A 16-Core Voltage-Stacked System With Adaptive Clocking and an Integrated Switched-Capacitor DC–DC Converter

Sae Kyu Lee, *Student Member, IEEE*, Tao Tong, *Student Member, IEEE*, Xuan Zhang, *Member, IEEE*, David Brooks, *Fellow, IEEE*, and Gu-Yeon Wei, *Member, IEEE*

***Abstract*— This paper presents a 16-core voltage-stacked system with adaptive frequency clocking (AFClk) and a fully integrated voltage regulator that demonstrates efficient on-chip power delivery for multicore systems. Voltage stacking alleviates power delivery inefficiencies due to off-chip parasitics but adds complexity to combat internal voltage noise. To address the corresponding issue of internal voltage noise, the system utilizes an AFClk scheme with an efficient switched-capacitor dc–dc converter to mitigate noise on the stack layers and to improve system performance and efficiency. Experimental results demonstrate robust voltage noise mitigation as well as the potential of voltage stacking as a highly efficient power delivery scheme. This paper also illustrates that augmenting the hardware techniques with intelligent workload allocation that exploits the inherent properties of voltage stacking can preemptively reduce the interlayer activity mismatch and improve system efficiency.**

***Index Terms*— Adaptive frequency clocking (AFClk), dc–dc converter, multicore, power delivery, voltage noise, voltage stacking.**

I. Introduction

EFFICIENT power delivery is a critical design target for modern computing systems from high-performance servers to mobile devices. Continued decreases in supply voltages and aggressive power reduction techniques (e.g., clock and power gating) under a fixed power budget have led to increases in average current draw and worsening current transients, forcing stringent requirements on the power delivery impedance. Today's 100-W high-performance processors operate under 1 V, draw excess of 100 A, and require $<$1-mΩ impedance for 10% voltage noise margin, which is extremely challenging to achieve. Furthermore, significant I^2R power loss in the off-chip parasitic resistance of the power delivery network can greatly degrade the overall system efficiency, while electromigration due to high current levels is a concern. Exacerbating the issue, the off-chip components ostensibly have not scaled, contrary to the ever-decreasing power delivery impedance requirements needed to keep up with the high current demands of modern computing systems.

Voltage stacking is an on-chip power delivery solution that delivers a high voltage to the chip by vertically stacking voltage domains in series and recycling charge through the stacked layers, thereby reducing the overall chip current demands [1]–[11]. For the same chip power, an n-way stacked system reduces the current draw of the chip proportionally by n, which reduces IR drop by factor of n, I^2R power loss by n^2, and alleviates the off-chip impedance requirements for the same voltage noise margin. It also obviates a high step-down off-chip dc–dc converter, which improves off-chip regulator efficiency.

There has been growing interest in integrating dc–dc converters on-chip to perform voltage conversion from the high input voltage levels provided to the chip [12]–[15]. However, such integration usually suffers from inferior conversion efficiency due in part to poor quality inductors and capacitors available on-chip. By stacking voltage domains and obviating the explicit voltage conversion stage, voltage stacking achieves high efficiency power delivery. Ideally, if the power consumption of all stacked layers perfectly matches, the layer voltages evenly subdivide the high input voltage, and voltage stacking achieves optimal intrinsic step-down voltage conversion with no loss. In practice, interlayer switching activity mismatch exists and results in interlayer voltage noise due to the series-connected nature of voltage stacking; wherein, this susceptibility to voltage noise negatively impacts system performance and energy efficiency, and poses system reliability concerns.

To address the corresponding voltage noise issue associated with voltage stacking, this paper presents a 16-core four-way voltage-stacked test chip implemented in TSMC's 40 G process that integrates industry grade microprocessor cores with a multioutput *integrated voltage regulator* (IVR) and implements adaptive frequency clocking (AFClk) to mitigate the impact of voltage noise. The IVR provides minimum voltage guarantees while AFClk allows the cores to operate efficiently with minimal margin. With voltage stacking,

Manuscript received June 1, 2016; revised October 2, 2016; accepted November 3, 2016. Date of publication December 19, 2016; date of current version March 20, 2017. This work was supported in part by NSF under Grant CCF-0903437 and Grant CCF-1218298, and in part by DARPA under Grant HR0011-13-C-0022. This paper was recommended by Associate Editor I. Savidis.

S. K. Lee, T. Tong, D. Brooks, and G.-Y. Wei are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: saekyu@eecs.harvard.edu; taotong@seas.harvard.edu; dbrooks@eecs.harvard.edu; guyeon@eecs.harvard.edu).

X. Zhang was with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA. He is now with Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail: xuan.zhang@wustl.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.2016.2633805

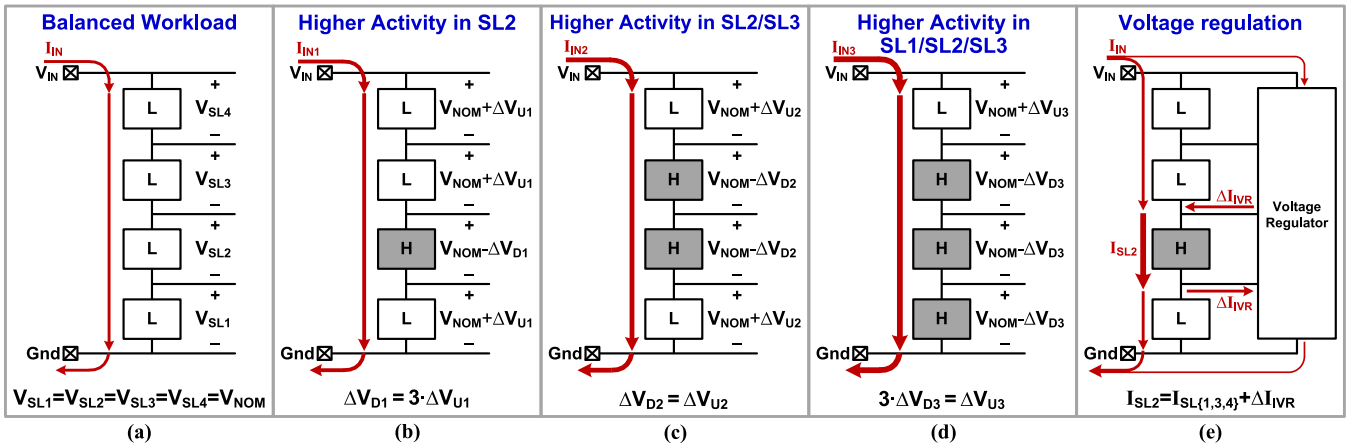


Fig. 1. Block diagram of various workload scenarios in a four-core four-way voltage-stacked system illustrating the basic properties of voltage stacking. (a) Balanced scenario. (b) Higher activity in SL2. (c) Higher activity in SL2/SL3. (d) Higher activity in SL1/SL2/SL3. (e) Voltage regulation.

IVR losses only apply to the mismatch-related power the IVR provides, which is a small fraction of the total power for high-throughput multicore systems. Measurement results demonstrate system-wide power delivery efficiency greater than 94% for high-throughput workload scenarios.

The remainder of this paper is organized as follows. Section II provides a brief background on the properties of voltage noise in voltage-stacked systems and existing noise mitigation schemes. Section III then presents the overall architecture and implementation details of the test chip. Finally, extensive measurement results from the test chip, presented in Section IV, demonstrate the system-level performance and efficiency advantages of a voltage-stacked system with an IVR and adaptive clocking.

II. Voltage Stacking Background

A. Voltage Noise in Voltage-Stacked Systems

Characterizing and managing voltage noise is crucial to synchronous digital systems, as the clock frequency must be chosen to ensure correct operation under worst case voltage noise conditions. Typically, in conventional digital systems, voltage noise is caused by the interaction between the load current fluctuations and the parasitic inductance and resistance of the power delivery network. Voltage stacking alleviates these conventional voltage noise issues by reducing the overall load current. However, series-connected voltage domains are susceptible to interlayer voltage noise with inherently different attributes. Fig. 1 shows an example of interlayer voltage noise using a simple four-core, four-way voltage-stacked system. For simplicity, we assume ideal off-chip connections and assume synchronous digital loads with no leakage current. The system operates off of a fixed V_{IN} of $4 \times V_{\mathrm{NOM}}$, where V_{NOM} is the nominal operating voltage of the cores.

KCL dictates that identical current must flow through the series-connected cores in the stack. If the power consumption of each stack layer (SL) perfectly matches, the SL voltages $V_{\mathrm{SL}\{1:4\}}$ will subdivide V_{IN} evenly, as shown in Fig. 1(a), where all four cores execute the same workload **L** with the same activity factor α. However, when there is imbalance in workload activity across the SLs, layer voltages deviate from V_{NOM} to maintain identical current flow through the stack. Fig. 1(b) shows such a scenario where the core in SL2 executes a higher activity workload **H** with activity factor $(\alpha + \Delta\alpha)$. To maintain identical current flow through the stack in this scenario, V_{SL2} decreases, while V_{SL1}, V_{SL3}, and V_{SL4} increase, causing voltage droop in V_{SL2}. With V_{IN} fixed, the droop in V_{SL2} equals the sum of voltage increases in V_{SL1}, V_{SL3}, and V_{SL4}. For each scenario shown in Fig. 1(b)–(d), I_{IN} can be expressed as

$$\begin{aligned} I_{\mathrm{IN}n} &= \alpha \cdot C \cdot f \cdot (V_{\mathrm{NOM}} + \Delta V_{\mathrm{U}n}) \\ &= (\alpha + \Delta\alpha) \cdot C \cdot f \cdot (V_{\mathrm{NOM}} - \Delta V_{\mathrm{D}n}) \end{aligned} \tag{1}$$

where $\alpha \cdot C$ and $(\alpha + \Delta\alpha) \cdot C$ are the total switched capacitance, f is the clock frequency, and V_{Dn} and V_{Un} are the magnitude of voltage droop and voltage increase in the SLs for each scenario as shown in Fig. 1(b)–(d). The scenarios presented in Fig. 1(b)–(d) and (1) reveal intuitive yet important insights into the determining factors of noise in voltage stacking.

1) Activity Mismatch: For all three scenarios shown in Fig. 1(b)–(d), layer voltages deviate from V_{NOM} to compensate for the activity mismatch $\Delta\alpha$ between SLs and maintain identical current through the stack. Equation (1) shows that larger activity mismatch $\Delta\alpha$ results in larger voltage deviations.

2) Current Dependence on Voltage: In addition to activity mismatch $\Delta\alpha$, interlayer voltage noise also depends on the magnitude of the nominal current flowing through the stack (I_{IN}), set by α. Higher current flow through the stack results in lower supply rail impedance for the SLs [4]. Hence, smaller values of ΔV_{Un} and ΔV_{Dn} can neutralize the same $\Delta\alpha$ when overall activity and power is higher. Moreover, if clock frequency f of each layer can vary proportionally with voltage, dynamic current has a quadratic, rather than linear, dependence on voltage, which further reduces voltage noise. In other words, the overall power consumption and the choice of clocking strategy can both impact voltage noise on the SLs.

3) Workload Profile: Finally, comparison of Fig. 1(b)–(d) shows how different workload conditions affect voltage droop, even for fixed values of α and $\Delta\alpha$. Since all layer

voltages must add up to V_{IN}, sum of ΔV_{Dn} must equal the sum of ΔV_{Un}. Therefore, the relative magnitudes of ΔV_{Dn} and ΔV_{Un} depend on the workload scenario. For example, Fig. 1(b) shows the worst case voltage droop, because the droop in V_{SL2} (ΔV_{D1}) must balance the sum of ΔV_{U1} in the other three layers. As more cores execute the higher activity workload in Fig. 1(c) and (d), ΔV_{Dn} becomes progressively smaller because more layers share the burden of voltage droop, while ΔV_{Un} becomes larger. This also results in larger I_{IN} flowing through the stack due to more cores executing higher activity workloads. In general, activity mismatch due to higher activity in only a single SL leads to worse voltage noise than activity mismatch in multiple layers.

B. Noise Mitigation in Voltage Stacking

While software techniques can implement workload balancing to mitigate interlayer activity mismatch, the timescales of such techniques are usually too long. It is infeasible to rely wholly on software techniques to provide the operating voltage guarantees necessary to avoid noise-related failures, such as SRAM instability or timing violations. Moreover, software techniques cannot easily address situations where entire layers must be powered down. Hence, we explore using efficient voltage regulators to neutralize activity mismatch and mitigate interlayer voltage noise. Fig. 1(e) shows how an on-chip IVR can provide the extra current to the higher activity core in SL2 and neutralize interlayer voltage noise. Notice the IVR only provides the differential current (ΔI_{IVR}), which means IVR losses only apply to a small fraction of the overall power delivered to the stack.

There are several examples of prior work that have proposed different voltage regulator topologies to regulate the stacked layers in a voltage-stacked system. Push–pull linear regulators have been proposed to handle the interlayer activity mismatch [1]. Although linear regulators have small area overhead and are easy to integrate, they suffer from inherently low conversion efficiency, especially for high step-down ratios. Off-chip inductor-based switching regulators have been shown for voltage stacking applications [5], [8], but the inherent difficulty of integrating high-quality inductors on-chip hinders full integration of inductor-based regulators. On the other hand, high-quality capacitors are much easier to integrate, making switched-capacitor (SC) converters attractive. For example, [3] demonstrated a 2-to-1 SC converter that regulates the intermediate voltage of two stacked layers. Multiple 2-to-1 SC converters have been used to support more SLs at the expense of higher complexity and overhead [6]. Building on prior work, this paper proposes using a symmetric-ladder topology to implement a fully integrated 4-to-1 SC converter that mitigates voltage noise on four SLs simultaneously.

One thing to point out, however, is that SC converters cannot regulate the SLs to an exactly even distribution of V_{IN}. This is because the SC converter relies on a certain amount of voltage difference between the output and the flying capacitors to deliver charge to the output. N-to-1 SC converters typically regulate the output voltage to ($V_{\mathrm{IN}}/N - \Delta V$), and the converter efficiency and its ability to deliver charge diminishes as ΔV approaches zero [14]. To overcome this limitation and augment the SC converter to improve system efficiency, this paper also explores per-layer AFClk, where clock frequency of cores in each SL tracks the fluctuations in layer voltage. AFClk is particularly a good fit with voltage-stacked systems. By having the per-layer clock frequency track voltage fluctuations, current consumption has a stronger dependence on voltage, which further alleviates voltage noise.

The remainder of this paper presents the design of a 16-core voltage-stacked test chip that demonstrates efficient voltage noise mitigation using an SC IVR and an AFClk scheme. We evaluate the noise mitigation and performance advantages the IVR and AFClk provide, and demonstrate the advantages of voltage stacking for efficient power delivery in multicore systems.

III. System Architecture

Fig. 2(a) presents an overview of the test chip implemented in TSMC's 40 G process. It comprises 16 processor cores organized into a 4 × 4 stacked array. The core array operates off a single 3.6 V V_{IN}, with cores in each SL (SL{1:4}) operating nominally at layer voltages ($V_{\mathrm{SL\{1:4\}}}$) of 0.9 V, which is the nominal operating voltage for this process. Each layer relies on triple wells to isolate nMOS transistors in each layer from body-bias effects. As described in Section II-B, when power consumption in all layers perfectly matches, voltage stacking evenly distributes internal voltage rails to 0.9 V per layer. However, as only V_{IN} is connected to a fixed external supply rail, the internal rails V_{UPP}, V_{MID}, and V_{LOW} fluctuate when there is interlayer activity mismatch between the cores occupying different layers, resulting in voltage noise. The test chip implements AFClk for the cores in each layer and integrates an IVR on the same die to reduce interlayer voltage noise. The test chip also integrates additional circuitry for testing and debug purposes: layer-shifting circuits for signaling between different SLs; scan chain to configure the digital blocks; and voltage monitoring circuitry to probe internal voltage rails and measure real-time voltage noise.

An annotated die photo of the test chip showing the floorplan of the cores and the IVR is presented in Fig. 2(b). A hierarchical design flow was used to construct the SL consisting of four cores and the clock generation and distribution logic. To isolate the entire layer from body-bias effects, a triple-well guard ring was created and placed around the SL. While the p-n junctions between the deep n-well and common GND substrate are reverse-biased at maximum 3.6 V, it is well below the breakdown voltage. The SLs were organized to facilitate power grid connection between the series-connected domains and with the IVR, and top two metal layers were reserved for power grid connections. Chip summary is presented in Table I.

A. Integrated Voltage Regulator

The test chip integrates a 4-to-1 SC IVR with the core array to neutralize voltage noise due to activity mismatch between the SLs, as shown in Fig. 2(a). The IVR implements a symmetric-ladder topology consisting of ten SC ladder units,

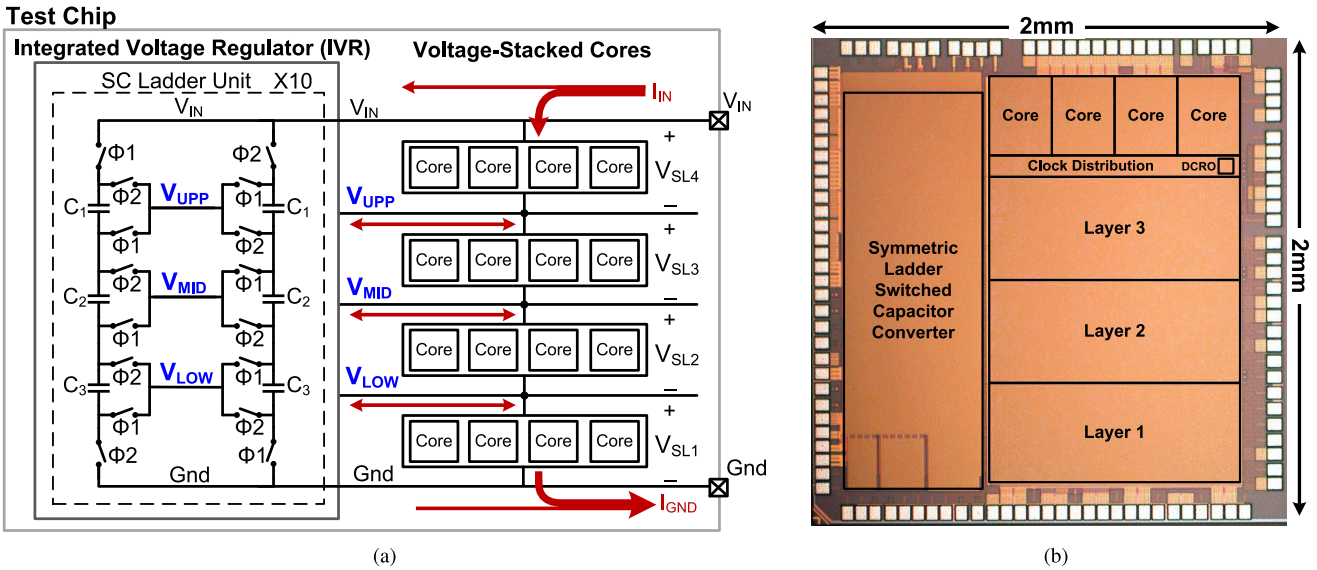


Fig. 2. (a) Block diagram of the voltage-stacked system showing the 16 four-way stacked cores and symmetric-ladder switched-capacitor IVR. (b) Annotated die photo.

TABLE I
CHIP SUMMARY

Process Technology	TSMC 40nm 1P10M
Die Size	2mm x 2mm
Core Array Area	1700um x 1200um
Core Operating Voltage	0.9V (Nominal)
Core F_{MAX}	250MHz @0.9V
Core Dynamic Power	5.65mW - 14.27mW (@0.9V)
DCRO Dynamic Power	2.31mW @0.9V
IVR Area	1675um x 495um
IVR Total Capacitance	4.5nF
IVR Power Density	21.1mW/mm^2 at 65% efficiency
Leakage Power	7.8mW per layer (@0.9V)

each controlled by one of ten phase-interleaved switching signals to reduce voltage ripple. The symmetric-ladder topology is a natural choice for voltage stacking, because in a sense, it too employs voltage stacking. By connecting to the internal rails V_{UPP}, V_{MID}, and V_{LOW}, the 4-to-1 symmetric-ladder converter becomes a multioutput regulator that pushes and pulls current to and from the stacked core array to smooth out imbalances and reduce voltage noise in the four stacked layers simultaneously. Because the IVR only neutralizes activity *mismatch* between the layers, the maximum power the IVR needs to provide is the power consumed by four cores in a single layer. It is important to note this is one quarter of the power a conventional single-layer, 16-core system would require from an IVR.

Powered from the external 3.6 V V_{IN}, the SC ladder unit operates with respect to two nonoverlapping clocks $\Phi 1$ and $\Phi 2$. The operation of the IVR is similar to other SC-based converters: the flying capacitors are charged in one phase and discharged in the other in symmetric fashion while current flows from the input to the output nodes through the capacitors and the power switches. Rather than regulating the internal rails to a specific voltage, the IVR implements a single-bound feedback control to keep each layer voltage above a prescribed reference voltage. Each layer implements a 2.5-GHz digital clocked comparator circuit that compares the layer voltages $V_{SL\{1:4\}}$ to a reference voltage V_{REF} generated on-chip for each layer. If any of the layer voltages fall below V_{REF}, the associated comparator generates a pulse signal that is processed by the feedback control logic to produce the clock signals $\Phi 1$ and $\Phi 2$ for each of the ten SC ladder units, which switch at a maximum frequency of 250 MHz. As a result, the IVR operates to keep the lower bound of each layer voltage at V_{REF} for all SLs.

The reference voltage V_{REF} presents a tradeoff between tolerable voltage droop and IVR loss. Higher V_{REF} reduces the voltage droop, improving V_{MIN}, but incurs larger IVR loss due to higher IVR activity, while IVR efficiency and power delivery capability may also suffer. On the other hand, if V_{REF} is too low, it can result in large interlayer voltage differences, as all SL voltages must add up to V_{IN}. Since the IVR only implements lower-bound control, there is a possibility that one or more layer voltages can exceed maximum rated voltage, causing reliability concerns. To resolve this issue, an upper bound can also be added to the control loop, such that the SC ladder units switch whenever the layer voltages deviate beyond the upper bound as well as the lower bound, and keep the layer voltages more evenly balanced.

The area and performance summary of the IVR is also presented in Table I. For consistency and to facilitate comparison to prior art, the IVR efficiency and power density are reported for the case of the IVR delivering power only to SL1. A more in-depth discussion of the IVR design and its implementation details can be found in [10] and [16]. This paper focuses on the system-level impact of the IVR on mitigating voltage noise, which is critical in a voltage-stacked multicore system.

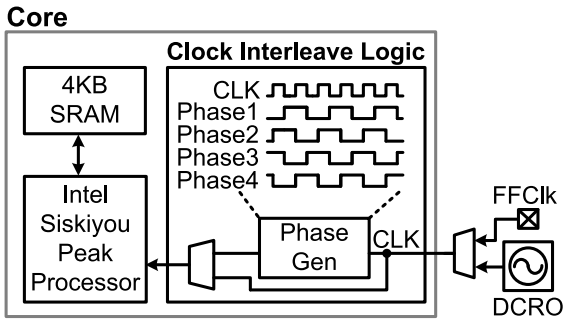


Fig. 3. Core block diagram.

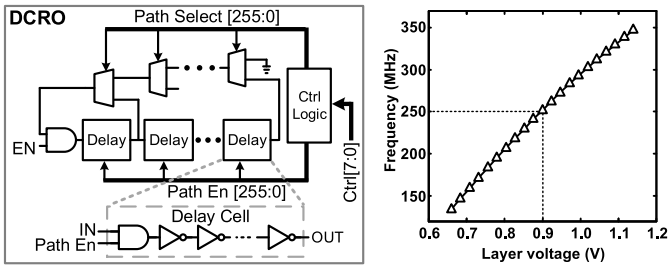


Fig. 4. Block diagram of the per-layer DCRO (left). Measured DCRO voltage versus frequency (right).

B. Core Design

Fig. 3 presents a block diagram of the *core* design. Each core contains an Intel Siskiyou Peak processor with 4-kB SRAM for both instruction and data. Our version of the Siskiyou Peak processor is a five-stage, single-issue, integer pipeline that implements a subset of the IA instruction set architecture and system software model. Each core can be independently programmed for multiprogram operation. Each core contains a clock interleave logic that allows each layer to implement clock-phase interleaving (ClkInt), which allows each of the four cores in each layer to operate off of 90° out-of-phase clock signals to smooth out high-frequency within-cycle voltage noise. To investigate the interaction between voltage noise and clocking schemes in voltage-stacked systems, the cores can operate in one of two clocking modes: 1) fully synchronous global fixed-frequency clocking (FFClk) or 2) per-layer AFClk. For FFClk, an external clock source drives all 16 cores. Per-layer AFClk utilizes a digitally configurable ring oscillator (DCRO) in each layer to generate the per-layer clock, whose frequency tracks the fluctuations in the SL voltage.

C. Digitally Configurable Ring-Oscillator Design

Prior works have demonstrated numerous AFClk schemes that utilize critical path tracking circuitry to change the clock frequency with voltage fluctuations [17]–[21]. The test chip implements a DCRO shown in Fig. 4, to act as a canary circuit that tracks the changes in critical path delay due to fluctuations in voltage. Many prior works have implemented similar designs, which have been shown to deliver good tracking accuracy [22]–[26]. The DCRO comprises a programmable array of delay cells configured via an 8-bit control code, which sets the length of the ring oscillator, with each control bit adding 125 ps to the clock period. During operation, the free-running ring oscillator essentially averages the effects of the layer voltage fluctuations to set the cycle-by-cycle clock period. The DCRO is configured during testing, such that the cores operate at the fastest DCRO setting without timing violations across the entire operating voltage range of 0.65–1.15 V for all cores in the layer. The DCRO setting also includes the necessary margins for potential high-frequency noise, local within-core voltage noise, and core-to-core variations. The frequency versus voltage relationship of the DCRO at the setting corresponding to the maximum core frequency is presented in Fig. 4. The DCRO runs at a nominal frequency of 250 MHz and consumes 2.31 mW at 0.9 V at this setting.

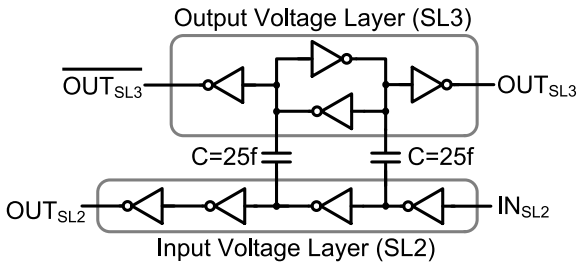


Fig. 5. Implementation of initializable capacitively coupled layer-shifter circuit. Schematic illustrates layer-shifter circuit conmmunicating data from SL2 to SL3.

D. Layer-Shifter Circuit

Voltage stacking requires efficient on-die signaling circuitry to enable communication between SLs that do not share common voltage reference planes. Direct communication between different SLs requires isolation to avoid current paths between stacked voltage rails. The test chip implements a capacitively coupled, initializable, layer-shifter circuit to communicate between different SLs. Fig. 5 presents the implementation of a layer-shifter circuit that shifts signals from SL2 to SL3 as an example. To enable domain crossing between any two SLs without breakdown concerns, the coupling capacitors are implemented using metal–oxide–metal (MoM) capacitors. The sizing of the coupling capacitors provides a tradeoff between power, reliability, area, and speed. Larger capacitors provide stronger coupling between layers, but also lead to larger area and power penalties. The insertion delay of the layer shifter is less than 100 ps, and the MoM capacitors are built using three lower metal layers M1 to M3. Similar circuitry was demonstrated in [1] using MOS capacitors.

IV. Measurement Results

This section provides the measurement results from the test chip fabricated in 40-nm CMOS. This section begins with a brief discussion of the various workloads that were characterized to run on the cores and used for all of the subsequent measurements. Sections IV-B–IV-F analyze voltage noise in the 4 × 4 voltage-stacked test chip and

TABLE II
INSTRUCTION POWER CHARACTERIZATION

Instruction	Description	Power per core (@250MHz/0.9V)
IMUL	Signed multiply	14.27 mW
ADC	Add with carry	10.02 mW
XCHG	Exchange data	5.65 mW

evaluate the noise mitigation offered by the IVR and adaptive clocking. Sections IV-B–VI-E present measurement results in the context of high-throughput multicore system assuming all 16 cores are active running highly parallel workloads. Section IV-B characterizes the inherent interlayer voltage noise for balanced and unbalanced workload scenarios assuming a single clock domain for all 16 cores and the IVR disabled. Section IV-C then evaluates the noise mitigation provided by turning ON the IVR, which reduces worst case voltage droop. To further improve efficiency, Section IV-D explores the benefits of AFClk. Section IV-E then takes a step back to understand improvements in system-wide power delivery efficiency made possible by voltage stacking. Finally, Section IV-F considers the impact of wider workload diversity due to core inactivity and the impact of workload allocation strategies in a multicore voltage-stacked system.

A. Stressmark Generation

As discussed in Section II, interlayer voltage noise in voltage-stacked systems is a direct result of activity mismatch between the SLs. The magnitude of voltage noise ultimately depends on: 1) workload behavior; 2) the instruction set architecture (ISA) and microarchitecture of the cores; and 3) the underlying process technology that determine the leakage and dynamic power consumption of the cores in the system. Therefore, to profile the worst case voltage noise of the test chip, the first step is to generate the power usage profile of the instructions in the ISA to find the maximum and minimum power instructions, and determine the worst case nominal power difference that can be generated by active cores. This is done by generating a power-profiling microbenchmark for each instruction that runs an endless loop of a 4-kB code block that runs each instruction in the ISA continuously with no dependencies between instructions. Table II presents the measured dynamic power consumption per core for the maximum (IMUL), minimum (XCHG), and medium (ADC) power instructions of the ISA. It should be noted that the power usage changes depending on the operand type for each instruction. Therefore, different operand types were also profiled to find the maximum and minimum power instructions.

Using this instruction power profile, we generate several microbenchmarks to stress the voltage stack. The microbenchmarks used for instruction profiling are utilized to generate static activity differences between the cores. We also generate several simple microbenchmarks that generate oscillatory activity behavior by periodically switching between two different instructions. The microbenchmarks used for testing are summarized in Table III.

TABLE III
MICROBENCHMARKS USED FOR NOISE PROFILING

Micro-benchmark	Description
MAX-CONT	IMUL instructions with random inputs in endless loop
MIN-CONT	XCHG instructions with random inputs in endless loop
MAX-MIN-*n*	*n* IMUL instructions followed by *n* XCHG instructions in endless loop
MID-MIN-*n*	*n* ADC instructions followed by *n* XCHG instructions in endless loop

B. Inherent Voltage Noise From Voltage Stacking

To demonstrate the inherent properties of voltage stacking in the context of high-throughput multicore systems, we first characterize voltage noise with all 16 cores active, running synchronously off of an FFClk, and with the IVR and ClkInt turned OFF. Fig. 6 presents snapshots of measured V_{SL2} for various workload scenarios as well as per-layer box plots (which delineates a distribution's minimum, first quartile, median, third quartile, and maximum values) of all four layer voltages. The box plots show voltage distribution measured over a 1-ms execution window. To ensure all cores operate correctly with no timing errors, the worst case voltage noise, the corresponding minimum SL voltage V_{MIN}, and maximum operating frequency F_{MAX} were characterized for all cores. For all test results in this section, cores were run at the worst case $F_{MAX} = 130$ MHz that guaranteed correct functionality under worst case voltage noise.

Fig. 6 compares the resulting voltage noise between a balanced workload scenario and multiple unbalanced scenarios. In the balanced scenario [Fig. 6(a)], all cores ran the *MIN-CONT* microbenchmark in lockstep. The V_{SL2} waveform and the boxplots show that all layers evenly subdivide to 0.9 V, as expected, since all layers have identical activity. Fig. 6(b)–(e) presents results for four unbalanced workload scenarios with the four cores in SL2 running a higher power microbenchmark while all other cores run *MIN-CONT*, demonstrating significant voltage droop in V_{SL2}. Comparison of Fig. 6(b) and (c) shows worse voltage noise for larger interlayer activity mismatch. Scenarios in Fig. 6(c)–(e) present noise profiles for worst possible activity mismatch (*MAX* versus *MIN* instructions), but with varying periodicity. Fig. 6(c)–(e) shows comparable noise magnitude for the three scenarios, while longer periods of activity mismatch results in longer periods of voltage droop akin to IR drop. As discussed in Section II, the scenarios shown in Fig. 6(c)–(e) all correspond to worst case voltage noise when all 16 cores are active, as they generate: 1) the maximum possible activity mismatch between SLs and 2) only a single layer has higher activity.

As all layer voltages must add up to V_{IN} in voltage stacking, all layer voltages are inherently interdependent. Box plots of Fig. 6(b)–(e) show that voltage droop in V_{SL2} results in increase in other layer voltages. To present a complete picture of how activity mismatch affects the layer voltages, Fig. 7 presents the waveforms of V_{SL2} for a variety of other activity

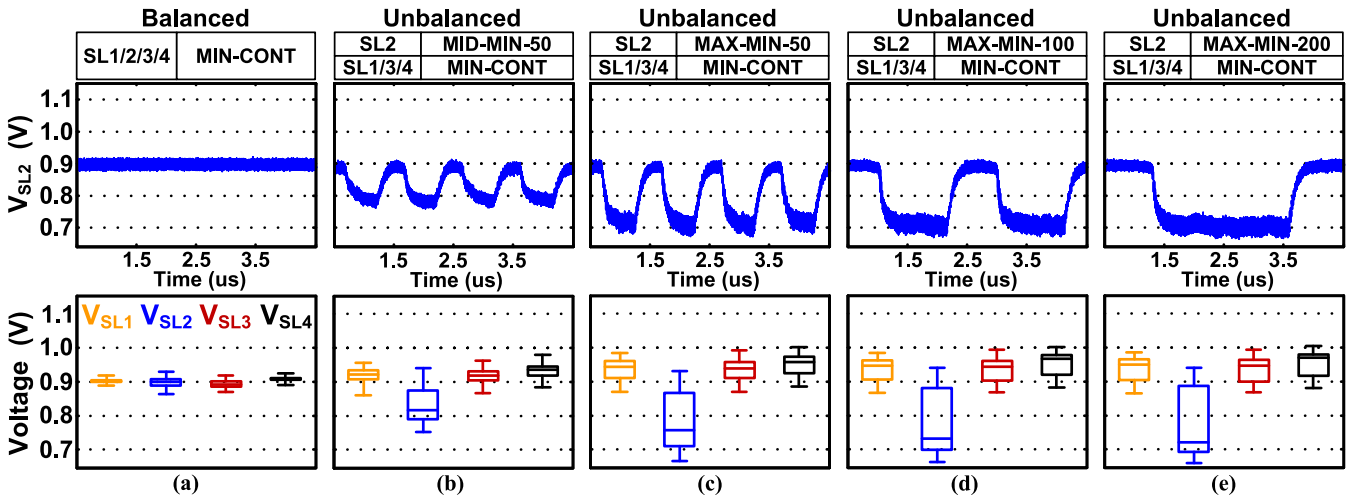


Fig. 6. Measured transient waveforms of V_{SL2} and boxplots of V_{SL1}–V_{SL4} for balanced and unbalanced workload scenarios. (a) Balanced: all cores running MIN-CONT microbenchmark. (b) Unbalanced: cores in SL2 running MID-MIN-50. (c) Unbalanced: cores in SL2 running MAX-MIN-50. (d) Unbalanced: cores in SL2 running MAX-MIN-100. (e) Unbalanced: cores in SL2 running MAX-MIN-200.

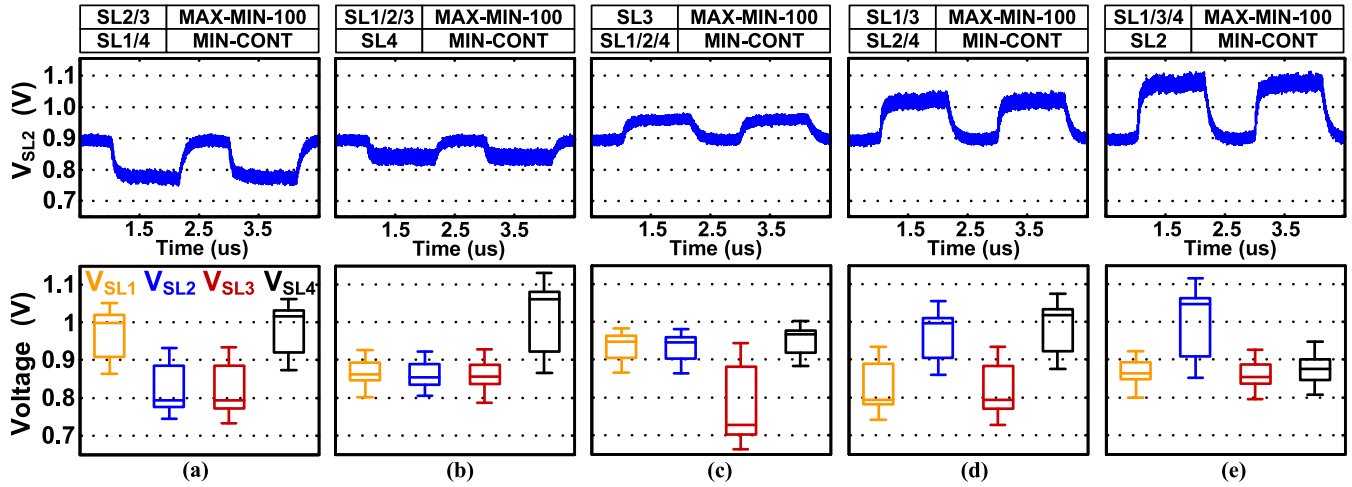


Fig. 7. Measured transient waveforms of V_{SL2} and boxplots of V_{SL1}–V_{SL4} for various unbalanced workload scenarios. (a) Cores in SL2/SL3 running MAX-MIN-100. (b) Cores in SL1/SL2/SL3 running MAX-MIN-100. (c) Cores in SL3 running MAX-MIN-100. (d) Cores in SL1/SL3 running MAX-MIN-100. (e) Cores in SL1/SL3/SL4 running MAX-MIN-100.

mismatch conditions. Box plots of all four SL voltages are also presented for completeness. As discussed previously, a comparison of Fig. 7(a) and (b) shows that voltage droop reduces when more layers execute higher activity workloads. Other scenarios, shown in Fig. 7(c)–(e), lead to increases in V_{SL2} due to higher activity workloads running on cores in other layers. We can infer the general behavior of voltage noise in all layers from the measured noise profiles presented in Figs. 6 and 7.

Since the voltage across a capacitor cannot change instantaneously, the rate of voltage transition due to interlayer activity mismatch for any layer depends on the time it takes to discharge the capacitance of the SLs, which includes the intrinsic capacitance of the digital logic, decoupling capacitance, and the nonswitching flying capacitance of the IVR. The capacitance is discharged through the power grid and the switching digital logic, which represents a resistive path to ground. Therefore, the transition time for any layer voltage is dictated by the *RC* time constant due to the capacitance of the SLs and the resistive discharge path to ground. Typically, inductance of on-chip power grid is small enough that inductive effects are negligible for interlayer noise in voltage stacking. Fig. 6 shows that in the case of the test chip presented in this paper, it takes ~300 ns for the voltage to transition from 0.9 V to the worst case minimum voltage shown in Fig. 6(c)–(e). Adding more decoupling capacitance to each layer will decrease the rate of voltage transition, but for static activity mismatch, it ultimately cannot alleviate voltage droop, necessitating the need for active noise mitigation techniques.

C. Using an IVR to Mitigate Voltage Noise

The worst case voltage noise for the high-throughput workload scenario observed in Fig. 6(c)–(e) is especially problematic for global fixed-frequency operation (FFClk), because the global clock frequency must be chosen to ensure

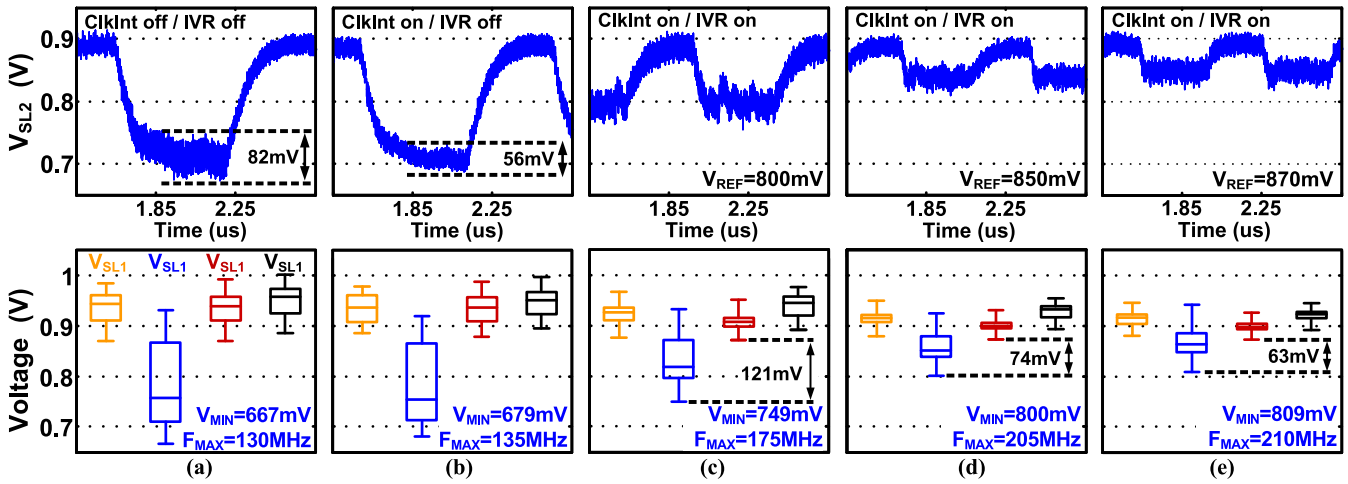


Fig. 8. Measured transient waveforms of V_{SL2} and boxplots of V_{SL1}–V_{SL4} for unbalanced workload scenario running FFClk at various test chip settings. (a) ClkInt OFF, IVR OFF. (b) ClkInt ON, IVR OFF. (c) ClkInt ON, IVR ON (V_{REF} = 800 mV). (d) ClkInt ON, IVR ON (V_{REF} = 850 mV). (e) ClkInt ON, IVR ON (V_{REF} = 870 mV).

correct operation for all cores at the minimum operating voltage (V_{MIN}) under worst case voltage noise conditions, requiring large voltage margins. Moreover, there are additional constraints on V_{MIN}. For example, it must be high enough to ensure reliable operation of voltage-sensitive circuits such as on-chip SRAMs, which can exhibit SRAM read/write failures under low voltage conditions. Consequently, the test chip includes an IVR to provide certain V_{MIN} guarantees.

To demonstrate the IVR's ability to reduce interlayer voltage noise, let us begin with the worst case noise scenario previously seen in Fig. 6(c). A close-up of V_{SL2}, in Fig. 8(a), shows the minimum voltage observed over a 1-ms execution window is 667 mV. Based on memory BIST tests, all memories were functional down to 650 mV and, therefore, the test scenario avoids issues due to SRAM instability. However, core performance degradation is a concern. To accommodate the worst case voltage noise, voltage margin in excess of 230 mV is required, limiting the maximum operating frequency (F_{MAX}) to only 130 MHz. Fig. 8(b) first demonstrates the benefits of ClkInt. Although within-cycle current smoothing reduces voltage ripple, the minimum voltage is still less than 700 mV. In contrast, also turning ON the IVR provides much more pronounced benefits, as shown in Fig. 8(c)–(e) for different V_{REF} settings. These results confirm that the IVR can maintain V_{SL2} around V_{REF} and provide V_{MIN} guarantees, which then allows the cores to operate at a higher F_{MAX}. However, it is important to note the IVR cannot regulate the layer voltages to exactly V_{NOM}. The IVR's ability to deliver power diminishes as V_{REF} setting approaches V_{NOM}, resulting in diminishing returns in V_{MIN} improvements when V_{REF} is set too aggressively. Furthermore, Fig. 8(c)–(e) shows that cores in SL1, SL3, and SL4 must operate at a higher power for the same performance level, due to the extra voltage margin caused by the elevated SL voltages, which degrades the overall system energy efficiency.

To quantify the benefits gained from ClkInt and turning ON the IVR, Fig. 9 presents the aggregate throughput, energy, and EDP of the test chip settings shown in Fig. 8, normalized to

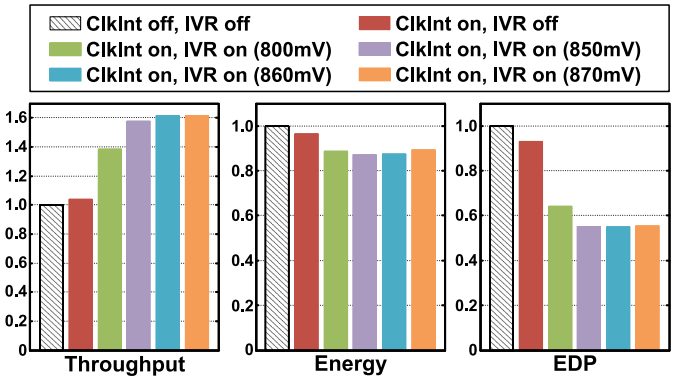


Fig. 9. Normalized system throughput, energy, and EDP of unbalanced testing scenarios running FFClk under various test chip settings. IVR V_{REF} is noted in parenthesis for the cases when IVR is ON.

the baseline case when none of the features are turned ON [shown in Fig. 8(a)]. The results show that the V_{MIN} improvements gained from turning ON the IVR provide average throughput enhancements, which then translate to large EDP reductions. Higher V_{MIN} also reduces voltage margins to provide energy improvements despite incurring IVR energy overheads. However, aggressively increasing V_{REF} beyond 860 mV exacerbates IVR losses without improving V_{MIN} or throughput.

D. Adaptive Frequency Clocking

To further reduce voltage margin and thereby further improve system performance and energy efficiency, the test chip implements AFClk. The clock frequency of the DCRO in each layer tracks SL voltage fluctuations to independently minimize voltage margins per layer. Consequently, AFClk does not penalize SLs for voltage noise in other layers. Fig. 10 presents the voltage waveform of V_{SL2}, box plots of all layer voltages, and the per-layer DCRO clock frequencies for the same unbalanced scenario presented before in Fig. 8 with

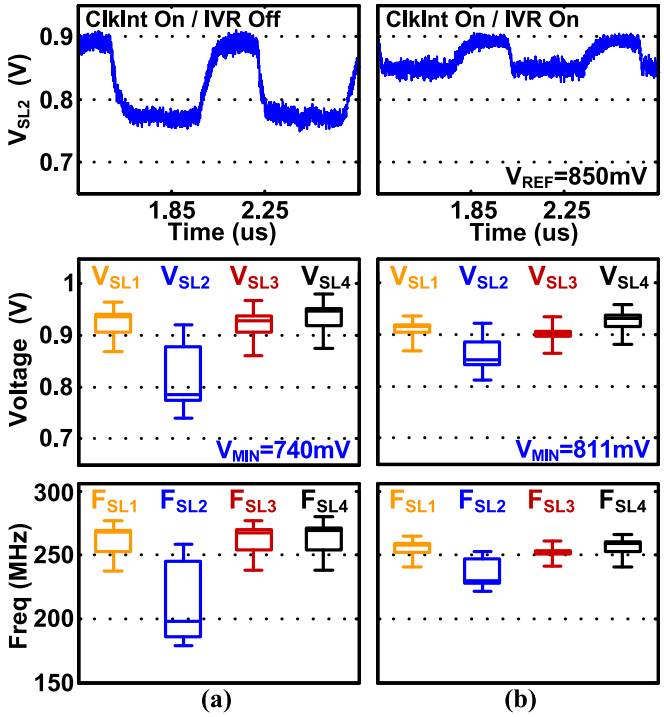


Fig. 10. Measured transient waveforms of V_{SL2} (top), box plots of V_{SL1}–V_{SL4} (middle), and box plots of per-layer DCRO frequency distribution (bottom) for unbalanced workload scenario running AFClk at two different test chip settings. (a) ClkInt ON, IVR OFF. (b) ClkInt ON, IVR ON (V_{REF} = 850 mV).

ClkInt turned ON. With the IVR OFF [Fig. 10(a)], AFClk results in smaller voltage droop (higher V_{MIN}) compared with FFClk operation. This is because per-layer dynamic current consumption has a stronger dependence on voltage compared with FFClk, since AFClk frequency depends on voltage. Moreover, the spread in clock frequencies, seen in Fig. 10(a), demonstrates that rather than being limited to the worst case frequency set by noise in V_{SL2}, cores in SL1, SL3, and SL4 ran at higher frequencies set by the DCRO in each layer operating at higher voltages, thereby improving overall system performance and energy efficiency. Fig. 10(b) demonstrates AFClk operation with the IVR turned ON and V_{REF} set at 850 mV. The results show that AFClk allows the cores to operate with minimal voltage margin, while the IVR maintains V_{MIN} above 810 mV. This provides a minimum performance guarantee for all layers and protection against potential reliability issues (e.g., SRAM instability).

Putting the measured results together, Fig. 11 presents the resulting aggregate throughput, energy, and EDP for AFClk. The results are normalized to that of FFClk operation with the IVR ON and V_{REF} set to 860 mV, shown in Fig. 9, which is the highest performance and energy efficiency V_{REF} setting for FFClk. Since cores in each layer operate at different voltages/frequencies due to voltage noise for AFClk, we calculate the system energy as the sum of total energy consumed by all cores to complete the same amount of work

$$E = \sum_{i=\mathrm{SL1}}^{\mathrm{SL4}} P_i t_i + P_{\mathrm{IVR}} t_{\mathrm{SL2}} \tag{2}$$

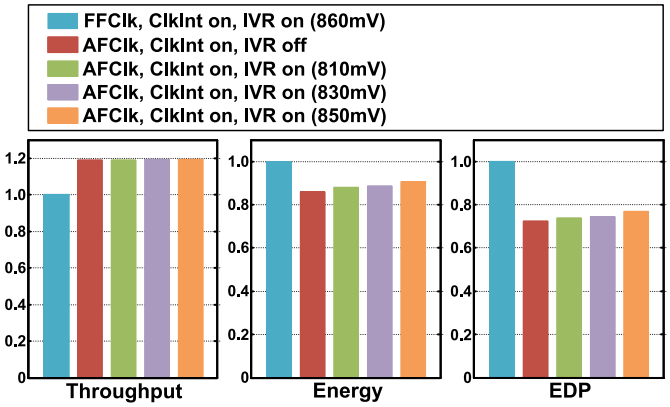


Fig. 11. Normalized aggregate throughput, energy, and EDP comparison of AFClk operation (AFClk) versus fixed-frequency operation (FFClk) for various test chip settings under worst case noise condition. IVR V_{REF} is noted in parenthesis for the cases when IVR is ON.

where $P_{\mathrm{SL}\{1:4\}}$ is the power consumption of the entire layer that includes all four cores and the DCRO. IVR energy overhead is calculated by multiplying IVR power loss (P_{IVR}) with the time it takes for the cores in SL2 to complete the task, because the IVR is supplying extra current only to the cores in SL2 in this workload scenario. EDP is also calculated in a similar fashion as

$$\mathrm{EDP} = \sum_{i=\mathrm{SL1}}^{\mathrm{SL4}} P_i t_i^2 + P_{\mathrm{IVR}} t_{\mathrm{SL2}}^2. \tag{3}$$

The results show that even when compared against the best-case FFClk scenario, AFClk provides significant improvements in throughput, energy, and EDP due to reduction in voltage margin for all cores. Notice the aggregate throughput of the system for AFClk stay relatively constant independent of whether the IVR is ON or not. This is because in the high-throughput scenario with all 16 cores active, loss of performance (frequency) in any layer is compensated by higher frequency in other layers. Overall, AFClk achieves reduction in energy and EDP compared with the best-case FFClk, even when accounting for extra power consumed by the DCRO and the IVR energy overhead. Setting higher IVR V_{REF} for higher V_{MIN} translates to increases in energy and EDP, as expected, due to higher IVR energy overhead.

The results shown in Figs. 10 and 11 clearly demonstrate the advantages of AFClk over FFClk for the voltage-stacked system presented in this paper. For all subsequent test results, we use AFClk as the default clocking mode of operation to achieve the best possible performance and efficiency for the system.

E. System-Wide Power Delivery Efficiency

As shown in the energy comparisons presented previously, voltage regulation using the IVR comes with an energy overhead. Typically, IVR losses are quantified by the IVR efficiency. For conventional systems in which the IVR provides the entirety of the power consumed by the system, IVR efficiency represents the overall on-chip power delivery efficiency. However, voltage-stacked systems are fundamentally

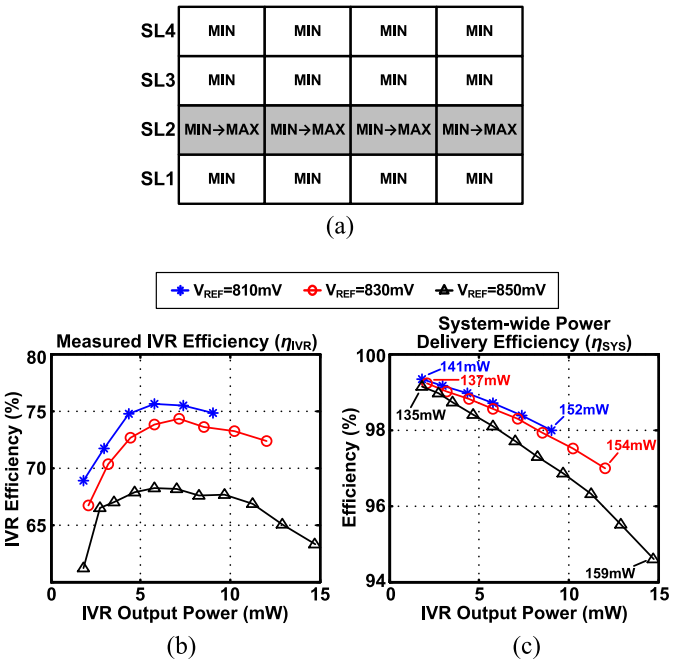


Fig. 12. (a) Block diagram illustrating the test setup to characterize IVR and system-wide power delivery efficiency. (b) Plot of measured IVR efficiency when delivering power to SL2 for three V_{REF} levels. (c) Corresponding system-wide power delivery efficiency measurements with annotations of the total power consumed by the test chip for the first and the last values of each V_{REF}.

different in that they require the IVR to only supply the load current necessary to compensate for interlayer activity mismatch. Therefore, IVR losses only apply to the mismatch-related power, while stack current that is common to all layers, such as leakage, is recycled efficiently through the stack. For this reason, IVR efficiency does not sufficiently capture the overall power delivery efficiency in voltage stacking.

To properly quantify power delivery efficiency in voltage-stacked systems, we define the metric *system-wide power delivery efficiency* (η_{SYS}) as

$$\eta_{SYS} = \frac{\text{Power consumed by all cores}}{\text{Total power delivered to test chip}}.$$

To illustrate this important attribute, Fig. 12 presents the measured IVR efficiency (η_{IVR}) and the system-wide power delivery efficiency (η_{SYS}) for an experiment running the test chip in high-throughput mode with all 16-cores active. Fig. 12(a) shows the experimental setup. With all cores in SL1, SL3, and SL4 executing the *MIN-CONT* microbenchmark, power consumption of the four cores in SL2 was swept from minimum active power to maximum by executing total of 20 ISA characterization microbenchmarks used for instruction power profiling, from *MIN-CONT* (minimum power) to *MAX-CONT* (maximum power). Running progressively higher power microbenchmarks in SL2 leads to larger voltage droop in V_{SL2} that requires more IVR activity and output power to keep V_{SL2} at V_{REF}, once V_{SL2} droops below the prescribed V_{REF}. Fig. 12(b) plots the η_{IVR} of the IVR versus the total power supplied by the IVR to the cores in SL2, for three IVR V_{REF} settings. The results demonstrate the tradeoff associated

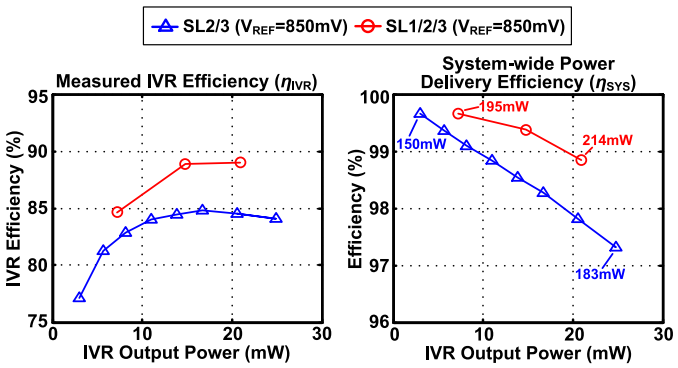


Fig. 13. Measured IVR efficiency (left) and corresponding system-wide power delivery efficiency (right) for two workload scenarios with V_{REF} set to 850 mV.

with V_{REF}. Higher V_{REF} translates to smaller voltage droop but also lower η_{IVR}, because the IVR must work harder to drive the layer voltage to higher levels.

Fig. 12(c) presents the corresponding η_{SYS}, again plotted versus the IVR output power for the three V_{REF} settings. The total power consumption of the test chip is annotated for the maximum and minimum efficiency points for each V_{REF} setting. With all 16 cores active, the current consumed by the cores in SL4, SL3, and SL1 is supplied from off-chip and efficiently recycled through the SLs, while also contributing to the current consumption of the cores in SL2. Therefore, the IVR only supplies the additional power necessary to support the higher activity of the cores in SL2 at V_{REF}. Thus, IVR losses only apply to a small fraction of total power consumed by all of the cores. For example, Fig. 12(c) shows that with V_{REF} set to 850 mV, the IVR provides only 15 mW of the total 159 mW consumed by the test chip even for the worst case activity mismatch, while rest of the power is recycled through the SLs. As a result, η_{SYS} is greater than 94%, even though η_{IVR} is limited to less than 65%, demonstrating that voltage stacking effectively masks the IVR inefficiency and achieves high system-wide power delivery efficiency.

Recall from Section II, spreading out higher activity workloads across multiple layers leads to less interlayer voltage droop than concentrating all of the higher activity to a single layer. Moreover, the symmetric-ladder SC IVR implemented on the test chip is more efficient at supplying power to multiple layers simultaneously than to a single layer. Due to its topology, delivering power to multiple layers reduces the total amount of charge flowing through the flying capacitors and power switches, reducing the losses [16]. To demonstrate this aspect of the system, Fig. 13 plots the measured IVR efficiency and system-wide power delivery efficiency when the IVR delivers power to multiple layers simultaneously. This experiment is similar to the prior experiment run for Fig. 12, but with the power swept from minimum to maximum in two layers (SL2 and SL3) and three layers (SL1, SL2, and SL3). IVR V_{REF} is set to 850 mV for both cases. Overall, both η_{IVR} and η_{SYS} improve when the IVR delivers power to multiple layers, even when the total test chip power is higher than the conditions corresponding to the results in Fig. 12.

TABLE IV
WORKLOAD SCENARIOS FOR EFFICIENCY TESTS

Workload Scenario	Description
RAND-LYR	Random kernel allocated per layer
RAND-CORE	Random kernel allocated per core
BALANCED	All cores running the same kernel

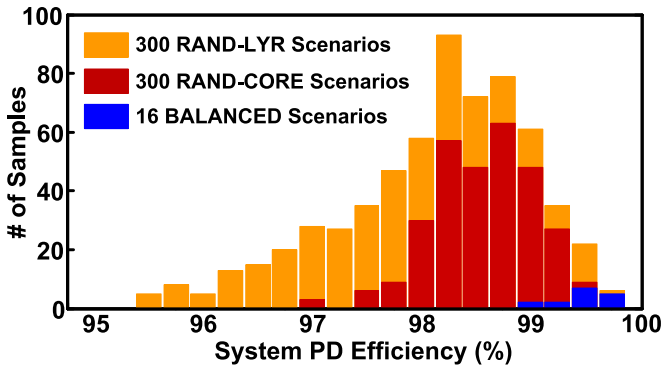


Fig. 14. Histogram of system-wide power delivery efficiency for 616 workload scenarios described in Table IV.

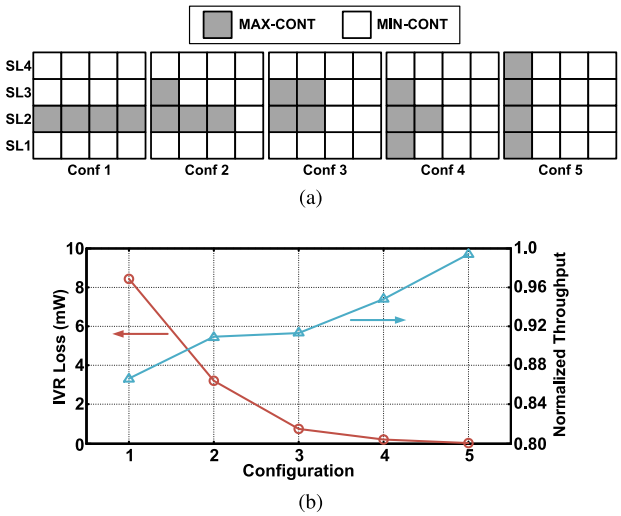


Fig. 15. (a) Five possible workload allocation scenarios with the same set of workloads. (b) Normalized average per-core throughput for the four cores running MAX-CONT and IVR loss.

To further expand the experiment beyond the worst case scenario and demonstrate the efficiency of the system running real kernels, we chose 16 kernels to run on the cores from the *Machsuite* benchmark suite [27]. These kernels cover a broad range of applications, including matrix multiplication, sort, string matching, AES encryption, molecular dynamics, and Viterbi algorithm, and offer a diverse mixture of workload behaviors (compute-intensive, memory-intensive, and so on). To analyze the effects of different workload behaviors on efficiency, we generate three different groups of workload scenarios using the 16 kernels to run on the system, as summarized in Table IV. With the test chip running in high-throughput mode with all 16 cores active, Fig. 14 plots the resulting histogram of system-wide power delivery efficiency for the three workload groups, which is greater than 95% for all workload scenarios. *RAND-LYR* scenarios exhibit the worst efficiency, mainly because they have the worst interlayer activity mismatch. When each of the 16 cores run one of 16 randomly chosen kernels (*RAND-CORE*), there is an averaging effect on the per-layer core activities, which ends up reducing overall voltage noise and IVR losses. Finally, with the *BALANCED* scenarios, efficiency is 99% or better, because the IVR rarely turns ON.

F. Workload and Core Allocation

The test results so far have demonstrated the IVR and AFClk to be effective hardware solutions to improve system throughput and energy efficiency and to provide the necessary minimum voltage guarantees required by the voltage-stacked system. We now take a step back to explore a software solution that relies on intelligent interlayer workload balancing to further improve overall efficiency. This section explores the impact of workload and core allocation in multicore voltage-stacked systems.

Recall the worst case workload scenario presented in Fig. 12, where all cores in SL2 execute the *MAX-CONT* microbenchmark while all other cores run *MIN-CONT*. Fig. 15(a) presents five different workload allocation scenarios that are possible for the same set of workloads running on the 16 cores. The workload allocation becomes progressively more balanced going from configuration 1 to 5. To show the benefits of a more balanced workload allocation, Fig. 15(b) plots the IVR loss for all configurations as well as the normalized average throughput of the four cores that are running *MAX-CONT*. Measurement results show that better balancing of workloads across the layers leads to lower voltage noise, lower IVR loss, and higher throughput.

In addition to workload differences, SLs may also suffer imbalances due to coarser-grained core inactivity. For instance, if a workload scenario calls for only four cores to be active with all other cores powered down, allocating the workload to the four cores in a single layer will limit the system-wide power delivery efficiency to the IVR efficiency, severely degrading the overall system energy efficiency. Therefore, further expanding on the concept of intelligent workload allocation, we deduce that for any number of active cores, there is an optimum allocation that evenly distributes work to minimize interlayer activity mismatch and maximize efficient current recycling through the stack. Fig. 16(a) shows the optimum order that work should be allocated across the 16 cores. For example, only the five cores labeled 1–5 will be active for configuration 5, as shown in Fig. 16(b), while the rest are inactive. For configuration 6, Core B-SL2 (labeled 6) will be activated in addition to the cores in configuration 5. As more cores are turned ON, the principle is to activate cores vertically along a column rather than horizontally along the same layer. For voltage stacking, this achieves the most balanced core allocation possible for any number of active cores. Note that for each column, the cores in SL3 and SL2 are activated before turning ON the cores in SL1 and SL4. This is to take advantage of the fact that the IVR is more efficient in delivering power to the middle two layers [10], [16]. Allocating cores this way ensures that the maximum activity

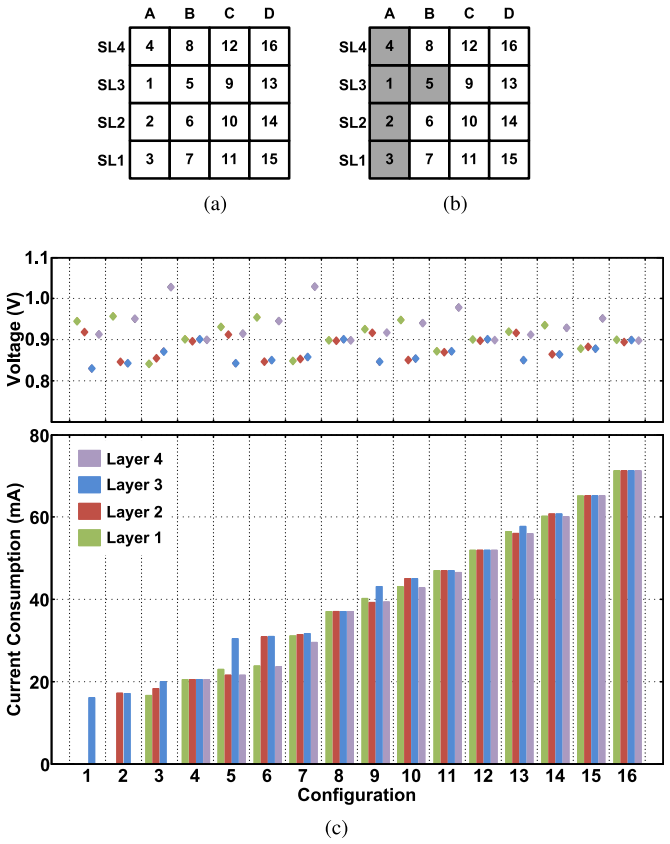


Fig. 16. (a) Block diagram illustrating the order in which the cores are activated when not all 16 cores are active, and (b) example of the active cores (shaded) for configuration 5 when only five cores are active and all other cores are powered down. (c) Average per-layer operating voltage and current consumption for each configuration.

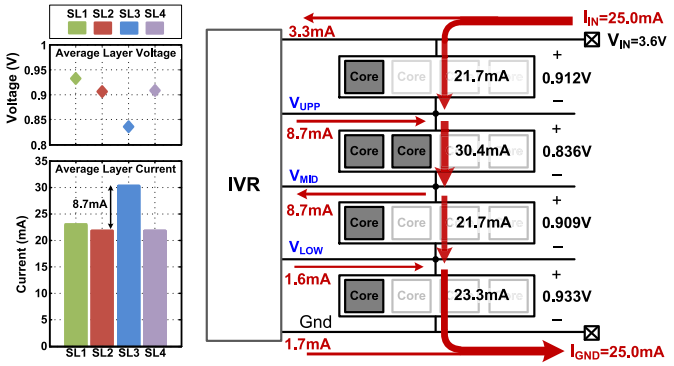


Fig. 17. Block diagram illustrating the current flow through the test chip for configuration 5 in Fig. 16.

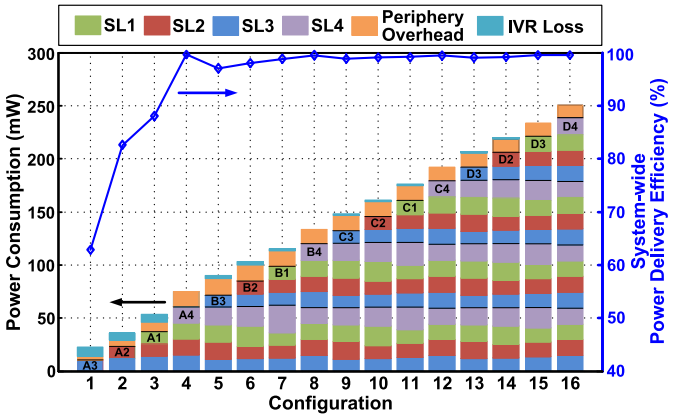


Fig. 18. Power consumption of the test chip and the corresponding system-wide power delivery efficiency for the 16 different core activation configurations. For each configuration, the additional core that is activated in addition to the previous configuration is marked with the coordinates as described in Fig. 16(a).

mismatch between any layers does not exceed the activity of a single core, preemptively capping the power that the IVR has to process.

We can now run an experiment to understand the effects of coarse-grained core activity/inactivity and workload allocation. First, all inactive cores are fully powered OFF, so that leakage current consumption of these cores is negligible. Second, all cores execute the *MAX-CONT* microbenchmark. Fig. 16(c) presents the average operating voltage and current consumption for each SL for all 16 configurations. Fig. 17 presents an illustration of the current flow through the SLs for configuration 5 presented in Fig. 16. Average per-layer voltage and current are also shown for clarity. Given the extra load in SL3, V_{SL3} droop below V_{REF} and the IVR maintains the minimum V_{SL3} close to V_{REF} (set to 850 mV). The average current consumption of 21.7 mA common to all SLs is recycled through the stack, while the IVR only provides the additional 8.7 mA of current that is required to support the two cores in SL3, and the additional 1.6 mA of current that is required to support the core in SL1 operating at an elevated voltage. With the IVR delivering additional current to support the extra load in SL3, the distribution of V_{SL1}, V_{SL2}, and V_{SL4} is determined by the static behavior of the IVR, which dictates that different amounts of charge must flow through the flying capacitors of the symmetric ladder, resulting in higher V_{SL1} than V_{SL2} and V_{SL4}.

The average per-layer voltage and current presented in Fig. 16(c) provide insight into the amount of power that the IVR has to process for different numbers of active cores under this core allocation scheme. For configurations 1–3, where one or more layers are powered OFF, no current can be recycled through the stack. Therefore, the IVR must supply all of the power consumed by the active cores. Beyond configuration 4, however, when at least a full column of cores is active, current consumption common to all layers is recycled through the stack while the IVR only supplies the additional current required to support the higher activity due to additional active cores in the SLs. Since all cores are running the same microbenchmark in this experiment, configurations 4, 8, 12, and 16 are balanced scenarios and the layer voltages evenly subdivide V_{IN} to 0.9 V. A comparison of results for configurations 1, 5, 9, and 13 in Fig. 16(c) confirms the earlier observation in Section II that higher overall current flow through the stack alleviates voltage noise for the same interlayer activity mismatch.

Fig. 18 presents a stacked bar chart of the power consumption for each configuration as well as the corresponding system-wide power delivery efficiency. IVR loss and periphery overhead, which includes the DCRO, clock distribution, and all other uncore power, such as layer-shifter power,

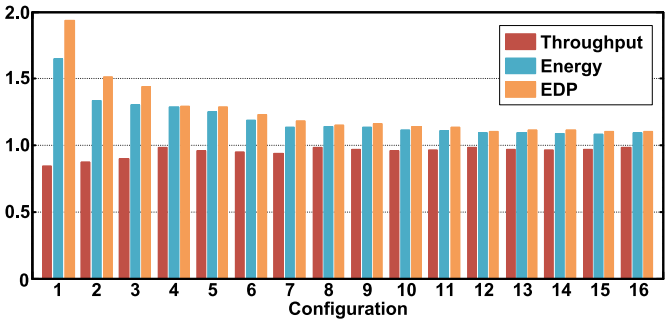


Fig. 19. Normalized average per-core throughput, energy, and EDP for the 16 different core activation configurations.

are shown separately. Fig. 18 provides a measure of the IVR losses incurred by the system in relation to the overall power delivered. For configurations 1–3, where the IVR supplies all of the power consumed by the cores, the IVR loss is substantial, and the system-wide power delivery efficiency equals the IVR efficiency. As more cores are activated, however, interlayer charge recycling increases and IVR losses reduce. Overall system-wide power delivery efficiency exceeds 95% for all configurations when four or more cores are active.

Finally, Fig. 19 presents the average per-core throughput, energy, and EDP for all 16 configurations. The results are normalized to an ideal baseline scenario where all layer voltages are losslessly driven to V_{NOM}. For one to three active cores, energy and EDP penalties are high due to large IVR loss and DCRO overhead and margins, coupled with the loss of performance. As more cores turn ON, the overall energy and EDP of the system improves due to decreasing IVR loss, amortized per-core DCRO overhead, and more balanced voltage distribution, leading to smaller loss in throughput. Note that due to the additional margin required to account for variations and nonideal voltage tracking capabilities of the DCRO, the normalized throughput of the system is less than one even for the balanced configurations. The energy and EDP includes the losses related to this margin.

V. Conclusion

Voltage stacking alleviates off-chip issues that could hinder efficient power delivery for future computing systems, while offering the potential for high efficiency system-level power delivery via charge recycling. However, voltage stacking suffers from internal voltage noise due to interlayer activity mismatch. This paper demonstrates a test chip that implements a 16-core four-way voltage-stacked system that integrates industry grade microprocessor with a symmetric-ladder SC IVR and AFClk to address the critical issue of interlayer voltage noise. The adaptive frequency operation in conjunction with the IVR achieves efficient voltage noise mitigation on the four SLs and measurement results demonstrate significant improvements in throughput and energy efficiency of the system. In high-throughput systems, the IVR only has to process a small fraction of the overall power while majority of the power is efficiently recycled through the stack. Measurement results show that the system-wide power delivery efficiency is higher than 94% even for the worst case noise condition in the context of high-throughput workload scenarios. In addition to the hardware techniques implemented on the test chip, this paper shows that intelligently allocating and scheduling workload at the software-level results in even higher system efficiency. By exploiting the synergy between intelligent software-level workload management and robust hardware solutions that provide the necessary fail-safe measures to voltage noise, voltage stacking offers a promising power delivery solution for future multicore systems.

Acknowledgment

The authors would like to thank TSMCs University Shuttle Program for chip fabrication and Intel Corporation for Siskiyou Peak IP.

References

[1] S. Rajapandian, K. L. Shepard, P. Hazucha, and T. Karnik, "High-voltage power delivery through charge recycling," *IEEE J. Solid-State Circuits*, vol. 41, no. 6, pp. 1400–1410, Jun. 2006.

[2] P. Jain, T.-H. Kim, J. Keane, and C. H. Kim, "A multi-story power delivery technique for 3D integrated circuits," in *Proc. Int. Symp. Low Power Electron. Design*, Aug. 2008, pp. 57–62.

[3] L. Chang, R. K. Montoye, B. L. Ji, A. J. Weger, K. G. Stawiasz, and R. H. Dennard, "A fully-integrated switched-capacitor 2:1 voltage converter with regulation capability and 90% efficiency at 2.3A/mm^2," in *Proc. Symp. VLSI Circuits*, Jun. 2010, pp. 55–56.

[4] S. K. Lee, D. Brooks, and G.-Y. Wei, "Evaluation of voltage stacking for near-threshold multicore computing," in *Proc. ACM/IEEE Int. Symp. Low Power Electron. Design*, 2012, pp. 373–378.

[5] K. Kesarwani, C. Schaef, C. R. Sullivan, and J. T. Stauth, "A multi-level ladder converter supporting vertically-stacked digital voltage domains," in *Proc. 28th Annu. IEEE Appl. Power Electron. Conf. Expo. (APEC)*, Mar. 2013, pp. 429–434.

[6] R. Zhang, K. Mazumdar, B. H. Meyer, K. Wang, K. Skadron, and M. Stan, "A cross-layer design exploration of charge-recycled power-delivery in many-layer 3D-IC," in *Proc. 52nd Annu. Design Autom. Conf.*, 2015, p. 133.

[7] R. Zhang, K. Mazumdar, B. H. Meyer, K. Wang, K. Skadron, and M. R. Stan, "Transient voltage noise in charge-recycled power delivery networks for many-layer 3D-IC," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Jul. 2015, pp. 152–158.

[8] P. S. Shenoy and P. T. Krein, "Differential power processing for DC systems," *IEEE Trans. Power Electron.*, vol. 28, no. 4, pp. 1795–1806, Apr. 2013.

[9] J. McClurg, Y. Zhang, J. Wheeler, and R. Pilawa-Podgurski, "Rethinking data center power delivery: Regulating series-connected voltage domains in software," in *Proc. IEEE Power Energy Conf. Illinois (PECI)*, Feb. 2013, pp. 147–154.

[10] S. K. Lee, T. Tong, X. Zhang, D. Brooks, and G.-Y. Wei, "A 16-core voltage-stacked system with an integrated switched-capacitor DC-DC converter," in *Proc. VLSI Circuits Symp. (VLSIC)*, Jun. 2015, pp. C318–C319.

[11] E. K. Ardestani, R. T. Possignolo, J. L. Briz, and J. Renau, "Managing mismatches in voltage stacking with coreunfolding," *ACM Trans. Archit. Code Optim.*, vol. 12, no. 4, p. 43, Jan. 2015.

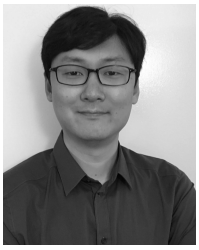
[12] W. Kim, D. Brooks, and G.-Y. Wei, "A fully-integrated 3-level DC-DC converter for nanosecond-scale DVFS," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 206–219, Jan. 2012.

[13] T. M. Andersen *et al.*, "A sub-ns response on-chip switched-capacitor DC-DC voltage regulator delivering 3.7W/mm^2 at 90% efficiency using deep-trench capacitors in 32nm SOI CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 90–91.

[14] H.-P. Le, S. R. Sanders, and E. Alon, "Design techniques for fully integrated switched-capacitor DC-DC converters," *IEEE J. Solid-State Circuits*, vol. 46, no. 9, pp. 2120–2131, Sep. 2011.

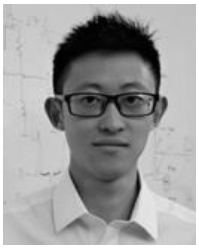
[15] T. Tong, X. Zhang, W. Kim, D. Brooks, and G.-Y. Wei, "A fully integrated battery-connected switched-capacitor 4:1 voltage regulator with 70% peak efficiency using bottom-plate charge recycling," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Sep. 2013, pp. 1–4.

[16] T. Tong, S. K. Lee, X. Zhang, D. Brooks, and G.-Y. Wei, "A fully integrated reconfigurable switched-capacitor DC-DC converter with four stacked output channels for voltage stacking applications," *IEEE J. Solid-State Circuits*, vol. 51, no. 9, pp. 2142–2152, Sep. 2016.
[17] M. E. Dean, "Strip: A self-timed RISC processor," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 1992.
[18] A. Grenat, S. Pant, R. Rachala, and S. Naffziger, "Adaptive clocking system for improved power efficiency in a 28nm x86-64 microprocessor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 106–107.
[19] C. R. Lefurgy *et al.*, "Active guardband management in power7+ to save energy and maintain reliability," *IEEE Micro*, vol. 33, no. 4, pp. 35–45, Jul. 2013.
[20] B. Zimmer *et al.*, "A RISC-V vector processor with tightly-integrated switched-capacitor DC-DC converters in 28nm FDSOI," in *Proc. Symp. VLSI Circuits (VLSI Circuits)*, Jun. 2015, pp. C316–C317.
[21] C. Tokunaga *et al.*, "A graphics execution core in 22nm CMOS featuring adaptive clocking, selective boosting and state-retentive sleep," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 108–109.
[22] G.-Y. Wei and M. Horowitz, "A fully digital, energy-efficient, adaptive power-supply regulator," *IEEE J. Solid-State Circuits*, vol. 34, no. 4, pp. 520–528, Apr. 1999.
[23] T. D. Burd, T. A. Pering, A. J. Stratakos, and R. W. Brodersen, "A dynamic voltage scaled microprocessor system," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1571–1580, Nov. 2000.
[24] X. Zhang, T. Tong, D. Brooks, and G.-Y. Wei, "Supply-noise resilient adaptive clocking for battery-powered aerial microrobotic system-on-chip in 40nm CMOS," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Sep. 2013, pp. 1–4.
[25] Y. Ikenaga *et al.*, "A 27% active-power-reduced 40-nm CMOS multimedia SoC with adaptive voltage scaling using distributed universal delay lines," *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 832–840, Apr. 2012.
[26] K. A. Bowman, C. Tokunaga, T. Karnik, V. K. De, and J. W. Tschanz, "A 22nm dynamically adaptive clock distribution for voltage droop tolerance," in *Proc. Symp. VLSI Circuits (VLSIC)*, 2012, pp. 94–95.
[27] B. Reagen, R. Adolf, Y. S. Shao, G.-Y. Wei, and D. Brooks, "Machsuite: Benchmarks for accelerator design and customized architectures," in *Proc. IEEE Int. Symp. Workload Characterization (IISWC)*, Oct. 2014, pp. 110–119.



Sae Kyu Lee (S'10) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, the M.S. degree in electrical and computer engineering from The University of Texas at Austin, Austin, TX, USA, and the Ph.D. degree from Harvard University, Cambridge, MA, USA.

He was with Intel Corporation, Austin, TX, USA, and Advanced Micro Devices, Boxborough, MA, USA, where he has been involved in mobile microprocessors designs. His current research interests include VLSI design for efficient on-chip power delivery solutions.



Tao Tong (S'10) received the B.E. degree from Tsinghua University, Beijing, China, the M.S. degree from Oregon State University, Corvallis, OR, USA, and the Ph.D. degree from Harvard University, Cambridge, MA, USA.

He was with MediaTek Wireless Inc., Woburn, MA, USA, and Lion Semiconductor Inc., San Francisco, CA, USA, designing analog-to-digital converters and fully integrated dc-dc converters for mobile applications. His current research interests include integrated voltage regulators and their applications in energy efficient computing systems.

Xuan Zhang (M'11) received the B.Eng. degree in electrical engineering from Tsinghua University, Beijing, China, in 2006, and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2011.

She was a Post-Doctoral Fellow with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, where she has been involved in the design of the brain for an insect-scale microrobot. She is currently an Assistant Professor with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA. She is involved in the fields of robotics, system engineering, VLSI, and computer architecture. Her current research interests include miniaturization and optimization of autonomous systems for performance, reliability, security, and energy efficiency, with diverse applications in microrobotics, Internet-of-Things, wearable/implantable devices, ubiquitous computing, and resilient cyber-physical systems.

Dr. Zhang has received an Intel Fellowship from 2008 to 2009 and a recipient of the Design Contest Award at the IEEE ISLPED in 2013.

David Brooks (F'16) received the B.S. degree from the University of Southern California, Los Angeles, CA, USA, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, USA, all in electrical engineering.

He was a Research Staff Member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. He is currently the Haley Family Professor of Computer Science with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His current research interests include resilient and power efficient computer hardware and software design for high performance and embedded systems.

Prof. Brooks has received several honors and awards, including the ACM Maurice Wilkes Award, the ISCA Influential Paper Award, the NSF CAREER Award, the IBM Faculty Partnership Award, and the DARPA Young Faculty Award.

Gu-Yeon Wei (M'00) received the B.S., M.S., and Ph.D. degrees from Stanford University, Stanford, CA, USA, in 1994, 1997, and 2001, respectively, all in electrical engineering.

He is currently the Gordon McKay Professor of Electrical Engineering and Computer Science with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His current research interests span multiple layers of a computing system: mixed signal integrated circuits, computer architecture, design tools for efficient hardware, identifying synergistic opportunities across these layers to develop energy-efficient solutions for a broad range of systems from flapping-wing microrobots to large-scale servers.