (a) Traditional topology. (b) Reconfigurable and always-on topologies.

respectively, to 0.43 and 0.33 ns. As mentioned before, in these topologies, the decoupling capacitors within a power-gated plane remain charged, thereby reducing the turn-on time. As mentioned in [16], a shorter wake-up time enables larger leakage power savings. The overall time required to turn on the entire plane is 1.17, 0.84, and 0.74 ns for, respectively, traditional, reconfigurable, and always-on topologies.

4) Effect of Number of Planes: The effect of the number of planes on the efficacy of the decoupling capacitor topologies is discussed. The simulation setup described in Section IV-A is extended to increase the number of planes with the same physical characteristics. Switching circuit loads within the top plane (closest to the package) are maintained active, whereas the additional planes beneath the first plane are power gated. Peak power supply noise on the top plane is shown in Fig. 12 as the number of power-gated planes increases. As shown in this figure, for the reconfigurable and always-on topologies, the decoupling capacitors within the second and third planes are highly effective in reducing the supply noise of the top plane. If, however, the number of planes further increases, the supply noise starts to slightly increase. The decoupling capacitors within the fourth and farther planes are not effective for the top plane due to greater impedance.

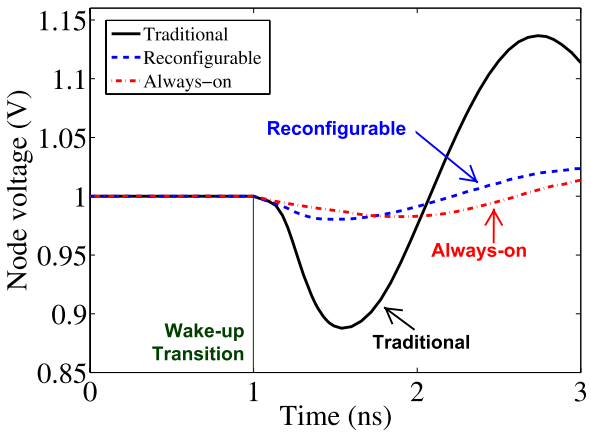


Fig. 10. Transient behavior of power gating noise at a specific node within the bottom plane for each topology for Scenario 3 (top plane is power gated and middle plane transitions from OFF to ON state at 1 ns).

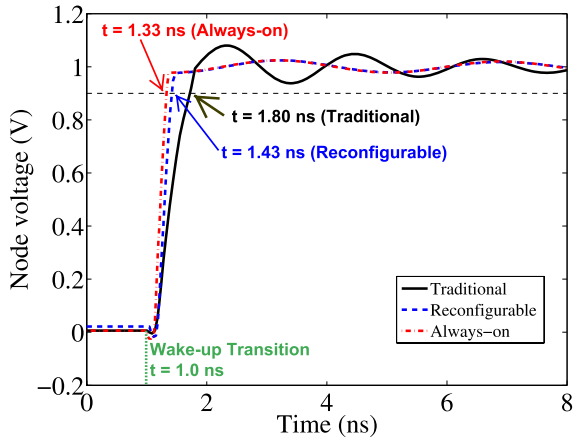


Fig. 11. Power supply voltage variation on the virtual grid of one of the circuit blocks (located within the middle plane) during the wake-up process.

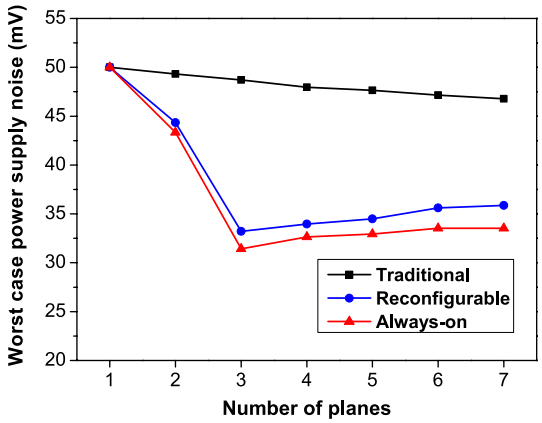


Fig. 12. Peak power supply noise on the top plane as the number of power-gated planes increases.

Since these capacitors are implemented as MOS-C and are connected to the global grid, the overall current drawn from the power supply slightly increases with increasing number of planes (due to nonnegligible MOS-C leakage current). Thus, the power noise of the top plane slightly increases if the number of planes increases beyond three. For the traditional topology, power noise slightly decreases with increasing number of power-gated planes due to the parasitic capacitance of the power grid and TSVs within the power-gated planes. Note that the negative impact of high MOS-C leakage current can be alleviated if capacitors are implemented with the MIM technique. This reduction in leakage current is achieved at the expense of a significant increase in physical area, as analytically determined in Section IV-C1.

TABLE VIII
OVERALL AVERAGE POWER CONSUMPTION (WHEN ALL DECOUPLING CAPACITORS ARE IMPLEMENTED AS MOS CAPACITORS)

	Traditional	Reconfigurable		Always-on	
	Power (mW)	Power (mW)	Overhead (%)	Power (mW)	Overhead (%)
Scenario 1	26.56	27.01	1.69%	27.14	2.18%
Scenario 2	17.62	18.15	2.99%	18.52	5.11%
Scenario 3	8.86	9.43	6.46%	9.80	10.64%
Scenario 4	17.53	18.12	3.37%	18.35	4.68%
Scenario 5	8.98	9.60	6.90%	9.86	9.80%

5) Power Overhead: It is important to quantify the power overhead of the decoupling capacitor topologies to ensure that the proposed topologies do not undermine the reduction in leakage current. Each topology is simulated for each scenario and the overall average power consumption is determined. All of the decoupling capacitors are implemented as MOS-C. Results are listed in Table VIII. The smallest overhead occurs in Scenario 1 where all of the planes are active. This overhead is due to increased capacitance and switches (for the reconfigurable topology only). Power overhead increases in Scenarios 3 and 5 where two planes are power gated. In particular, for the reconfigurable topology, power overhead is approximately 6∼7% due to the leakage current of MOS capacitors that are connected to the global grid. For the always-on topology, the power overhead increases to approximately 10% since more capacitance is required in this topology. However, note that power overhead in Scenario 1 is more important since peak power is consumed in this scenario. Also note that if MIM capacitors are utilized, the power overhead can be significantly reduced. For example, assuming a leakage current of 1 nA/cm^2 for an MIM capacitor based on [32], the power consumption increases by only 1.25% and 1.38%, respectively, for the reconfigurable and always-on topologies in Scenario 3. This small power overhead is achieved at the expense of a significant increase in area, as listed in Table V.

V. CONCLUSION

The 3-D ICs are expected to be heavily power gated due to higher integration and substantial subthreshold leakage current in modern CMOS processes. In 3-D ICs with power gating, system-wide power integrity can be compromised if traditional decoupling capacitor placement topology is utilized since these capacitors are typically placed sufficiently close to the switching circuit, i.e., connected to the virtual power network. When a block within a plane or the entire plane is power gated, related decoupling capacitors cannot provide charge to the neighboring, active planes, degrading both power supply noise and power gating noise. Two characteristics of

TSVs are exploited to alleviate this issue: 1) low resistivity and 2) ability to bypass plane-level power network when delivering the power supply voltage to farther planes. Utilizing these two characteristics, two decoupling capacitor topologies are investigated with significant reductions in power supply and gating noise at the expense of a moderate increase in area and peak power consumption. The turn-on time of the proposed topologies and the effect of number of planes on the efficacy of these topologies are also investigated.

REFERENCES

[1] V. F. Pavlidis and E. G. Friedman, *Three-Dimensional Integrated Circuit Design*. San Mateo, CA, USA: Morgan Kaufmann, 2010.

[2] A. W. Topol *et al.*, "Three-dimensional integrated circuits," *IBM J. Res. Develop.*, vol. 50, nos. 4–5, pp. 491–506, Jul./Sep. 2006.

[3] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*. New York, NY, USA: McGraw-Hill, 2012.

[4] (2011). *International Technology Roadmap for Semiconductors (ITRS)*. [Online]. Available: http://www.itrs.net/

[5] E. Salman, M. H. Asgari, and M. Stanacevic, "Signal integrity analysis of a 2-D and 3-D integrated potentiostat for neurotransmitter sensing," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Nov. 2011, pp. 17–20.

[6] G. Huang, M. Bakir, A. Naeemi, H. Chen, and J. D. Meindl, "Power delivery for 3D chip stacks: Physical modeling and design implication," in *Proc. IEEE Conf. Elect. Perform. Electron. Packag.*, Oct. 2007, pp. 205–208.

[7] S. M. Satheesh and E. Salman, "Power distribution in TSV-based 3-D processor-memory stacks," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 4, pp. 692–703, Dec. 2012.

[8] M. B. Healy and S.-K. Lim, "Distributed TSV topology for 3-D power-supply networks," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 11, pp. 2066–2079, Nov. 2012.

[9] P. Zhou, K. Sridharan, and S. S. Sapatnekar, "Optimizing decoupling capacitors in 3D circuits for power grid integrity," *IEEE Des. Test Comput.*, vol. 26, no. 5, pp. 15–25, Sep. 2009.

[10] Z. Xu *et al.*, "Decoupling capacitor modeling and characterization for power supply noise in 3D systems," in *Proc. 23rd Annu. SEMI Adv. Semicond. Manuf. Conf.*, May 2012, pp. 414–419.

[11] K. Kim, J. S. Pak, H. Lee, and J. Kim, "Effects of on-chip decoupling capacitors and silicon substrate on power distribution networks in TSV-based 3D-ICs," in *Proc. IEEE 62nd Electron. Compon. Technol. Conf.*, May/Jun. 2012, pp. 690–697.

[12] H. Jiang, M. Marek-Sadowska, and S. R. Nassif, "Benefits and costs of power-gating technique," in *Proc. IEEE Int. Conf. Comput. Design*, Oct. 2005, pp. 559–566.

[13] A. Todri *et al.*, "A study of tapered 3-D TSVs for power and thermal integrity," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 21, no. 2, pp. 306–319, Feb. 2013.

[14] M. Popovich, M. Sotman, A. Kolodny, and E. G. Friedman, "Effective radii of on-chip decoupling capacitors," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 7, pp. 894–907, Jul. 2008.

[15] E. Salman, E. G. Friedman, R. M. Secareanu, and O. L. Hartin, "Worst case power/ground noise estimation using an equivalent transition time for resonance," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, no. 5, pp. 997–1004, May 2009.

[16] T. Xu, P. Li, and B. Yan, "Decoupling for power gating: Sources of power noise and design strategies," in *Proc. ACM/IEEE Design Autom. Conf.*, Jun. 2011, pp. 1002–1007.

[17] H. Wang and E. Salman, "Power gating topologies in TSV based 3D integrated circuits," in *Proc. ACM/IEEE Great Lakes Symp. VLSI*, May 2013, pp. 327–328.

[18] H. Wang and E. Salman, "Resource allocation methodology for through silicon vias and sleep transistors in 3D ICs," in *Proc. IEEE Int. Symp. Quality Electron. Design*, Mar. 2015.

[19] E. Wong, J. Minz, and S. K. Lim, "Decoupling-capacitor planning and sizing for noise and leakage reduction," in *Proc. IEEE Int. Conf. Comput.-Aided Design*, Nov. 2006, pp. 395–400.

[20] W. Wang, M. Anis, and S. Areibi, "Fast techniques for standby leakage reduction in MTCMOS circuits," in *Proc. IEEE Int. Syst.-on-Chip Conf.*, 2004, pp. 21–24.

[21] D.-S. Chiou, S.-H. Chen, and S.-C. Chang, "Sleep transistor sizing for leakage power minimization considering charge balancing," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 17, no. 9, pp. 1330–1334, Sep. 2009.

[22] *FreePDK45*. [Online]. Available: http://www.eda.ncsu.edu/wiki/FreePDK45:Contents, accessed Feb. 2014.

[23] R. Jakushokas and E. G. Friedman, "Multi-layer interdigitated power distribution networks," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 5, pp. 774–786, May 2011.

[24] K. Nabors and J. White, "FastCap: A multipole accelerated 3-D capacitance extraction program," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 10, no. 11, pp. 1447–1459, Nov. 1991.

[25] *FastHenry*. [Online]. Available: http://www.fastfieldsolvers.com/, accessed Feb. 2014.

[26] M. S. Gupta, J. L. Oatley, R. Joseph, G.-Y. Wei, and D. M. Brooks, "Understanding voltage variations in chip multiprocessors using a distributed power-delivery network," in *Proc. Design, Autom., Test Eur. Conf. Exhibit.*, Apr. 2007, pp. 1–6.

[27] H. H. Chen and D. D. Ling, "Power supply noise analysis methodology for deep-submicron VLSI chip design," in *Proc. 34th ACM/IEEE Design Autom. Conf.*, Jun. 1997, pp. 638–643.

[28] X. Zhang, T. Tong, S. Kanev, S. K. Lee, G.-Y. Wei, and D. Brooks, "Characterizing and evaluating voltage noise in multi-core near-threshold processors," in *Proc. Int. Symp. Low Power Electron. Design*, Sep. 2013, pp. 82–87.

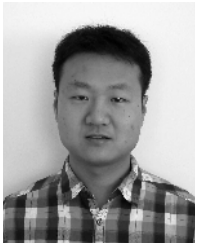
[29] Q. K. Zhu, *Power Distribution Network Design for VLSI*. New York, NY, USA: Wiley, 2004.

[30] H. Wei, T. F. Wu, D. Sekar, B. Cronquist, R. F. Pease, and S. Mitra, "Cooling three-dimensional integrated circuits using power delivery networks," in *Proc. IEEE Int. Electron Devices Meeting*, Dec. 2012, pp. 14.2.1–14.2.4.

[31] *Cadence Spectre*. [Online]. Available: http://www.cadence.com/products/cic/spectrecircuit

[32] S.-U. Park *et al.*, "Analysis of reliability characteristics of high capacitance density MIM capacitors with SiO_2–HfO_2–SiO_2 dielectrics," *Microelectron. Eng.*, vol. 88, no. 12, pp. 3389–3392, 2011.

[33] K. Kawasaki, T. Shiota, K. Nakayama, and A. Inoue, "A sub-μs wake-up time power gating technique with bypass power line for rush current support," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1178–1183, Apr. 2009.



Hailang Wang received the B.S. degree in microelectronics from Xidian University, Xi'an, China, in 2010. He is currently pursuing the Ph.D. degree in electrical engineering with Stony Brook University, Stony Brook, NY, USA.

His current research interests include low-power integrated circuit design methodology and power distribution in 3-D integrated circuits.

Emre Salman (S'03–M'10) received the B.S. degree in microelectronics engineering from Sabanci University, Istanbul, Turkey, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from the University of Rochester, New York, NY, USA, in 2006 and 2009, respectively.

He was with STMicroelectronics, Istanbul, Turkey, Synopsys, Mountain View, CA, USA, and Freescale Semiconductor, Tempe, AZ, USA, working in the fields of custom integrated circuit design, noise, and timing analysis. He has been an Assistant Professor with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY, USA, since 2010, where he is the Director of the Nanoscale Circuits and Systems Laboratory. He has authored a comprehensive tutorial book entitled *High Performance Integrated Circuit Design* (McGraw-Hill, 2012), which unifies interconnect-centric design methodologies for nanoscale integrated circuits. His current research interests include analysis, modeling, and design methodologies for high performance, and energy efficient integrated circuits with emphasis on power, signal, and sensing integrity.

Prof. Salman was a recipient of the National Science Foundation CAREER Award in 2013 and the Outstanding Young Engineer Award from the IEEE Long Island in 2014. He serves as a Regional Editor of the *Journal of Circuits, Systems and Computers*. He also serves on the Editorial Boards of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS and the *Journal of Low Power Electronics and Applications*.