A Cross-Layer Design Exploration of Charge-Recycled Power-Delivery in Many-Layer 3D-IC

Runjie Zhang†, Kaushik Mazumdar‡, Brett H. Meyer*, Ke Wang†, Kevin Skadron†, Mircea Stan‡

†Dept. of Computer Science
University of Virginia
Charlottesville, VA, USA

*Dept. of Elec. & Comp. Eng.
McGill University
Montréal, QC, Canada

‡Dept. of Elec. & Comp. Eng.
University of Virginia
Charlottesville, VA, USA

{runjie, km3sj, kewang, skadron, mircea}@virginia.edu, brett.meyer@mcgill.ca

ABSTRACT

3D-IC technology brings both the opportunities to continue the historical trend of integration-level scaling and the challenges to deliver power reliably and efficiently. Voltage-stacking (V-S), a charge-recycled power delivery scheme that connects the different layers' supply/ground nets into a series stack, provides a scalable solution to the 3D-IC power delivery wall. While prior work has extensively discussed the implementations of V-S at circuit-level, a cross-layer study that examines its system-level implications is missing. In this paper, we start with a circuit implementation of a charge-recycled voltage regulator and build an architecture-level model to study the costs and benefits of utilizing V-S in 3D-IC. Our study shows that by significantly improving the EM-lifetime of C4 and TSV array (e.g., up to 5x) while only marginally increasing the average-case voltage noise (e.g., 0.75% Vdd IR drop), V-S provides a scalable solution for many-layer 3D-IC's power delivery challenge.

Categories and Subject Descriptors

B.7.2 [**Design Aids**]: Simulation

General Terms

Design and Reliability

Keywords

3D stacking, Power distribution network, Voltage noise

1. INTRODUCTION

Because the benefits of Dennard scaling (devices that are simultaneously smaller, faster and lower power) are quickly vanishing, three-dimensional integrated circuits (3D-IC) are becoming an essential path to maintain exponential growth in device integration. However, 3D-IC raises several fundamental technical difficulties in addition to the fabrication challenges. Because the number of physical layers in a 3D-IC stack is expected to increase in the future, the already serious problems of delivering power to and removing heat from the 3D stack will be even worse. The main culprit is the fundamental mismatch between the volumetric (cubic) aspect of power consumption and dissipation in 3D-IC, and the fact that power *delivery* and affordable heat *removal* are limited to only the top or bottom 2D surface (quadratic). With the advance in the development of volumetric cooling technologies (e.g., micro-channel cooling [15]), power delivery may become an even more serious constraint.

To alleviate the power delivery constraints in the era of 3D-IC, previous research proposals [6, 9] suggest using the idea of voltage-stacking (V-S) to build the power delivery network (PDN) for 3D-IC. V-S simply refers to the power delivery arrangement of two or more circuit blocks such that the ground of one block becomes the power supply connection for the next: the blocks are connected as a series stack for power delivery, with all of them sharing the same current while their Vdd values are added. With the help of voltage-stacking's ability to "recycle" current between blocks, adding more layers to a 3D-IC only requires increasing the off-chip supply voltage while the current density within the PDN stays constant. For this reason, V-S provides a scalable solution to break the mismatch between 3D volume power dissipation and 2D surface power delivery.

To make the envisioned 3D-IC V-S practical, explicit voltage regulation is required for the general case when the currents of the various layers are not perfectly matched. While circuit solutions have been proposed for these explicit regulators [6, 9], a cross-layer tradeoff study that examines the benefits of voltage-stacking's current reduction, the area overhead and power efficiency of explicit voltage regulation, and the supply voltage noise under different workload conditions, is missing from the literature. For example, it is intuitive that compared with "regular" PDNs, V-S PDNs are more robust to electromigration (EM) wearout due to the reduced current density in through-silicon-via (TSV) and Controlled Collapse Chip Connection (C4) pads. However, it is not clear how 3D-IC scaling (i.e., more layers) affects the TSV/C4 array's EM lifetime, or whether designers can improve regular PDNs' EM-robustness to match V-S PDNs' lifetime by allocating more power supply TSVs and pads.

The major contributions of this paper are:

- A system-level PDN model for 3D-ICs that supports the study of EM-induced reliability and supply voltage noise for both regular and voltage-stacked PDN. With a fine-grained modeling granularity and the ability to capture on-chip voltage regulators' power efficiency and output voltage drop, our model can help system designers to evaluate the benefits and costs of design scenarios with different number of regulators and different TSV/C4 pad allocations.

- A detailed analysis of voltage stacking's impact on power-supply C4 pad and TSV array's EM-induced

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
DAC'15, June 07–11, 2015, San Francisco, CA, USA.
Copyright is held by the owner/author(s). Publication rights licensed to ACM
ACM 978-1-4503-3520-1/15/06$15.00
http://dx.doi.org/10.1145/2744769.2744774

lifetime. Our analysis indicates that although stacking more silicon layers quickly degrades regular PDNs' EM-lifetime, V-S PDNs' EM robustness is much less sensitive to silicon-layer count. For an 8-layer 3D processor, V-S improves the EM-induced lifetime of C4 pad and TSV array by up to 5x.

- Demonstrating the importance of workload imbalance (i.e., the power consumption difference between two adjacent layers in a 3D-IC) as a V-S design consideration and quantifying its impact on supply voltage noise, power efficiency and area overhead. Simulation results show that with the same total area overhead and an average workload imbalance ratio extracted from full applications, a V-S PDN's IR drop is only marginally larger (i.e., 0.75% Vdd) than a regular PDN.

These findings in turn suggests that V-S provides a scalable and practical solution to the power delivery challenge in the era of many-layer 3D-IC.

2. BACKGROUND AND RELATED WORK

2.1 Voltage Regulation in V-S PDN

Compared with the conventional power delivery scheme for an N-layer 3D-IC, V-S reduces the off-chip and cross-layer current density by up to N times through recycling charges between layers. This not only reduces the resistive noise (i.e., IR drop) across the PDN, but also significantly improves PDN's EM-induced reliability. However, a major design challenge of V-S arises from the fact that V-S will try to compensate for any current-consumption mismatch between the stacked loads by re-distributing the intermediate voltage. This effect of workload imbalance gives rise to voltage noise, which can disrupt the functionality of the stacked circuits. Explicit regulators have been proposed to handle this accumulation of charge imbalance at the intermediate nodes. Unlike conventional regulation schemes, where the regulators provide 100% of the current required by the loads (e.g., [19]), V-S requires differential converters that only handles the current-mismatch between the layers, and thereby converters with smaller passives can attain higher efficiency than conventional regulation. These differential converters have a "push-pull" ability that can either source or sink charges depending on the behavior of the loads.

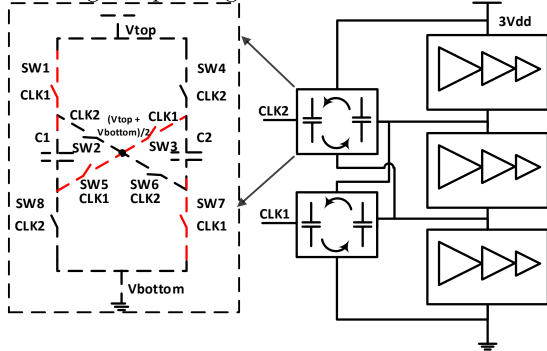


Figure 1: Stacked loads (three layers) with stacked SC converters (two), ideally providing Vdd voltage headroom to each load. Zoomed up single cell of 2:1 push-pull SC converter shown on the left.

To regulate workload imbalance, pioneering work proposed a push-pull linear regulator [13] for V-S PDN. Although linear regulators have low area overheads, they suffer from poor power efficiency due to their resistive nature, especially when the current imbalance is large. More recent work proposed a push-pull switched-capacitor (SC) to recycle the charge imbalance between the stacked loads [9]. Because of their energy-storage capability, these SC converters provide higher power-efficiency at the cost of a larger silicon area dedicated to the capacitors. Various surveys and comparisons of switching regulators in the literature [17] show that, with the rapid improvement of capacitive technology, switched-capacitors are going to surpass inductive converters. We therefore focus on SC converters in this paper and leave the study of inductive converters for future work. Fig. 1 shows the structure of the SC converter we implement/model in this paper. It involves two fly-capacitors (C1 and C2) interchanging their positions periodically, thereby shuttling excess charge between the stacked loads to "source" or "sink" them as the loads demand. To support many-layer 3D-IC, we extend this converter for two stacked loads [9] into a scalable, multi-output ladder SC.

2.2 System-level Evaluation of V-S

Although researchers have previously identified V-S as a promising solution to alleviate the power delivery constraints in 3D-IC [9], the impact of V-S on PDN current density and the resulting implications to PDN reliability has not been closely investigated. In this paper, we build a system-level PDN model and simulate an example many-core 3D processor to directly compare the EM-induced lifetime of regular and V-S PDNs. Another important aspect of V-S design is the voltage noise at the intermediate nodes. Zhou et al. proposed a whole-system PDN model to study SC converters' impact on power delivery noise [19]. However, they only studied the traditional 2D-IC case without voltage stacking. To the best of our knowledge, a system-level noise evaluation for SC-converter-supported V-S PDNs is missing from the literature. Adopting a design methodology similar to [19], we combine a resistive model of SC converters [14] with our 3D extension to an existing PDN model [18] to evaluate V-S PDN's noise level. Since the supply noise in V-S PDN is strongly correlated with the workload-imbalance between the adjacent layers [9], we also examine a large range of workload-imbalance and quantify its impact on voltage noise, system power efficiency, and PDN area overhead.

3. MODELING METHODOLOGIES

3.1 SC Converter Modeling

A cross-layer design exploration of the benefits and overheads of V-S in 3D-IC requires incorporating circuit-level insights with architecture-level study. To accurately capture the power efficiency, output voltage drop and area overhead of SC converters, we implement a 2:1 push-pull SC converter (as shown in Fig. 1) in a commercial 28nm CMOS technology. It has integrated fly capacitors (8nF total), an optimum switching frequency of 50MHz, and 4-way interleaving. Each SC converter can provide up to 100mA current to the load. Using the Cadence ADE environment and Spectre simulator, we simulate this converter and use the results to derive a compact model for system-level exploration.

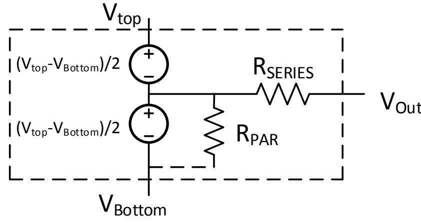


Figure 2: Compact model for SC converters.

Fig. 2 shows the efficiency and noise model for the SC converters. We adopt an analytical methodology introduced in [14]. Based on the switch topology, the charge multiplier vectors ($a_{c,i}$ and $a_{r,i}$) are derived to calculate the slow (R_{SSL}) and fast switching (R_{FSL}) asymptotic limits of SC converter output impedance. The optimized R_{SSL} and R_{FSL} are given as:

$$R_{SSL} = \frac{1}{C_{tot} f_{SW}} \left(\sum_{i}^{n} |a_{c,i}| \right)^2 \quad (1)$$

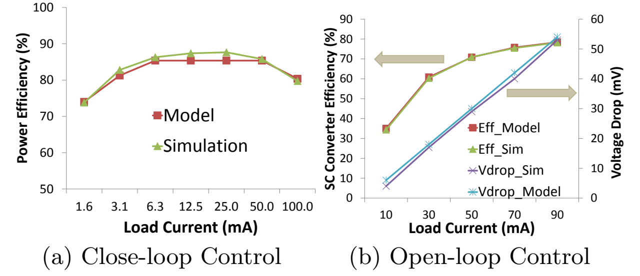
$$R_{FSL} = \frac{1}{G_{tot} D_{cyc}} \left(\sum_{i}^{n} |a_{r,i}| \right)^2 \quad (2)$$

where C_{tot} is the fly capacitance, G_{tot} is the total switch conductance, f_{SW} is the switching frequency, and D_{cyc} is the duty cycle (assumed 50%). The R_{SERIES} in Fig. 2 captures the switching and conductance losses while R_{PAR} captures the various parasitic losses of switch parasitic capacitance, bottom-plate capacitance and gate-drive loss. This model also captures the resistive voltage drop of the SC converters through R_{SERIES}, which can be calculated as: $R_{SERIES} = \sqrt{R_{SSL}{}^2 + R_{FSL}{}^2}$. For the SC converter we implemented, $R_{SERIES} = 0.6\Omega$.

As shown in Fig. 1, the voltage-headroom (i.e., the potential difference between V_{Top} and V_{Bottom}) of the SC converters in many-layer 3D-ICs is dependent on the adjoining layers' workload imbalance. In order to incorporate this dependency in our cross-layer study, we make both V_{Top} and V_{Bottom} as inputs to our SC converter model and calculate the ideal output voltage (i.e., without the IR drop on R_{SERIES}) as $(V_{Top} + V_{Bottom})/2$.

To verify the accuracy of this model, we compare the estimated power-efficiency and output voltage drop against circuit simulation results of a SC converter for a 2-layer 3D-IC under fixed capacitance and different load current. We test two different frequency modulation strategies. The closed-loop scheme modulates SC converter's switching frequency dynamically with the load current while the open-loop control scheme keeps the frequency constant at all time. Fig. 3 shows that our model accurately captures power-efficiency and output voltage drop for both control policies. According to Fig. 3, closed-loop converters have higher power-efficiency. However, because it requires the implementation of feedback loops, the closed-loop policy is more complex to model. For simplicity, we use open-loop SC converters and leave the evaluation of closed-loop control for future work.

We implement our SC converters with MIM capacitors and the resulting area of each converter is $0.472mm^2$. Considering the fact that the fly-caps contribute to the majority of SC-converters' area and MIM capacitors have low density, we also calculate the converters' area overhead with other high-density integrated capacitors. For example, if implemented with ferreolectric [17] or trench capacitors [12], the area of each converter would be $0.102mm^2$ or $0.082mm^2$.

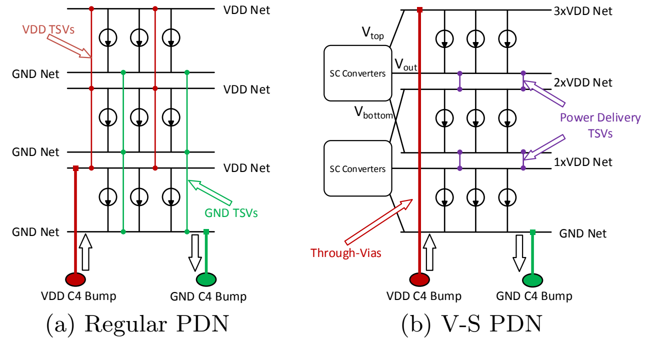


(a) Close-loop Control (b) Open-loop Control

Figure 3: Model validation results.

3.2 PDN Modeling for 3D-IC

The PDN of a modern processor usually consists of millions of nodes, which require a significant amount of time for detailed simulation. In order to quickly explore the multi-dimensional space of 3D-IC's PDN design and evaluate the cost and benefits of different design scenarios, we perform our early-stage PDN simulation using VoltSpot, a pre-RTL PDN model [18]. VoltSpot uses ideal current sources to model load (i.e., both dynamic and leakage power of the switching transistors) and RLC elements to model the on-chip PDN metal stack, C4 pads and chip package. Since VoltSpot only models 2D chips, we extend it to support 3D-IC. Fig. 4 illustrates our extensions.



(a) Regular PDN (b) V-S PDN

Figure 4: PDN structure for 3D-IC.

To model the traditional PDN for 3D-IC, we simply add more layers of silicon on top of each other and connect all layers' Vdd nets and ground nets with TSVs (Fig. 4a). To model V-S PDN, we connect all layers' Vdd nets and ground nets in series with regular TSVs and provide the off-chip supply voltage (i.e., the single layers' Vdd multiplied by the number of layers) to the top layer using through-vias (Fig. 4b). TSVs are modeled as resistors. The resistive model for SC converters has been described in Sec. 3.1. We uniformly distribute them within each core.

With the fine-grained pre-RTL modeling capability inherited from VoltSpot, our 3D-IC PDN model provides a detailed current profile for both the C4 pad and TSV arrays. It also captures on-chip IR drop for both regular PDN and V-S PDN under given workload behaviors. This model provides a key link to the tool chain that allows designers to explore the complex tradeoff space that involves power delivery architecture, C4 pad/TSV allocation, voltage regulation scheme, PDN noise/reliability, and workload characteristics.

3.3 EM-induced System Lifetime Calculation

Due to the momentum transfer between electrons and metal atoms, high-density and continuous current flow can incur gradual mass transportation in metal conductors and eventually cause open or short circuit. This phenomenon is referred to as electromigration. A metal conductor's EM-induced lifetime follows a lognormal distribution and the mean-time-to-failure can be estimated with Black's equation [4]. For a group of conductors (e.g., C4 pad or TSV array), all elements are subject to EM-induced wearout. Therefore we adopt a mean-time-to-failure (MTTF) calculation method from [18] that considers the failure chance of multiple conductors: $P(t) = 1 - \prod_i (1 - F_i(t))$, where $P(t)$ is the whole group's failure-probability cumulative distribution function (CDF), and $F_i(t)$ is each conductor's chance of failure after time t. Using the detailed per-pad/TSV current information generated by our PDN model, we first determine $F_i(t)$ for each pad/TSV's CDF. After that, we calculate $P(t)$ and use the time value which makes $P(t) = 0.5$ as a lifetime estimation that represents the whole pad/TSV array's expected lifetime until the first EM-induced failure. We will use this metric (expected EM-damage-free-lifetime) in the remainder of this paper to evaluate PDN's robustness against EM stress.

4. SIMULATION SETUP

4.1 Many-core Processor Modeling

In order to establish realistic 3D-IC design scenarios for our cross-layer exploration, we select a 40nm, dual-core ARM Cortex A9 IP implementation running at 1.0 GHz [1] and replicate it 8 times to build a single-layer, 16-core processor. The reason for selecting ARM processors is that they are power-efficient and therefore can be used to build many-layer 3D-ICs without relying on aggressive, volumetric cooling solutions. We use McPAT [8], an architecture-level power and model to derive the area and power consumption of the single-layer processor. The processor floorplan was generated by ArchFP [5]. With a supply voltage of 1V, this 1GHz single-layer processor has a peak power consumption of 7.6 W and an area of 44.12 mm^2.

Although many-layer, especially many-logic-layer 3D-ICs pose various fabrication challenges, the possibility of manufacturing 3D stacks economically has been exemplified by existing commercial products (e.g., the Micron hybrid memory cube with 4-8 layers [10]). To study the voltage noise in both short-term and long-term future 3D-ICs, and to evaluate how 3D scaling affects PDN design tradeoff, we build a series of example 3D systems with 2 to 8 layers stacked together. With the help of a pre-RTL thermal model, HotSpot [16], we find that we can build 3D-ICs with up to 8 layers of our example 16-core processor while maintaining the hotspot temperature below 100 Celsius (which is a typical upper limit [16]) with a conventional air-cooling solution.

4.2 PDN Modeling and TSV Configurations

One of the major extensions we made to VoltSpot is adding an explicit model for TSVs. The diameter, pitch and resistance values of TSVs come from prior work [7]. As suggested by Pathak et al. [11], the thermal stress generated by TSVs could potentially impact the electrical performance of the nearby transistors. Therefore each TSV requires a keep-out zone (KoZ) to space away other active devices. We use the size of this KoZ to calculate the TSV array's total area occupancy. We assume that all TSVs have equal size and resistance, and they are uniformly distributed within each silicon layer. Other PDN modeling parameters are adopted from previous work [18] and listed in Table 1.

Parameter	Value
C4 Pad Pitch (μm)	200
C4 Pad Resistance ($m\Omega$)	10
Minimum TSV Pitch (μm)	10
TSV Diameter (μm)	5
Single TSV's Resistance ($m\Omega$)	44.539
TSV Keep-Out Zone's Side Length (μm)	9.88
On-chip PDN's Pitch,Width,Thickness (μm)	810,400,720

Table 1: Major PDN modeling parameters

The number of TSVs allocated for PDN is a design parameter for system designers. More TSVs provide more vertical current delivery channels, therefore reducing both average TSV current and the effective inter-layer PDN resistance. At the cost of higher area overhead due to the KoZs, increasing the number of power-supply TSVs not only reduces voltage noise, but also improves the TSV array's reliability against EM-induced wearout. To explore the tradeoff between power delivery quality and TSVs' area overhead, we examine three TSV topologies in our study that represent a conservative (Dense), an aggressive (Few) and an average (Sparse) design scenario. Table 2 gives more details about each configuration's TSV count and area overhead.

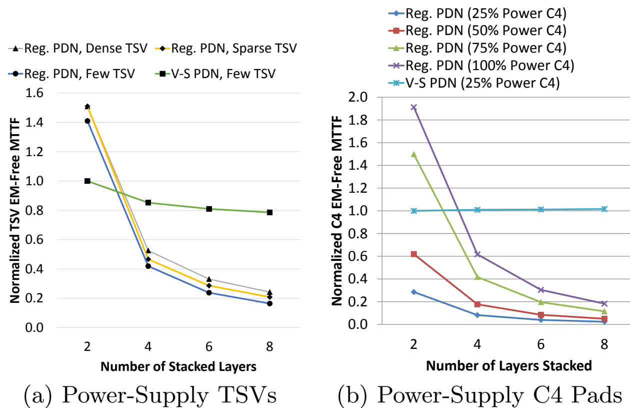
	Effective Pitch(um)	Number of TSVs per Core	Total Area Overhead
Dense TSV	20	6650	24.2%
Sparse TSV	40	1675	6.1%
Few TSV	240	110	0.4%

Table 2: TSV configurations used in this study.

5. RESULTS

5.1 EM-Induced TSV/C4 Pad Lifetime

Using the methodology described in Sec. 3.3, we evaluated the expected EM-damage-free lifetime for both regular and V-S PDN's TSV (Fig. 5a) and C4 pad (Fig. 5b) arrays. As we stack more layers, the increasing current density significantly reduces the lifetime of the regular PDN's TSV array by up to 84%. At the same time, the V-S PDN's MTTF only slightly degrades. This is because while the current density of TSVs in the V-S PDN is independent of layer count, adding more layers still requires more TSVs to support them, which increases the risk of TSV failures. We also observe that the V-S PDN's TSV array has a shorter lifetime compared to the regular PDN when the number of stacked layers is small (e.g., 2 layers). This is because in the V-S PDN, we connect each Vdd C4 pad with only one TSV, to provide supply voltage/current directly to the top-layer. Since the number of Vdd pads (32 per-core in this case) is smaller than the number of Vdd TSVs in a regular PDN (55 per core in the "Few TSV" case), the Vdd TSVs in the V-S PDN have higher average current, which limits the whole-system MTTF. Regardless of this side effect, the EM-lifetime of V-S PDNs in 3D-ICs with more layers still surpasses that of the regular PDN by more than 3x.



(a) Power-Supply TSVs (b) Power-Supply C4 Pads

Figure 5: EM-induced lifetime evaluation. All results are normalized to the lifetime of the 2-layer V-S PDN.

Similarly, the regular PDN's C4 MTTF quickly degrades with 3D-IC scaling. For the V-S PDN, stacking more layers neither increases the number of total pads, nor raises the total off-chip current demand, and therefore its C4 array's EM-damage-free lifetime is independent of layer count. For the 8-layer 3D processor, the gap in the C4 array's MTTF between the V-S PDN and the regular PDN can be up to 5x. This indicates that, because V-S extends the pad array's EM lifetime, it reduces the requirement for power supply pads and allows more pads to be used for I/O. Please note that since the C4 array's EM robustness is insensitive to the TSV topology, we use a fixed topology in all the evaluations.

Another interesting observation is that, for the regular PDN, adding more TSVs or C4 pads only marginally increases MTTF. Even with aggressive allocations (e.g., "Dense TSV" topology or even allocating 100% of pads as power supply), the regular PDN's MTTF is still far inferior to that of the V-S PDN. We therefore conclude that for many-layer 3D-ICs, it is not feasible to improve the regular PDN's EM-robustness to the same extent as with the V-S PDN by simply allocating more power-supply TSVs and C4 pads.

5.2 Load-Imbalance-Induced Voltage Noise

Integrated-voltage-regulation is necessary in V-S PDN, because when the current consumptions of two adjacent layers do not match, the voltage regulators need to either provide or sink the difference. This introduces extra voltage noise due to the regulators' output voltage drop and the lateral impedance of the on-chip PDN. While larger workload-imbalance increases noise with higher current demand for the SC converters, having more regulators distributed across the silicon die reduces IR drop by amortizing the per-converter current load and reducing the average load-to-regulator distance. Fig. 6 shows the noise levels of PDNs for a 8-layer 3D-IC under different regulator configurations and workload behavior conditions. We assume that the power consumption of the silicon layers has an interleaved "high-low" pattern, where the high-power layers are always fully active and the low-power layers consume X% lower dynamic power (e.g., 100% imbalance means that the low-power layers are idle and only consume leakage power). This pattern serves as a good benchmark, because it requires the converters on all layers to source/sink the same amount of current, therefore imposing the most stress on the PDN. We note that since our SC converter has a maximum load limit of 100 mA, Fig. 6 skips all data points that violate this constraint.

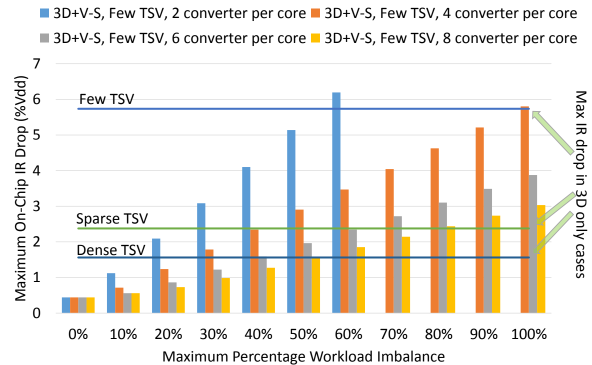


Figure 6: Voltage noise evaluation of our 8-layer processor. For 3D-ICs without V-S, the worst-case IR drop happens when all layers are fully active. Therefore the assumption about workload-imbalance does not affect those evaluations.

The lines in Fig. 6 illustrate the maximum on-chip IR drop of regulator PDNs with different TSV configurations. Regular PDNs rely on TSVs to provide all current to all layers, and therefore the worst-case IR drop always happens when all layers are fully active. For this reason, regular PDNs' maximum IR drop results are irrelevant to the imbalance of workloads. Since adding one SC converter to an ARM core incurs around 3% area overhead (assuming the converters are implemented with high-density capacitors discussed in Sec. 3.1), a V-S PDN with 8 converters per core and "Few TSV" topology occupies the same area as a regular PDN with "Dense TSV" topology. If we compare the voltage noise of theses two cases, we find that the V-S PDN has lower IR drop when the workload-imbalance ratio is lower than 50%. When larger imbalance exists, V-S PDN's IR drop surpasses regular PDN by up to 1.58% Vdd.

To give an example of workload-imbalance in full applications, we simulate the Parsec 2.0 benchmark suite [2] with performance simulator Gem5 [3] and adopt the methodology of statistical sampling from prior work [18]. We simulate one thousand 2k-cycle samples from each application and calculate their average power consumptions with McPAT. Fig. 7 shows the distribution of each application's power consumption. The top/bottom bars in Fig. 7 represent the max/min values of each distribution. The edges of the boxes are the 25th and 75th percentiles, and the central marks are the medians. We observe that although the samples from different applications have large differences in power consumption, the samples from the same application show much smaller

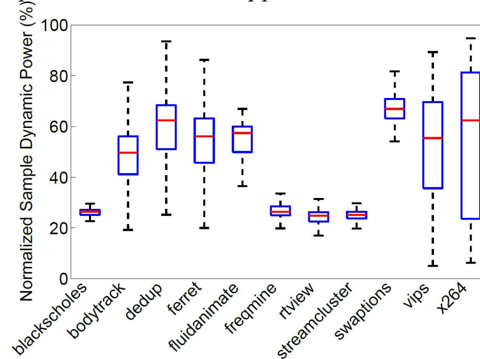


Figure 7: A box-plot that shows the distributions of workload imbalance within and across different applications.

variance. For example, while the maximum workload imbalance among all samples is more than 90%, the best-case application (blackscholes) shows a maximum imbalance of 10% across all its samples. On average, the applications have a maximum-imbalance ratio of 65%, which makes the V-S PDN's IR drop only 0.75% larger than the regular PDN. These results indicate that by scheduling different instances of the same application, or different threads from the same instance onto the cores in the same core-stack, we can reduce the workload-imbalance and a V-S PDN's noise.

5.3 System Power Efficiency

Figure 8 shows the power efficiency (i.e., the total power consumed by the processors divided by the total power drawn from the off-chip power source) results for 3D-ICs with V-S PDN. As the amount of workload-imbalance increases, the SC converters need to deliver more power for compensation. Consequently, the power overhead of voltage regulation increases. When we compare V-S PDNs with different numbers of SC converters, we observe that increasing the number of converters reduces power efficiency. As we discussed in Sec. 3.1, this is because our open-loop converters do not modulate their switching frequency at run-time, and therefore each converter's efficiency reduces as more converters are allocated to share the current load. Closed-loop control is an area for future work. Considering the fact that placing more converters can reduce on-chip IR drop, the allocation of SC converters in V-S PDN becomes a tradeoff between on-chip voltage noise and system-level power efficiency. Our models can help designers to choose the optimal design point based on their specific design objectives.

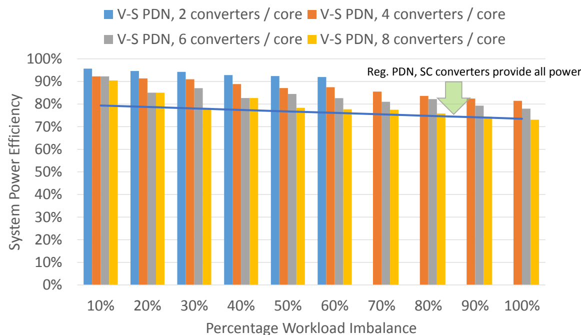


Figure 8: Power regulation efficiency of 3D processors.

Figure 8 also shows the power efficiency of using SC converters in 3D processors with regular PDN. Unlike V-S PDN, where the voltage regulators only need to compensate for the differential power consumption between layers, SC converters in regular PDN have to provide current to all layers. As a result, V-S PDNs have higher power efficiency.

6. CONCLUSIONS

3D-IC provides an essential mechanism for the industry to stay on the historical scaling trend of device integration while raises power delivery challenges with reduced EM-lifetime and increased voltage noise. In this paper, we build a system-level PDN model for 3D-ICs to study a charge-recycled, voltage-stacking PDN structure and compare it with the regular, non-voltage-stacked PDNs in the context of 3D-IC. Our EM-robustness analysis on both C4 pad and TSV arrays indicates that V-S PDN's EM-induced MTTF significantly surpasses (e.g., 5x longer) the regular PDN's lifetime. By implementing and validating a resistive model for V-S PDN's voltage regulator (i.e., SC converters), we analyze the on-chip voltage noise and observe that with the same total area overhead, the V-S PDN has lower IR drop than the regular PDN when the workload-imbalance ratio is below 50%. Under the average workload imbalance ratio extracted from full applications (65%), a V-S PDN's IR drop is no greater than 0.75% Vdd beyond the noise level of a regular PDN. Combined with the observation that both EM-lifetime and IR drop of V-S PDNs are insensitive to many-layer 3D-ICs' layer count, our study demonstrates that V-S provides a scalable and practical solution to the power delivery challenge in the era of many-layer 3D-IC.

Acknowledgment

This work is supported by NSF grant CNS-0916908, and DARPA MTO under contract no. HR0011-13-C-0022.

7. REFERENCES

[1] ARM. http://www.arm.com/products/processors/cortex-a/cortex-a9.php.
[2] C. Bienia. *Benchmarking Modern Multiprocessors*. PhD thesis, Princeton University, Jan 2011.
[3] N. Binkert et al. The gem5 simulator. *SIGARCH Comput. Archit. News*, 39(2), Aug 2011.
[4] J. Black. Electromigration: A brief survey and some recent results. *IEEE Transactions on Electron Devices*, 16(4):338–347, 1969.
[5] G. Faust, R. Zhang, K. Skadron, M. Stan, and B. Meyer. ArchFP: Rapid prototyping of pre-RTL floorplans. *VSLI-SoC*, 2012.
[6] P. Jain, T.-H. Kim, J. Keane, and C. H. Kim. A multi-story power delivery technique for 3d integrated circuits. In *ISLPED*, 2008.
[7] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene. Electrical modeling and characterization of through silicon via for three-dimensional ics. *IEEE Transactions on Electron Devices*, 57(1), 2010.
[8] S. Li et al. Mcpat: An integrated power, area, and timing modeling framework for multicore and manycore architectures. In *MICRO*, 2009.
[9] K. Mazumdar and M. Stan. Breaking the power delivery wall using voltage stacking. In *GLSVLSI*, 2012.
[10] Micron. http://www.micron.com/products/hybrid-memory-cube.
[11] M. Pathak, Y. J. Lee, T. Moon, and S. K. Lim. Through-silicon-via management during 3d physical design: When to add and how many? In *ICCAD*, 2010.
[12] C. Pei et al. A novel, low-cost deep trench decoupling capacitor for high-performance, low-power bulk CMOS applications. In *ICSICT*, 2008.
[13] S. Rajapandian, Z. Xu, and K. L. Shepard. Implicit dc-dc downconversion through charge-recycling. *Solid-State Circuits, IEEE Journal of*, 40(4), 2005.
[14] M. D. Seeman. A design methodology for switched-capacitor DC-DC converters. Technical report, DTIC Document, 2009.
[15] D. Sekar et al. A 3D-IC technology with integrated microchannel cooling. In *IITC*, 2008.
[16] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, D. Tarjan, and K. Sankaranarayanan. Temperature-aware microarchitecture. In *ISCA*, 2003.
[17] M. Steyaert, T. Van Breussegem, H. Meyvaert, P. Callemeyn, and M. Wens. DC-DC converters: From discrete towards fully integrated cmos. In *ESSDERC*, 2011.
[18] R. Zhang, K. Wang, B. H. Meyer, M. R. Stan, and K. Skadron. Architecture implications of pads as a scarce resource. In *ISCA*, 2014.
[19] P. Zhou, D. Jiao, C. H. Kim, and S. S. Sapatnekar. Exploration of on-chip switched-capacitor dc-dc converter for multicore processors using a distributed power delivery network. In *CICC*, 2011.